Borderline consciousness, phenomenal consciousness, and artificial consciousness: a unified approach

Chuanfei Chin
Trinity College

Faculty of Philosophy
University of Oxford

Submitted for the degree of Doctor of Philosophy
Michaelmas 2015

Borderline consciousness, phenomenal consciousness, and artificial consciousness: a unified approach

Chuanfei Chin
Trinity College

Submitted for the degree of Doctor of Philosophy
Michaelmas 2015

Abstract

Borderline conscious creatures are neither definitely conscious nor definitely not conscious. In this thesis, I explain what borderline consciousness is and why it poses a significant epistemological challenge to scientists who investigate phenomenal consciousness as a natural kind. When these scientists discover more than one overlapping kind in their samples of conscious creatures, how can they identify *the* kind to which all and only conscious creatures belong? After assessing three pessimistic responses, I argue that different groups of scientists can legitimately use the concept of phenomenal consciousness to refer to different kinds, in accord with their empirical interests. They can thereby resolve three related impasses on the status of borderline conscious creatures, the neural structure of phenomenal consciousness, and the possibility of artificial consciousness.

The thesis has three parts: First, I analyse the concept of borderline consciousness. My analysis counters several arguments which conclude that borderline consciousness is inconceivable. Then I explain how borderline consciousness produces the multiple kinds problem in consciousness science. Second, I assess three recent philosophical responses to this problem. One response urges scientists to eliminate the concept of consciousness, while another judges them to be irremediably ignorant of the nature of consciousness. The final response concludes that scientific progress is limited by the concept's referential indeterminacy. I argue that these responses are too pessimistic, though they point to a more promising approach. Third, I propose that empirically constrained stipulation can solve the multiple kinds problem. Biologists face the same problem because of their longstanding controversy over what counts as a species. Building on new arguments for stipulating the reference of species concepts, I demonstrate that this use of stipulation in biology is neither epistemologically complacent nor metaphysically capricious; it also need not sow semantic confusion. Then I defend its use in consciousness science. My approach is shown to be consistent with our understanding of natural kinds, borderline cases, and phenomenal consciousness.

Acknowledgements

The research for this thesis was supported by a PhD scholarship from the British Society for the Philosophy of Science.

Various chapters from the thesis were presented at these conferences:
(a) *The Collective Dimension of Science* (2011), Archives Poincaré, University of Nancy;
(b) *Workshop on History and Philosophy of Science in Practice* (2012), Tembusu College, National University of Singapore;
(c) *Knowing, Making, Governing: Asia-Pacific Science, Technology, and Society Network Biennial Conference* (2013), Tembusu College, National University of Singapore;
(d) *Phenomenal Mindreading: Attributing Conscious Experiences to Oneself and Others* (2013), Ruhr-Universitat Bochum;
(e) *42nd IUC Philosophy of Science Conference* (2015), Inter-University Centre, Dubrovnik.

I thank my supervisors Bill Child, Tim Bayne, and Katherine Morris. I also thank Gavin Maughfling, Lim Chong Ming, and Alexandra Serrenti.

Contents

Chapter 1

Introduction

1.1 Borderline consciousness: what it is and why it matters

Are octopuses conscious? The answer is not obvious if our question is about phenomenal consciousness – the kind of consciousness that philosophers and scientists associate with feeling pain and pleasure, perceiving the world from a perspective, and experiencing emotions. Looking around, we can point to friends who are conscious in this phenomenal sense, and flora which are not. But do octopuses have feelings and other conscious experiences? Do they count, in some sense, as subjects? When we consider their behaviour and brains, octopuses seem to be neither definitely conscious nor definitely not conscious. We can describe them as *borderline conscious*.

Other borderline conscious creatures include fishes, frogs, late-stage human foetuses, and some vegetative-state patients. Allen (2011) reports that, for most philosophers and scientists who study phenomenal consciousness, 'reptiles, amphibians, and fish constitute an enormous grey area' (§1). Borderline cases of consciousness are akin to more familiar examples of vagueness. For instance, a man is borderline bald if he is neither definitely bald nor definitely not bald. A few grains of sand make a borderline heap because they are neither definitely a heap nor definitely not a heap. We recognise these to be vague phenomena on the basis of our hesitant responses to them. And we can do so without adopting any view about

the fundamental nature of their vagueness. More specifically, recognising these borderline cases does not commit us to the further claim that there is no fact of the matter about their status as conscious creatures, bald men, or heaps.

Indeed, we are inclined to believe that there must be a fact of the matter whether a borderline conscious creature has conscious experience at all. That is why we expect science to classify, more definitely, borderline conscious creatures. In this respect, borderline cases of consciousness differ from the other examples of vagueness. We do not expect science to help similarly with borderline bald people or borderline heaps. To borrow a phrase from Sorensen (2012): these more familiar cases are 'inquiry resistant'. It does not seem possible that more conceptual analysis or empirical research will discover how few strands of hair a person must have to be bald, or how many grains of sand are needed to make a heap. For practical purposes, we might stipulate that baldness requires a certain amount of hair, or that a heap is made from a minimum number of grains; but no one imagines that these stipulations advance scientific inquiry into baldness or heaps.

In this thesis, I shall clarify what borderline consciousness is and explain why its study is central to the science of consciousness. Borderline consciousness poses a significant epistemological challenge to scientists who investigate phenomenal consciousness as a natural kind. When these scientists discover more than one overlapping kind in their samples of conscious creatures, how can they identify *the* kind to which all and only conscious creatures belong?

As I will show, this multiple kinds problem leads to at least three theoretical impasses on the nature of phenomenal consciousness. These impasses arise in debates on the neural structure of phenomenal consciousness (Irvine 2013), the development of foetal consciousness (Derbyshire & Raja 2011; Chin 2011), and the possibility of artificial consciousness (Prinz 2003, 2005). My arguments will focus on neural or neurofunctional theories of phenomenal consciousness (Crick & Koch 1990, 2003), although they may apply more widely to other theories that leave room for neural and functional constraints.

I will propose that empirically constrained stipulation can solve the multiple kinds problem. Through stipulation, different groups of scientists can legitimately use the concept of phenomenal consciousness to refer to different kinds. Each group of scientists will refer to one of the overlapping kinds, in accord with their empirical interests. After stipulation, each group can classify at least some borderline conscious creatures more precisely, in accord with their conceptual decisions. By adopting this approach, scientists can resolve their impasses on borderline consciousness, phenomenal consciousness, and artificial consciousness.

My thesis argues that this approach is more viable than it seems to many philosophers and scientists. I shall draw on and develop our philosophical resources in three areas: natural kinds, borderline cases, and phenomenal consciousness. Since stipulation is rarely mentioned in the epistemology of natural kinds, I will demonstrate how biologists use it to solve the multiple kinds problem. I will also explain why their use of stipulation makes sense in metaphysical and semantic terms. Then I want to defend the use of stipulation in consciousness science. In

particular, I will argue that its use is consistent with philosophical accounts of vagueness and current conceptions of phenomenal consciousness.

1.2 Four objections to stipulation

Using stipulation to re-classify borderline conscious creatures may strike some as *obviously* wrong – so wrong that it needs no debate. What intuitions and arguments support this stance? They are not so obvious. My contribution to the debate includes distinguishing four key objections to this use of stipulation, and clarifying their philosophical grounds. I plan to respond to these objections after I develop an account of stipulation's role in the epistemology of natural kinds. But let me sketch them here since they highlight the aims of my account.

First, it is said that stipulation is *naturalistically unsound*. This objection is based on both the methodological and ontological stances in philosophical naturalism (Papineau 2007; Jacobs 2009). According to the methodological stance, we should respect the established methods of science, especially empirical testing and inference to the best explanation. Stipulation will only hinder scientific discovery. De Brigard and Prinz (2010) warn explicitly of this danger: those who stipulate 'precisified definitions' of consciousness and other related concepts 'often replace the folk concept with idiosyncratic definitions that settle crucial questions by fiat rather than facilitating the process of scientific investigation and discovery' (52).

The ontological stance reinforces this distrust of stipulation. Natural kinds are groupings that reflect 'the structure of the natural world rather than the interests and actions of human beings' (Bird & Tobin 2015). Because they reflect the world's structure, natural kinds ground scientific explanations and inductions about the world. If phenomenal consciousness is a natural kind, then its nature awaits scientific discovery. Stipulation will only produce an artificial kind with limited use in making explanations and inductions. Block (2002) spells out this stance. According to him, someone who believes that phenomenal consciousness is real and has a 'scientific nature' should not define it through a 'decision' on 'extrapolating a concept of consciousness grounded in our physical constitution to other physical constitutions' (421).

Second, it might be said that stipulation is *metaphysically presumptuous* about vagueness. Stipulation is licensed by only two controversial theories of vagueness. According to the first theory, our semantic indecision explains borderline cases (Lewis 1993; Barnes 2009). So we have yet to decide on the exact boundaries of the vague term 'conscious'. The second theory holds that a special kind of context sensitivity explains borderline cases (Soames 1999; Shapiro 2006; Åkerman 2012). For instance, the boundaries of the term 'conscious' may depend on the speaker's judgements in each context. If either theory is correct, then scientists may legitimately stipulate more precise boundaries for the term.

However, there are at least two other theories of vagueness, with different epistemological, metaphysical, and semantic implications. One of them claims that we are irremediably ignorant of where the vague term's precise boundaries lie

(Williamson 1994; Sorensen 2006). Another blames metaphysical indeterminacy – indeterminacy in the world itself, rather than in our representation of the world (Williams 2008; Barnes 2009). Neither theory supports stipulation as a response to borderline cases. To recommend that scientists stipulate in consciousness science, must we presume that borderline consciousness arises from semantic indecision or context sensitivity?

Third, it is suggested that stipulation is *semantically shifty*. Even if we put aside worries about the roots of vagueness, stipulation is a risky response to it (Williamson 1994: 214; Prinz 1998, §40; Sorensen 2012, §1). It is likely to shift the reference of the term 'phenomenal consciousness', and so abruptly change the subject of consciousness science. Again Block (2002) warns: 'Stipulations need not stick when it comes to the phenomenal realist conception of consciousness' (422). After scientists use stipulation to re-classify borderline conscious creatures, our original question about their experience will arise in different terms. For instance, we might ask if the so-called phenomenal consciousness of a re-classified creature '*feels the same*' as our phenomenal consciousness.

Fourth, it is said that stipulation is *subjectively absurd*. This objection denies that subjectivity can be treated like species. Even if stipulation has a role in the epistemology of other natural kinds, it cannot be used for phenomenal consciousness. From introspection, it seems obvious that we either have or do not have conscious experience. Our specific states of experience may be faint or fleeting. But there is a fact of the matter whether we have experience at all. Therefore,

phenomenal consciousness is not a kind whose boundaries depend on someone else's decision.

Chalmers (1996) draws on intuition to support this objection: 'Does a mouse have conscious experience? Does a virus? These are not matters for stipulation. Either there is something that it is like to be a mouse or there is not, and it is not up to us to define the mouse's experience into or out of existence' (105). Here is how Papineau (2002) describes our intuitive view about the status of octopuses and intelligent robots: 'surely, it seems, there must be a fact of the matter whether it is *like anything at all* for such creatures' (202). In Papineau (2003), he uses a metaphor to illuminate this view: 'Surely a light is on, or it is not' (219).

My defence against these charges of naturalistic unsoundness, metaphysical presumption, semantic shiftiness, and subjective absurdity will draw on the arguments made in the following chapters. To those who are undecided about stipulation's role in scientific classification, these arguments offer grounds to support it. To those who are already sympathetic, these arguments clarify the conditions under which stipulation can contribute to scientific classification, especially in biology and consciousness science. Finally, to those who are set against any use of stipulation, the arguments indicate that their objections need to be strengthened.

1.3 Thesis outline

My thesis has three parts. The first part comprises Chapters 2 and 3. It builds a conceptual framework on borderline consciousness, in order to broach the epistemological problems that borderline consciousness poses in consciousness science. The second part, consisting only of Chapter 4, evaluates three philosophical responses to the multiple kinds problem in consciousness science. The final part, Chapters 5 and 6, explains my unified approach to the impasses on borderline consciousness, phenomenal consciousness, and artificial consciousness. It argues that empirically constrained stipulations can solve the multiple kinds problem in biology, then assesses whether they can solve the same problem in consciousness science.

Here is a more detailed outline of the following chapters. Chapter 2 offers a conceptual analysis of borderline consciousness. First, I analyse the concept of borderline consciousness in terms of phenomenal consciousness and borderline cases. My analysis explains how our ordinary classification of creatures with respect to phenomenal consciousness produces borderline cases. Second, I defend this concept of borderline consciousness against two challenges by McGinn (1996) and Antony (2006b, 2008). Both argue that borderline consciousness is inconceivable – McGinn on conceptual grounds, Antony on psychological ones. My defence draws on conceptual resources from my analysis. At the same time, it clarifies the metaphysical and methodological commitments of my analysis.

Chapter 3 turns to two epistemological challenges that borderline consciousness poses. First, I discuss the problem of classifying borderline conscious creatures more precisely. I explain how scientists can solve this problem by building neural theories of phenomenal consciousness. Then I clarify some epistemological norms and metaphysical assumptions behind this strategy. Following Block (2007a, b) and Shea and Bayne (2010), I interpret it to be the investigation of phenomenal consciousness as a natural kind. Second, I explain why borderline consciousness leads scientists to discover multiple kinds in their samples of conscious creatures. To clarify this problem, I contrast it with the phenomenon of multiple realisability in mental kinds. Third, I demonstrate how this multiple kinds problem produces three theoretical impasses on the neural structure of phenomenal consciousness (Irvine 2013), the development of foetal consciousness (Derbyshire & Raja 2011; Chin 2011), and the possibility of artificial consciousness (Prinz 2003, 2005).

Chapter 4 explains and evaluates three pessimistic responses to the multiple kinds problem in consciousness science. They are from Irvine (2013), Prinz (2003, 2005), and Papineau (2002, 2003). First, I address Irvine's recommendation that scientists eliminate the concept of consciousness from their research. Second, I assess Prinz's view that scientists are irremediably ignorant of the nature of phenomenal consciousness, even though they can discover a lot about human consciousness. Third, I examine Papineau's view that the concept of consciousness-as-such suffers from a special form of referential indeterminacy, which is tied to its function in tracking human consciousness. My evaluation counters their pessimism about the science of phenomenal consciousness, and brings up the possibility that scientists

can legitimately stipulate that the concept of phenomenal consciousness refers to one of the multiple kinds.

Chapter 5 demonstrates that empirically constrained stipulation can solve the multiple kinds problem in biological classification. First, I explain why the Kripke-Putnam model of natural kinds creates a presumption against stipulation. In this model, stipulation brings the risks of metaphysical caprice, epistemological complacency, and semantic confusion. Second, I explain why the 'species problem' among biologists means that they face the multiple kinds problem too. Then I evaluate the arguments in LaPorte (2004, 2010) that support stipulating the reference of species concepts. While his arguments help to clarify the biologists' disagreement, they do not make a convincing case. Third, I propose a new epistemological argument for stipulation. My argument is based on biological practices described by biologists (Cracraft 2000; Coyne & Orr 2004; de Queiroz 2007) and philosophers of biology (Stanford 1995; Ereshefsky 1992, 2010). I infer that, under certain conditions, biologists can legitimately stipulate the reference of species concept according to their empirical interests. Fourth, I make sense of their stipulations from the metaphysical and semantic perspectives.

Chapter 6 defends the use of empirically constrained stipulation in consciousness science. It also highlights strengths and weaknesses in my arguments, and identifies areas for further philosophical research. First, I draw on the arguments in Chapter 5 to show that this use of stipulation is naturalistically sound on both methodological and metaphysical grounds. Second, I argue that its use is consistent with all four theories of vagueness in the philosophical literature. Third, I examine its semantic

consequences. It need not abruptly shift the subject of consciousness science. Fourth, I argue that its use is consistent with our introspective and intuitive views about the determinacy of conscious experience. I conclude with a set of conceptual considerations that support my approach to the multiple kinds problem in consciousness science. Drawing on Lycan (2004), de Sousa (2004), Godfrey-Smith (2013), and others, I distinguish qualitative consciousness, perspectival consciousness, and first-personal consciousness as different modes of subjectivity.

By the end of this thesis, I will have explained why the study of borderline conscious creatures is central to consciousness science. I will also have explained how scientists can use stipulation to solve the multiple kinds problem in consciousness science. Under the right conditions, stipulation is neither naturalistically unsound, metaphysically presumptuous about vagueness, semantically shifty, nor subjectively absurd. It is the basis of my unified approach to the current impasses on borderline consciousness, phenomenal consciousness, and artificial consciousness.

Chapter 2

Borderline consciousness: a conceptual analysis

2.1 Introduction

In Chapter 1, I defined borderline consciousness as a kind of vagueness about conscious experience. It is attributed to creatures that are neither definitely conscious nor definitely not conscious. I also claimed that borderline consciousness poses a significant epistemological challenge to scientists who investigate the nature of phenomenal consciousness. Before I examine this epistemological challenge, I need to clarify the concept of borderline consciousness.

Why is a conceptual analysis needed? Some philosophers find the idea of borderline consciousness more troubling than I do. McGinn (1996) even claims that borderline consciousness is inconceivable:

> …the concept of consciousness does not permit us to conceive of genuinely borderline cases of sentience, cases in which it is inherently indeterminate whether a creature is conscious: either a creature definitely is conscious or it is definitely not. (14)

However, others happily allow for the possibility of borderline consciousness. Block (1992), for instance, claims that 'one who accepts phenomenal consciousness can also accept all sorts of borderline cases between consciousness and nonconsciousness' (176). In Block (2002b), he adds that a fish may be 'a borderline case of consciousness' (421). Despite these conflicting attitudes, few philosophers have attempted to analyse the concept of borderline consciousness. So there is no

clear basis from which we can assess their arguments on this concept.[1]

Here I shall analyse the concept of borderline consciousness from a naturalist perspective and defend it against two challenges. First, I analyse the concept in terms of phenomenal consciousness and borderline cases. This analysis explains how our ordinary classification of creatures with respect to phenomenal consciousness produces borderline cases. Second, I test my analysis against the conceptual challenge in McGinn (1996). According to him, the conceptual analysis of consciousness shows that it has no borderline cases. Third, I assess the psychological challenge by Antony (2006b, 2008). He argues that our understanding of the concepts of conscious state and conscious creature does not meet two psychological requirement for conceiving borderline cases. I argue that the concept of borderline consciousness in my analysis is not threatened by their challenges, though they help to clarify the metaphysical and methodological commitments of my analysis.

2.2 Phenomenal consciousness and borderline cases

What is it for us to have a concept of conscious experience which has borderline cases?[2] I shall analyse the concept of borderline consciousness in terms of

[1] I learnt most from the recent analyses in Antony (2006b, 2008). Besides Block, other philosophers admit the possibility of borderline consciousness: see Papineau (1993), §4.8; Papineau (2002), ch. 7; Papineau (2003); Shea and Bayne (2010), §5.1; Tye (1997), §8.1; and Unger (1988). Those who deny that the concept of consciousness has borderline cases include Antony (2006b, 2008); Chalmers (1996), 105; McGinn (1996), ch. 1; and Strawson (2009), §6.4.

For further references, see Antony (2008), notes 2 and 6, though some on his lists focus on *degrees* of consciousness rather than *borderline cases*. I argue later that these should be distinguished.

[2] I modeled this question after Peacocke (1984), 97: 'What is it to have the concept of a type of experience which can be enjoyed by myself and by others?' This conceptual problem of other minds is elaborated in Avramides (2000), ch. 8, and Hyslop (2014), §1.2.

phenomenal consciousness and borderline cases. First, I define phenomenal consciousness through a set of conceptual distinctions and connections. Then I explain how our operational criteria for attributing phenomenal consciousness lead to borderline cases. Second, I clarify what borderline cases are and how they are related to vagueness. I survey four philosophical theories of vagueness, and show that they do not warrant a method for re-classifying borderline conscious creatures.

My analysis is carried out from a naturalist perspective, based on our current understanding of phenomenal consciousness and borderline cases. Its results are neither analytic truths about the meaning of the term 'borderline consciousness' nor necessary truths about borderline consciousness. They are more aptly seen as provisional clarifications about how we understand the concept of borderline consciousness and how we use it. Both philosophical and scientific advances can prompt us to revise this analysis.[3] Indeed, I will be proposing a significant revision at the end of this thesis.

Phenomenal consciousness

When a creature is borderline conscious, it is vague whether the creature has conscious experience. This experience is what Block (1995a) calls *phenomenal consciousness*.[4] So I shall focus on this concept of consciousness. Phenomenal consciousness can be attributed to a creature or its psychological states. A creature's

[3] This stance on conceptual analysis is defended by Kitcher (1992), 76: 'conceptual clarification can play a valuable role within the naturalist enterprise, even though it is clearly understood that the concepts in question might be superseded' (note 69). See also Papineau (2007), §2.2.
[4] His account is revised and defended in Block (1995b) and (2002a).

phenomenal consciousness is determined by its phenomenally conscious states. By Block's definition,

> P-conscious states are experiential states, that is, a state is P-conscious if it has experiential properties. (166)

The sum of a state's phenomenal properties constitutes the phenomenal content of that state. It is 'what it is like' to be in that state. And the sum of a creature's phenomenal states determines what it is like to be that creature: 'What it is for there to be something it is like to be me, that is for me to be P-conscious, is for me to have one or more states that are P-conscious' (179).

Suppose we describe what is it like to be in pain, to feel an itch, and to see, hear, smell, or taste something. Then we are referring to the phenomenal properties of our sensations and perceptions. Block believes that our thoughts, desires, and emotions also have phenomenal properties. All these properties create an 'explanatory gap' between consciousness and the rest of nature (167).

I am provisionally construing a creature's phenomenal consciousness as a determinable, which has phenomenal states as determinates.[5] In turn, these phenomenal states form a complex structure – depending on the nature of the phenomenal properties that they bear. Loar (1997) notes that such properties range from highly determinate phenomenal properties to 'phenomenal determinables' (§10). His examples of this range are drawn from visual experience: the properties of

[5] On the logical distinction between determinables and determinates, see Sanford (2014). I follow Papineau (1993, 2002, 2003) and Bayne (2007, 2009a) in applying it to phenomenal consciousness.
As Bayne notes, this concept of creature consciousness must not be confused with another which equates it with wakefulness. See also Van Gulick (2014), §2.1, and Piccinini (2007).

being crimson, dark red, red, warm coloured, coloured, and visual. But, as Van Gulick (2014) reminds us, phenomenal structure includes 'much of the spatial, temporal and conceptual organization of our experience of the world and of ourselves as agents in it' (§2.2). Let me put aside the philosophical debate on which states and properties fall within the 'reach' of phenomenal consciousness.[6] My focus here is on whether a creature has conscious experience *at all*.

To clarify what phenomenal consciousness is, Block (2002a) distinguishes it conceptually from access consciousness and self-consciousness.[7] First, as he defines it, *access consciousness* is a sophisticated kind of information-processing:

> A representation is A-conscious if it is broadcast for free use in reasoning and for direct 'rational' control of action (including reporting) (278).

He cites empirical and introspective evidence which suggests that a psychological state can be phenomenally conscious without being access conscious. It also seems conceptually possible that a psychological state can be access conscious without being phenomenally conscious. Second, *self-consciousness* is the 'possession of the concept of the self and the ability to use this concept in thinking about oneself' (Block 2002a: 287). Many philosophers and scientists agree that human neonates, monkeys, and dogs have experience. Yet they do not believe that these creatures have the conceptual sophistication required for self-consciousness.

[6] Some key issues are surveyed in Bayne (2009a), 479-481, Bayne (2009c), and Tye (2015), §2.
[7] As Block (1995a) notes, his conceptual distinctions do not rule out the 'empirical possibility' that phenomenal consciousness is identical to some form of information processing (202, note 10). He even concedes that access consciousness is a 'good candidate for a reductionist identification' with phenomenal consciousness (172). Block (2007b) adds that phenomenal consciousness entails a minimal level of accessibility, which he calls 'awareness' (536). However, this awareness is distinct from the cognitive accessibility needed for reporting.

Apart from these conceptual distinctions, philosophers also use the concepts of subjectivity and qualia to point to phenomenal consciousness.[8] What is meant by *subjectivity* here? It is the orientation of a creature's psychological states to a perspective. According to Nagel (1974), a subject of experience must have a perspective on the world. And this subjectivity is fundamental to consciousness:

> …fundamentally an organism has conscious mental states if and only if there is something that it is like to *be* that organism – something it is like *for* the organism. (519)

As he makes clear in a note, 'what it is like' to be a creature refers to 'how it is for the subject himself' (526). And facts about what it is like to be a creature 'embody a particular point of view' belonging to that creature (522). It follows that a creature's psychological states are subjective insofar as they are tied to its perspective.

So far as I can tell, Nagel recognises three levels of subjectivity. The most general level consists of 'the fact that an organism has conscious experience *at all*': this implies that there is at least 'something' it is like to be that organism (519). The intermediate level is the subjectivity associated with a particular species, such as 'what it is like to be a human being, a bat, or a Martian' (522). The most specific level of subjectivity consists of facts about an individual's experience – 'facts about what is it like *for* the experiencing organism' (522).

Qualia refers to the raw feels or feelings of experiences. The term 'raw feels' was originally used to describe those aspects of our psychological states that science

[8] Here I draw on the surveys of consciousness in Manson (2007), Bayne (2009a), Van Gulick (2014).

could not capture.[9] As Dennett (1988) notes, some philosophers continue to attribute to qualia the following set of scientifically problematic properties: ineffable, intrinsic, private, directly or immediately apprehensible in consciousness. Tye (2015) cites other philosophers who simply define qualia as the non-representational properties of experience. In this sense, qualia are most commonly associated with sensations and perceptual experiences, though some believe that they are present in cognitive and affective states.

However, the concept of qualia can be used more broadly. For example, Chalmers (1996) applies the concept of qualia to all experiences.

> "To be conscious" in this sense is roughly synonymous with "to have qualia," "to have subjective experience," and so on. Any differences in the class of phenomena picked out are insignificant. (6)

According to Chalmers, a psychological state is conscious if it has 'a *qualitative* feel – an associated quality of experience' (4). Such qualitative feels are what he calls 'phenomenal qualities, or *qualia* for short'. In a recent anthology on phenomenal consciousness, Alter and Walter (2007) describe 'conscious experience' similarly, in terms of 'phenomenal qualities from the experiencing subject's perspective' (3).

Neither subjectivity nor qualia are more obviously definable than phenomenality. In fact, philosophers connect these concepts – as mutually supportive means to point to the distinctive character of experience (Manson 2007: §13.2; Bayne 2009a; Van Gulick 2014, §4). Their glue is often Nagel's evocative phrase. Thus, in his definition of phenomenal consciousness, Block (1995a) mentions 'what it is like' to be in a

[9] For the origins of the term, see Kirk (1996), vi and 28. Smart (1959) already refers to 'the singular elusiveness of "raw feels" – why no one seems to be able to pin any properties on them' (150).

conscious state, such as what it is like to hear a sound as coming from the right. Chalmers claims that 'a mental state is conscious if there is something it is like to be in that mental state' (1996: 5). He defines qualia as 'those properties that characterize conscious states according to what it is like to have them' (2002: 268). Prinz (2003) refers to 'qualitative, phenomenal feelings – Nagel's famous what-it's-likes' (112). Tye (2013) defines the phenomenal character of experience as 'what it is like subjectively to undergo the experience' (§1).

Aside from their consensus that Nagel has highlighted something distinctive about experience, these philosophers disagree profoundly on the nature of experience. For now, let me acquiesce in their shared assumption: anyone who asks about the phenomenal or qualitative aspects of conscious states is attending to the subjective character of experience. Thus, to ask whether a creature is conscious in the phenomenal sense is to ask if it is a subject of experience, bearing a distinctive perspective on the world. By the end of this thesis, I will be in a better position to contest Chalmers' claim that there are no significant differences in the phenomena picked out by these concepts.

Thus far, I have relied on a set of contrasts and connections to analyse the concept of phenomenal consciousness. My analysis is not yet complete. I will also describe the basic epistemic conditions that we use to apply this concept of consciousness.[10]

[10] I take the phrase 'basic epistemic conditions' from Burge (1993). He distinguishes *concepts* from *conceptual definitions*. The former are components of thought contents; the latter explicate our understanding of concepts. 'Definitions also state basic epistemic conditions that the individual has for applying the concept, or the individual's best understanding of conditions for falling under the concept' (293). See also Burge (2007), 24-26.

Like Burge, I do not assume that a concept is constituted by its conditions of application. I include these conditions only in the provisional analysis of a concept.

What are these conditions? They correspond to the operational criteria that we ordinarily use to classify friends as conscious, fauna as not conscious, and octopuses as borderline conscious. We rely on these rough criteria now, though we expect science to revise them eventually.

Our operational criteria are primarily behavioural and neural. To be classified as having experience at all, a creature must be similar enough, in behaviour and brain, to conscious humans. We use *behavioural similarity to conscious humans* as one criterion. Humans use various behaviours to express their conscious states – say, when they suffer pain, feel an itch or a tickle, see a bright colour, smell a bad odour, or feel concern for others. These behaviours vary along many dimensions. They are not limited to bodily movements defined in non-psychological terms. So they may include, for instance, intentional action.[11]

However, behavioural similarity is not enough. *Neural similarity to conscious humans* serves as the other criterion. Several thought-experiments indicate that we are loath to attribute consciousness to artificial systems that behave exactly as conscious humans do, but lack our brains. For instance, Block (1978) raises a doubt about the experience of homunculi-headed systems that are behaviourally and functionally equivalent to conscious humans.[12] More recently, Knobe (2008) cites two experimental studies: they confirm our reluctance to attribute emotional experiences to robots that behave exactly as we do. Huebner (2010) confirms that neurophysiology plays a role in our ordinary attributions of experiences.

[11] See Block (1995b), 234: 'purposive action is evidence of [access consciousness], but it is also evidence, albeit indirect evidence, of [phenomenal consciousness].'
[12] He adds that 'our intuitions are in part controlled by the not unreasonable view that our mental states depend on our having the psychology and/or neurophysiology we have' (76).

By logical standards, these joint criteria are circular because they explicated in relation to conscious humans. By scientific standards, the criteria are imprecise. We are, in effect, using conscious humans as a 'paradigm case' of consciousness.[13] To classify another creature as conscious or not, we check how far it is behaviourally and neurally similar to this paradigm case.

These criteria explain our provisional classification of human and non-human species. We are fortunate: our judgements of behavioural and neural similarities agree enough for the criteria to be useful in three ways. First, the criteria secure a near-universal consensus that normal human adults and children are conscious, while plants are not. Second, they support a widespread consensus that mammals such as apes, monkeys, and dogs are conscious, while ants and aphids are not. Third, they highlight a set of borderline conscious creatures.[14] As I noted in §1.1, these typically include octopuses, fishes, frogs, late-stage human foetuses, and some vegetative-state patients. According to the criteria, they are neither definitely conscious nor definitely not conscious.

Borderline cases

Next, I shall clarify what borderline cases are and how they are related to

[13] See Bayne (2009a), 485-8, on the 'paradigm-case' approach to ascribing consciousness. He notes that this approach is 'impotent exactly where it is most needed', when scientists are dealing with 'hard cases'. McLaughlin (2003) highlights the 'anthropocentric' aspect of our concept of consciousness: 'we assume that we are justified in attributing consciousness to other beings if and only if they are relevantly like us' (182).

[14] The boundaries of the term 'borderline conscious creatures' may be vague: insects, for instance, seem to be borderline cases of borderline conscious creatures. I shall put aside this problem of higher-order vagueness: it afflicts vague terms generally, and does not affect my arguments on borderline consciousness.

vagueness.[15] A case that is borderline P is neither definitely P nor definitely not P – where 'definitely' is construed in epistemic terms, to capture our inability to offer a verdict one way or another. Let me start with some familiar borderline cases. Some people are neither definitely bald nor definitely not bald. A few grains of sand neither definitely make a heap nor definitely do not make a heap. Each of these examples depend on only one dimension of variation, either quantity of hair or of sand. But borderline cases of other phenomena may be more complicated. First, they may involve more than one dimension. Second, these dimensions may not be well specified. Niceness, for instance, is clearly multi-dimensional, though we have no 'clear-cut' list of its dimensions (Keefe & Smith 1996: 5). Some people are borderline nice 'by scoring well in some relevant respects but not in others'.

Borderline cases of consciousness are similarly complicated. As I noted above, we rely on two operational criteria to attribute phenomenal consciousness to other creatures: behavioural and neural similarities to conscious humans. These similarities depend on more than one dimension of behaviour and neurophysiology. We have yet to codify or demarcate these dimensions clearly. But we know that some creatures are borderline conscious because they score well in some dimensions and not in others.

What do all borderline cases have in common? Borderline cases do not just provoke doubt or uncertainty during classification. They prompt a distinctive pattern of responses. Asked if a borderline case falls under the relevant concept, we are loathe to say that it definitely does or definitely does not – even though we have the same

[15] My analysis draws on Williamson (1994, 2005). Other sources include Keefe and Smith (1996), Keefe (2000), Williams (2008), Barnes (2009), Sorensen (2012), and Åkerman (2012).

kinds of information about this case which would support our confident judgements about other, more definite, cases. Rather we 'hesitate in judging either way or deny both judgements or disagree with each other or change our mind over time' (Keefe 2000: 43). These responses explain why we want to say that the answer is 'indefinite' or 'indeterminate' in some sense, even that there may be 'no fact of the matter' at stake (Barnes 2010).[16]

Vagueness is defined in terms of such borderline cases. In philosophy, a term is vague 'to the extent that it has borderline cases' (Sorensen 2012). Relatedly, a concept is vague if its application leads to borderline cases. This is distinct from ordinary practice, which sometimes attributes vagueness to an under-specified or ambiguous term.[17] By citing examples of borderline bald men or borderline heaps, we can point to vague phenomena without committing to a view about the fundamental roots of vagueness. In particular, as Williamson (1994), Wright (1995), and Keefe (2000) emphasise, our pre-theoretical description of these borderline cases should be neutral on whether there is a fact of the matter about their status as bald men or heaps.

Is there a fact of the matter at stake? Philosophers disagree on this very question.

[16] See also Wright (2001), 70: 'more often – and more basically – the indeterminacy will be initially manifest not in (relatively confident) verdicts of indeterminacy but in (hesitant) differences of opinion (either between subjects at a given time or within a single subject's opinions at different times) about a polar verdict, which we have no idea how to settle, and which, therefore, we do not recognize as wrong.'

[17] See Williamson (1994), §2.1; Keefe and Smith (1996), 5. Vagueness is often associated with Sorites paradoxes. If some borderline cases differ along a dimension, and if their differences are small enough not to make any difference to their correct classification, then they produce a Sorites paradox. Some philosophers restrict vagueness to terms that are Sorites-susceptible (Williamson 1994; Williams 2008; Barnes 2009). Others require only borderline cases, though they acknowledge that Sorites-susceptibility is typical (Black 1937; Keefe 2000, ch. 1; Sorensen 2012).

I shall adopt the less restrictive definition. Those who want to be more restrictive can substitute 'borderline' for 'vague' in this thesis. It makes no difference to my arguments, which do not depend on Sorites paradoxes.

They propose four main theories on the roots of vagueness. Here I shall sketch these theories, focusing on their semantic, metaphysical, and epistemological implications. In the first theory, vagueness is the result of our *semantic indecision* (Lewis 1986, 1993; Keefe 2007; Barnes 2010; Sorensen 2012, §8). The term 'bald' is vague because we have yet to fix its precise boundaries. In a borderline case of baldness, we can know everything else about the person, including the quantity of his hair and the size of his head. But we cannot know whether he is bald. The sentence 'He is bald' has no definite meaning when applied to him.

According to this theory, there is more than one way to make more exact the meaning of the term 'bald'. Once we stipulate more precise boundaries for the term, it can be used to classify borderline cases more definitely. It is tempting to say that, before stipulation, there is no fact of the matter whether a borderline case is bald. But we should avoid confusing an indefiniteness in language with one in reality. The sentence 'He is bald' can be true or false of a borderline case, depending on how we stipulate new boundaries for the term 'bald'.

The second theory, known as epistemicism, attributes vagueness to our *irremediable ignorance* (Williamson 1994; Sorensen 2005). There is nothing incomplete about the meaning of the term 'bald': it already has precise boundaries. According to Williamson (1994), our use of the term 'bald' already determines the smallest quantity of hair that it takes for someone to be bald. But the differences between this number being the boundary and its neighbours being the boundary fall below our margin for error in knowledge. These differences in meaning are too small for us to discriminate reliably. So we cannot find out where the boundary lies. For each

borderline case, there is a fact of the matter whether that person is bald or not. Yet, in principle, we cannot discover this fact.

The third theory blames *metaphysical indeterminacy* (Prinz 1998; Williams 2008; Barnes 2009). This indeterminacy may be attributed to objects, properties, relations, or states of affairs (Williamson 2005). So the fault lies at least partly in the world: it is fuzzy, or 'fundamentally unsettled', in a way that cannot be blamed on our representation of the world. In a borderline case of baldness, there is no fact of the matter whether the person is bald or not. Advocates of this theory cite, as examples, indeterminacy in personal identity, ordinary objects, quantum mechanics, and the open future.

As Williams (2008) notes, metaphysical indeterminacy is consistent with semantic indefiniteness. In this theory, the term 'bald' lacks precise boundaries, and the sentence 'He is bald' has no definite meaning in borderline cases. But this semantic indefiniteness is produced by the fuzziness of the world, rather than our failure to choose more precise boundaries for the term 'bald'. Prinz (1998) alludes to this possibility when he suggests that, in some cases, 'the indeterminacy of our predicates is inherited from the indeterminacy of the properties they denote' (§1).

The fourth theory, known as *contextualism*, appeals to context sensitivity (Soames 1999; Shapiro 2006; Åkerman 2012). It suggests that the boundaries of the term 'bald' vary contextually in a way that has been neglected. Each significant shift of context changes these boundaries. So, before any context is fixed, borderline cases abound. Which contextual factors are relevant? For instance, the boundaries of the term may depend on our judgements about borderline cases in a particular

conversation. According to one contextualist model, the sentence 'He is bald' is true of a borderline case if a competent speaker judges it to be true and his audience accepts it to be so during their conversation.[18]

Soames (2002) likens this special context sensitivity in vague terms to that in indexicals, such as 'I' and 'now'. Each term has a context-invariant meaning that constrains its boundaries across different contexts. But, without the input of relevant contextual factors, each term remains 'partially defined'. For Shapiro (2006), this boundary-determining role for contextual factors implies that vague terms are 'open-textured': a competent speaker of the language can choose between different classifications of borderline cases 'without sinning against the meaning of the words and the non-linguistic facts' (vi).

These four theories of vagueness – semantic indecision, irremediable ignorance, metaphysical indeterminacy, and context sensitivity – may not be exhaustive. Both Williams (2008) and Barnes (2009) point out that no argument has been made to show that our theories are exhaustive. Each theory has been elaborated with more logical details.[19] I will set aside these details; my interest in the theories is in their semantic, metaphysical, and epistemological implications.

---

[18] Raffman (1996) proposes a more complicated model: the relevant contextual factors are those psychological states which account for the speaker's dispositions to judge borderline cases. See the analysis of this 'psychological contextualism' in Åkerman (2012), §2.5. To simplify my discussion, I shall focus on Soames and Shapiro's model.

[19] Supervaluationist theories offer a logic and semantics that fit *semantic indecision*. They preserve most of classical logic, but sacrifice classical semantics. See Williamson (1994), ch. 5; Keefe (2000), ch. 7; and Sorensen (2012), §8. Williamson (1994), ch. 7, argues that *epistemicism* has the virtue of preserving both classical logic and semantics.

Barnes (2009), 379-380, briefly evaluates some rival logics for *metaphysical indeterminacy*. Metaphysical indeterminacy is often associated with non-classical logics; see for instance Williamson (2004), §2. However, Williams and Barnes (2011) develop a classical logical framework for it. Shapiro (2006) develops a *contextualist* model for reasoning with vague terms. He believes that a fully general account requires a para-consistent logic (17).

To spell out the implications: (a) If borderline consciousness is caused by our imprecise language, then the sentence ‘Octopuses are conscious’ has no truth value yet. This sentence can be true or false, depending on how we stipulate a definite extension for the term ‘phenomenal consciousness’. (b) If borderline consciousness is caused by our imperfect knowledge, then there is a fact of the matter whether octopuses have experience or not. But we cannot, in principle, find out this fact because we cannot discover the definite extension of the term ‘phenomenal consciousness’. (c) If borderline consciousness is caused by the fuzzy world, then there is no fact of the matter whether octopuses have experience or not. Because consciousness is metaphysically indeterminate, the term ‘phenomenal consciousness’ lacks more precise boundaries. (d) If borderline consciousness is caused by our context-sensitive language, then the sentence ‘Octopuses are conscious’ has no truth value until a context is specified. We can stipulate a definite extension for the term ‘phenomenal consciousness’. When others in the same context accede to our stipulation, we jointly fix new boundaries for the term.

Although the theories based on irremediable ignorance and metaphysical indeterminacy differ on semantics and metaphysics, they agree on three points about epistemology. First, neither of them licenses us to stipulate new boundaries for the term ‘phenomenal consciousness’. Second, neither implies any method for classifying creatures at the borderline more precisely. Third, both seem to suggest that no method is forthcoming: either we cannot ever find the precise boundaries of the term ‘phenomenal consciousness’, or none are to be found because consciousness itself is fuzzy.

All three points distinguish them from the theories based on semantic indecision and context sensitivity. The latter theories imply an epistemology for borderline cases. If these theories apply to borderline consciousness, then they license us to stipulate new boundaries for the vague term 'phenomenal consciousness' – in order to classify borderline conscious creatures more precisely. However, neither theory specifies how stipulation ought to be implemented in science. In particular, neither says more about the conditions under which stipulation can contribute to science and the bases of stipulation under those conditions.

I conclude, on these grounds, that current philosophical investigations into the nature of vagueness do not warrant any method for classifying borderline conscious creatures more precisely. As my survey shows, there is no consensus among philosophers on which is the best or correct theory of vagueness. And, unfortunately, the four main theories differ in their epistemological implications.

Moreover, I think there is reason to doubt if the four theories of vagueness are directly applicable to borderline consciousness. Philosophers who debate these theories tend to focus on familiar borderline cases that result in the Sorites paradox, such as borderline heaps or borderline bald men. Like most of us, these philosophers assume that science cannot discover exactly how many grains of sand are needed to make a heap, or how few strands of hair a person must have to be bald.[20] So they do not investigate what happens to borderline cases in science. Indeed, as Weiner (2007) says, it is 'widely assumed' in the philosophical literature on vagueness that

[20] Two exceptions are Hart (1992) on heaps and Weiner (2007) on baldness. Williamson (1994) finds Hart's case – that four grains are needed – to be 'astonishingly plausible' (213). His astonishment supports my claim that we do not expect science to classify borderline heaps more precisely.

'the methods and results of science have no place among the data to which our semantics of vague predicates must answer' (355).

Yet, as I have emphasised, we expect science to find a more precise borderline between conscious and not-conscious creatures. So we expect it to classify, more precisely, at least some borderline conscious creatures. This suggests that a theory of vagueness which applies to borderline consciousness needs to take science into account. Borderline cases that are re-classifiable by science should be kept distinct from the familiar cases that are not re-classifiable. I do not deny that some semantic, metaphysical, and epistemological resources developed in the current theories may help us interpret what happens to borderline cases in science. But, in this thesis, I will heed Weiner's advice to look afresh at how scientists manage the borderline cases of their phenomena.

2.3 McGinn's conceptual challenge

Now let me address two challenges. Both McGinn (1996) and Antony (2006b, 2008) agree that borderline consciousness is, in some sense, inconceivable. McGinn (1996) claims that our concept of consciousness 'does not permit us to conceive of genuinely borderline cases' (14). Antony (2008) contends that our concepts of conscious state and conscious creature are sharp, so they 'can have no borderline cases' (239). If they are right, then my conceptual analysis of borderline consciousness is unsound. I shall show that McGinn and Antony's challenges pose no threat to my analysis. I will explain and evaluate their arguments in detail because

they help to clarify the metaphysical and methodological commitments of my analysis.

In McGinn (1996), the challenge to borderline consciousness is based on a conceptual analysis of consciousness.[21] McGinn acknowledges two difficulties in making this analysis, which he blames on the nature of consciousness. First, consciousness is ineffable: it 'belongs to that range of properties that can be grasped only by direct acquaintance' (13). For him, this means that the concept of consciousness is only available to those who are conscious: 'just as a man born blind cannot really know what it is to be red, so a being without consciousness cannot be taught what it is to be conscious' (14). The concept cannot be explained through non-circular definitions. Second, consciousness is elusive 'even to acquaintance' (14). In particular, the relation between our consciousness and the objects of which we are conscious is 'peculiarly impalpable and diaphanous'. When we try to investigate this relation through introspection, our attention is invariably drawn to the objects of consciousness instead. This makes it difficult for an analysis to say 'what consciousness intrinsically is'.

Despite these difficulties, McGinn believes that his analysis reveals one essential aspect of consciousness.

> There is, though, something instructive that we can say about the nature of consciousness – and this is that the possession of consciousness is not a matter of *degree*. Put differently, the concept of consciousness does not permit us to conceive of genuinely borderline cases of sentience, cases in which it is inherently indeterminate whether a creature is conscious: either a

[21] He does not specify phenomenal consciousness, but refers to consciousness with 'an "inner" subjective aspect' (15). In a later chapter, he speaks interchangeably of 'consciousness', 'conscious experience', 'subjective awareness', and 'experience' (40).

creature definitely is conscious or it is definitely not. (14, all italics in original)

In this passage, McGinn equates borderline cases of consciousness with degrees of consciousness. He ties the conceptual possibility of borderline consciousness to 'the question whether, *if* the creature is conscious, this can be a matter of degree' (14). McGinn also seems to assume that what we cannot 'conceive of' with the concept of consciousness reflects something about the 'nature' of consciousness. If he can prove on such conceptual grounds that consciousness has no borderline cases, then I would have to recant my analysis of borderline consciousness.

So how does McGinn make his case? According to his analysis, a creature possesses consciousness if it is a subject. There must be 'something "inner", some way the world appears *to* the creature' (15). McGinn claims that this concept of consciousness does not allow us to conceive of how the world appears to a borderline conscious creature – or, in Nagel's terms, what it is like to be a borderline conscious creature. To back up his claim of inconceivability, he appeals to our imagination: 'we cannot imagine the position of a creature for whom it is indeterminate whether there is such an "inner" subjective aspect'.

He also draws a conceptual contrast between life and consciousness. On one hand, we can think of life emerging through gradual transitions 'from the plainly inanimate to the indisputably living' (15). Amid these transitions are borderline living things such as bacteria. On the other hand, the emergence of consciousness 'must' be compared to 'a sudden switching on of a light, narrow as the original shaft must have been'. We can conceive of the 'minds of lowly creatures' with a 'small speck of

consciousness quite definitely possessed', but they are not 'in the partial possession of something admitting of degrees'. From his analysis, McGinn concludes that consciousness has an 'all-or-nothing character' (15). This conclusion 'seems' to be something that 'any account of consciousness must respect'.

I do not think that McGinn's arguments affect my conceptual analysis of borderline consciousness. Why? First, I question how he conflates vagueness of consciousness with degrees of consciousness. The claim that consciousness has borderline cases neither implies, nor is implied by, the claim that it is possessed in degrees. Consider the view proposed in Searle (1992):

> Consciousness is an on/off switch: a system is either conscious or not. But once conscious, the system is a rheostat: there are different degrees of consciousness. (83)

One might agree with Searle that there is a sharp boundary between having consciousness and lacking it, even though creatures with consciousness have it in various degrees.[22] Or one might accept that consciousness has borderline cases, but agree with McGinn that it is not possessed in degrees. Nothing in my conceptual analysis of borderline consciousness requires that creatures have consciousness in different degrees.

McGinn's conflation of borderline cases and degrees underlies the contrast he draws between life and consciousness. According to him, we cannot make sense of borderline consciousness because we cannot conceive consciousness as 'the partial possession of something admitting of degrees'. Yet he does not explain the relation

[22] See, for instance, Unger (1988), 308, who attributes this 'dimmer-switch model' to Mark Heller.

between consciousness being indeterminate, being partially possessed (as opposed to being 'definitely possessed'), and 'admitting of degrees'. Without an account of this relation, I do not know how to interpret McGinn's conceptual contrast, and assess its implications for borderline consciousness.

Second, I note that McGinn sees vagueness as metaphysical indeterminacy. As he specifies above, 'genuinely borderline cases' are cases in which it is 'inherently indeterminate' whether a creature is conscious (14).[23] I take 'inherently' to mean that any vagueness of consciousness must lie in the creatures, rather than be the result of our semantic and epistemic relations with them. My interpretation fits McGinn's reference to the 'position of a creature for whom it is indeterminate' (15). It is also supported by the fact that he explicitly disavows the epistemicist view of vagueness: his claim against borderline consciousness is meant to be 'a claim about what it is to *be* conscious, not a claim about our *knowledge* as to whether a creature is conscious' (14). So the real target of McGinn's challenge must be the claim that consciousness is metaphysically indeterminate.

Unlike McGinn, I do not define borderline cases so narrowly. Following current practice in philosophy of vagueness, I do not rule out by fiat the possibility that some borderline cases are produced by semantic indecision, irremediable ignorance, or context sensitivity. Indeed, my analysis suggests that none of the four theories are directly applicable to borderline consciousness. At most, McGinn's arguments show that consciousness is not metaphysically indeterminate. It remains possible that

[23] See also Tye (1996), 682: 'Consciousness is vague only if there are *objective* borderline cases – cases for which there is no definite fact of the matter as to whether they are instances of consciousness. Not knowing, or being able to tell, whether certain states (or creatures) are conscious is compatible with their definitely being conscious (in the metaphysical sense).'

borderline cases of consciousness arise from semantic indecision, irremediable ignorance, or context sensitivity. McGinn offers no argument against these possibilities since he assumes that vagueness is metaphysical indeterminacy.

Third, I deny that McGinn has proven that consciousness is not metaphysically indeterminate. So I see no reason to revise my analysis to accommodate his conclusion. His challenge to metaphysical indeterminacy is based on appeals to conceivability; even the conceptual contrast between life and consciousness rests on what is intuitively not conceivable about the 'minds of lowly creatures'. Although I am sympathetic to McGinn's claims about what is inconceivable, I see no reason to trust arguments that move directly from claims about what is inconceivable of consciousness to conclusions about the nature of consciousness. Such arguments are dubious from a naturalist perspective – especially if we expect science to make interesting discoveries about the nature of consciousness. Our beliefs about what is conceivable of a phenomenon tend to change as the science of that phenomenon develops.[24]

Moreover, McGinn's trust in these arguments sits oddly with his own sceptical stance on what we can discover about the nature of consciousness.[25] Elsewhere, he denies that we can infer from the inconceivability of a physical explanation between conscious states and brain states to the conclusion that consciousness is either

[24] I take this methodological point from Block (1978): 'That something is currently inconceivable is not a good reason to think that it is impossible. Concepts could be developed tomorrow that would make what is now inconceivable conceivable' (85).

[25] This stance, in turn, contrasts with his considerable faith in what philosophers of mind can achieve from the armchair: '…that we can do interesting philosophy of mind at all shows something important about mental concepts and hence mental phenomena. What it shows is that the essence of mental phenomena is contained a priori in mental concepts' (6).

miraculous or non-physical in nature. Instead, he proposes that 'consciousness is a *mystery* in the sense that it is beyond human powers of theory construction, yet there is no sense in which it is inherently miraculous' (42). So the fault lies in our cognitive limitations, which prevent us from grasping the physical properties that 'link' consciousness to the brain (43). My *ad hominem* counter-challenge is this: if McGinn acknowledges that inconceivability is no proof that consciousness is not physical, then why does he accept, on the basis of inconceivability, that consciousness is not indeterminate?

It may be that McGinn is aware of this tension. He abruptly weakens the strength of his challenge against metaphysical indeterminacy in consciousness. Initially, he concludes that the 'all-or-nothing' aspect of consciousness 'seems to be a feature that any account of consciousness must respect' (15). By the end of the same paragraph, he provides an escape clause for theories that deny the metaphysical determinacy of consciousness:

> It is therefore in place to ask, of any theory of the mind, whether it can accommodate this feature of consciousness – and if it cannot, what view it takes of the intuition that consciousness is so constituted. (16)

I take this to be a significant concession. First, McGinn is acknowledging that his conclusion about the metaphysical determinacy of consciousness rests on an intuition. Second, he suggests that this intuition can be overridden or explained away by a theory of consciousness. This seems to me methodologically sound. So I shall follow his recommendation. When I evaluate my approach to borderline consciousness, I will check if it is compatible with our intuitions on the determinacy

of consciousness. If it is incompatible, then I will ask whether we can override or explain away these intuitions.

2.4 Antony's psychological challenge

The challenge in Antony (2006b, 2008) differs from McGinn's in two ways. First, Antony does not rely on our imagination to establish that borderline consciousness is inconceivable. Neither does he rely on our intuition to conclude that consciousness is metaphysically determinate. Rather he specifies two psychological requirements for conceiving borderline cases, and shows that these requirements are unlikely to be met by our understanding of the concepts of conscious state and conscious creature. Second, he does not assume that vagueness reflects metaphysical indeterminacy. He relates borderline cases intuitively to 'borderline regions with blurred or fuzzy boundaries' (2008: 240). And his usage is meant to be 'neutral' between competing theories about the roots of vagueness.

In his analysis, Antony focuses on the concepts of conscious state and conscious creature, rather than the concept of consciousness. The concepts of conscious state and conscious creature are related because he assumes that 'a creature is conscious if and only if it can enjoy conscious states' (2008: 240). For him, the concept of conscious state refers to the '*general* state of being conscious', rather than more determinate conscious states such as pain and the experience of seeing red (243).

According to Antony (2006b), philosophers have two conflicting intuitions about

these concepts. On one hand, many intuitively believe that the concept of conscious state is not vague. This belief draws on our ordinary intuition that 'there can be no borderline conscious states' (516). It 'extends naturally' to our concept of conscious creature, because a borderline conscious creature would have to be in a borderline conscious state.[26] On the other hand, some of us intuitively believe that the concept of conscious creature is vague. This belief draws on our ordinary judgements that various creatures – 'fish, worms, insects, etc.' – are borderline conscious (518).

How should we adjudicate between these intuitions? Antony (2008) claims that the intuition of sharpness in the concepts 'seems strongest'. But he acknowledges a methodological problem with relying on this intuition: 'the trouble with intuitions, even forceful ones, is that it is all to easy to deny having them, or to deny their trustworthiness' (243). So he has to offer arguments to support the intuition. If these arguments work, then Antony can deny that our judgements about borderline conscious creatures are intelligible: 'One can of course *say* of a creature that it has borderline conscious states, but in doing so…one does not understand what is being attributed to the creature' (259).

To explicate his arguments, I must explain how Antony distinguishes a concept **F** from its associated conception [F]. Put in psychological terms, a concept is 'a mental representation that can be a constituent of thoughts' (2008: 239). A conception is a complex mental structure 'by means of which we identify, categorize, and often simply think about objects, events, properties, etc.' (244). It represents how we conceive of something. So the conception associated with a concept is the mental

[26] Unlike me, Antony does not explicitly allow for borderline cases that are re-classifiable by science.

structure that we depend on when we use the concept to think.[27] For instance, our conception [bald] is a complex mental structure with elements such as [person], [head], and [quantity of hair]. This conception represents how we conceive of baldness. We depend on it when we use the concept **bald** to make claims and classifications about baldness. To classify an individual as bald, borderline bald, or not bald, we determine if he satisfies the conception's elements in the right way.

Let me reconstruct and assess the two arguments against borderline consciousness in Antony (2006b, 2008). The first is a 'short argument' in Antony (2006b), which 'points to' the conclusion that the concept of conscious state is sharp (518). He begins with an observation about our vague concepts:

> 'any uncontroversially vague concept like **red** or **child** determines for us one or more *specific* borderline regions to which anyone competent with the concepts is sensitive; and the way in which such concepts are mentally represented excludes our even *entertaining* the possibility of borderline cases lying well outside those regions. (518-9)

Antony does not spell out what he means by 'the way in which such concepts are mentally represented'. I take him to be referring to the complex mental structures underlying our use of the concepts – in other words, the conceptions that are associated with the concepts. From his observation, he seems to derive a negative psychological requirement, which I call the *limited possibilities condition*. The condition requires that, for any vague concept of F which we understand, our associated conception [F] prevents us from imagining borderline Fs 'well outside' the

[27] Antony remains 'neutral' on the metaphysical relation between the concept and its associated conception (245-6). It is makes no difference to his argument whether the concept is identical to, individuated by, or only psychologically related to the conception.

borderline areas known to anyone who understands the concept.[28]

Antony's examples are the concepts of red and child. For those of us who understand the concept of red, we are familiar with the colour regions in which borderline reds appear. It is part of our understanding of this concept that our conception [red] prevents us from imagining that black might be a borderline red. Similarly, if we understand the concept of child, then we are familiar with the ages of borderline children. And our conception [child] is such that we cannot imagine a 70-year-old to be a borderline child.

He contrasts these examples with the concept of conscious creature. Although we may associate the concept of conscious creature with some borderline areas, our conception [conscious creature] does not prevent us from imagining borderline conscious creatures well outside these areas. Indeed, we seem able to entertain unlimited possibilities: 'one can imagine with Descartes that only humans are conscious, or that the boundary lies closer to fish, or that plants or even everything is conscious (panpsychism)' (518).

Why is this contrast significant? Antony draws out two implications. First, because our concept of conscious creature does not fulfil the limited possibilities condition, it 'appears not to be genuinely vague' (519). This suggests that the concept of conscious state also cannot be vague. If the concept of conscious state were vague, the concept of conscious creature would be too. For Antony, a borderline conscious creature just is a creature in a borderline conscious state. Second, our judgements

[28] I use the term 'imagining' here to refer to whatever Antony means by 'entertaining the possibility'. In this short argument, he uses the two terms interchangeably.

about borderline conscious creatures are conceptually suspect.

> Judgments that such cases are borderline…are not plausibly determined by one's concept **conscious creature** – at least not in the way in which a judgment that some color is borderline red is determined by the psychological nature of one's concept **red**. (518)

Antony suggests that our speculations about borderline conscious fishes and insects are, instead, explained by 'a prior commitment to a materialist theory of consciousness' (535). Because neural and functional concepts are usually vague, materialist theories entail the existence of borderline consciousness.

I am not convinced by either implication. My objection is to the limited possibilities condition itself: Antony never explains why this condition must hold true of all vague concepts that we understand. Why is a concept 'genuinely vague' only if its associated conception prevents us from 'entertaining the possibility' of borderline cases well outside the recognised borderline areas? When Antony compares the concepts of red and child with the concepts of conscious creature and conscious state, he does highlight a putative difference in what is conceivable using these concepts.[29] But he has yet to prove that this difference matters to the vagueness of our concepts. This missing step is crucial. Its absence may explain why he only claims that the argument 'points to' his conclusion.

Similarly, I do not see why our judgements about borderline Fs are 'plausibly determined' by our concept of F only if the limited possibilities condition holds true. In the sentence cited above, Antony shifts uneasily between the credible claim that

[29] I add 'putative' for this reason: if we think of the concepts of red and child as natural-kind concepts, does that enlarge the possibilities which we can entertain?

our judgements about borderline conscious creatures are not determined by our concept of conscious state 'in the same way' – because the concept does not rule out borderline conscious plants and rocks – and the controversial conclusion that these judgements are not 'plausibly determined' by the concept at all.

Of course, he is free to define a 'genuine' sense of vagueness such that the limited possibilities condition must hold true of all concepts with borderline cases. But we are just as free to introduce another sense of vagueness – call it *wagueness* – so that the limited possibilities condition need not hold true of all concepts with borderline cases.[30] Then it would be the wague concept of conscious creature that constitutes our judgements about borderline conscious creatures such as octopuses, fishes, frogs, late-stage human foetuses, and vegetative-state patients. Antony cannot impugn these judgements as unintelligible, unless he begs the question against wagueness.

My conceptual analysis in §2.2 adds a point in favour of wagueness. It highlights our expectation that science will re-classify some creatures now described as borderline conscious. I think this expectation may explain why the wague concept of conscious creature does not fulfil the limited possibilities condition. If we expect science to revise our boundary between conscious and not-conscious creatures, then it is not surprising that we can at least entertain the possibility of borderline conscious humans, fishes, plants, and rocks.

In Antony (2008), there is a second, more elaborate, argument for the claim that the

[30] Cf. Kripke (1980), 108, on schmidentity: 'This sort of device can be used for a number of philosophical problems.'

concepts of conscious state and conscious creature are not vague.[31] Here he proposes a positive psychological requirement for the conceiving of borderline cases. I call it the *common properties condition*: it requires that, for any vague concept **F** which we understand, our mental representations of Fs, borderline Fs, and not-Fs have elements in common drawn from the conception [F] (251). Each element refers to a property that we represent as 'clearly belonging' to Fs, borderline Fs, and not-Fs. Antony adds two clarifications. First, we need not always have representations of Fs, borderline Fs, and not-Fs that fulfil this condition when we are conceiving borderline Fs. But it should be '(normally) always psychologically possible to do so' (250). Second, we need not be aware that our representations fulfil this condition (251).

Consider again the concept **bald**. Our mental representations of people who are bald, borderline bald, and not bald are generated by the conception [bald]. These representations have the elements [person], [head], and [quantity of hair] in common, which are drawn from the conception [bald]. Correspondingly, we attribute the properties of being a person, having a head, and bearing a certain quantity of hair jointly to people who are bald, borderline bald, and not-bald. If we imagine a series of bald, borderline bald, and not-bald individuals, we think of their quantity of hair increasing gradually from individual to individual.

Does the common properties condition hold true of our concept **conscious state**? It requires that some elements from the conception [conscious state] appear jointly in our mental representations of conscious states, borderline conscious states, and not-conscious states. Here Antony constructs a dilemma. Either the elements from the

[31] I simplify it in two places, which are highlighted in the next two notes. These simplifications do not affect my evaluation of his argument.

conception [conscious state] refer to properties which conceptually imply consciousness, or they refer to properties which do not.[32] [Subjectivity] and [qualia] belong to the first horn of the dilemma. He denies that citing these elements can help to fulfil the common properties condition. Why? We cannot coherently attribute the properties of being subjective and having qualia to states that we are conceiving at the same time as borderline conscious or not-conscious.

For the second horn of the dilemma, the most plausible candidates refer to neural and functional properties. But Antony rejects such elements as 'illegitimate'.

> There is an obvious difficulty with this suggestion.…I wish to argue for the sharpness of a concept that is *pretheoretical* or *neutral* with respect to materialism, dualism and idealism. That being the case, one clearly cannot assume that representations as of such physical(/functional) properties are part of the conception associated with *that* concept, since such representations are inconsistent with the concept's neutrality. So the suggestion is illegitimate. (253)

Since elements of the form [neurophysiological property N] and [functional property F] refer to neurophysiological and functional properties, we would have to attribute these properties jointly to conscious, borderline conscious, and not-conscious states. And, according to Antony, we cannot do so without committing to materialism about consciousness.

Why is this metaphysical commitment problematic? Earlier Antony appeals to the 'traditional mind-body problem': for Descartes and others who ponder this problem, consciousness is conceived 'in such a way that materialism, dualism, and idealism are all entertainable (albeit not with equal ease)' (246). So the concepts used to pose

[32] He also mentions a third horn, in which the element is a borderline case of consciousness-entailing (252). But this is quickly dismissed as untenable.

this traditional problem must be 'neutral' with respect to materialism, dualism, and idealism. Antony assumes that our current concepts of conscious state and conscious creature are similarly neutral, for 'it is consciousness so conceived that is in play in debates about the mind-body problem' (253). Even materialists should focus on these concepts, for 'it is consciousness so conceived' that they 'aim to explain by appeal to complex physical properties'.

Antony's overall strategy is to rule out the 'natural' candidates for each horn of the dilemma (253). Aside from elements that refer to properties which conceptually imply consciousness, and elements that refer to neurophysiological and functional properties, he also considers [temporal extension] and [intensity]. These are much less plausible candidates for securing us a vague concept of conscious state: after all, even a low-intensity surge of consciousness is enough to make a creature conscious during that brief period.[33] He acknowledges that this strategy is not conclusive. It is open to others to cite elements for the second horn that he has yet to rule out. So, in effect, he is raising a challenge to those who believe that borderline consciousness is conceivable. He concludes that, 'at least until shown otherwise', the concept of conscious state has no possible borderline cases (257).

I have no objection to the common properties condition, but I do not think that Antony's arguments work. My main complaints are with his assessment of the second horn of his dilemma. This is not surprising. According to my analysis in §2.2, we classify creatures as conscious, borderline conscious, and not-conscious according to their behavioural and neurophysiological similarities to conscious

[33] See Antony (2008), §4.2 and §4.3. For brevity's sake, I leave out his criticisms of these candidates.

humans. To do this, we must attribute at least some behavioural and neurophysiological properties jointly to conscious, borderline conscious, and not-conscious creatures. So I cite elements which refer to behavioural and neurophysiological properties to fulfil the common properties condition. When I criticise Antony's argument, I am also defending this conceptual analysis.

So what is wrong with Antony's arguments? First, I deny that citing elements which refer to neurophysiological and functional properties commits us to materialism about consciousness. In fact, neither dualism nor idealism is excluded. Suppose that neurophysiological properties play the role of common properties. This is compatible with the metaphysics of property dualism and phenomenalism. Property dualism posits both physical and phenomenal fundamental properties in conscious states (Chalmers 2003, §9; Robinson 2012, §2.2). It is true that, according to this form of dualism, neurophysiological properties may turn out to be *contingent* properties of conscious states. But the common properties condition does not require that the relevant properties be *necessary* properties of conscious states, borderline conscious states, and not-conscious states – only that they be properties that we represent as 'clearly belonging' to these states. Phenomenalism is the form of idealism in which phenomenal properties constitute physical properties (Chalmers 2003, §11; see also Robinson 2012, §1.1). It admits neurophysiological properties but construes them as fundamentally phenomenal in nature.

Next, suppose that functional properties play the role of common properties. I do not know why Antony assumes, without argument, that the attribution of such properties to conscious states, borderline conscious states, and not-conscious states

is inconsistent with dualism and idealism. As Putnam (1967) emphasises, functional properties are role properties that can be realised by either physical or non-physical properties.[34] It is true that, in most functionalist theories of the mind, the functional properties that individuate mental states are specified partly in terms of physical inputs and outputs (Block 1978). But such theories need not rule out non-physical properties from realising the functional properties. As above, I see no trouble for phenomenalism. It allows for functional properties, as well as physical inputs and outputs, so long as they are fundamentally phenomenal in nature.

Second, even if citing elements which refer to neurophysiological and functional properties commits us to materialism, I do not find this metaphysical commitment as problematic as Antony does. Unlike him, I am ready to concede that debates about borderline consciousness might be intelligible among materialists only. Say that Antony is right on this historical point: when Descartes and others pose the traditional mind-body problem, consciousness is conceived to be 'neutral' between materialism, dualism, and idealism. I see no reason why materialists cannot theorise about 'consciousness so conceived' and thereby modify their associated conception to include representations of neurophysiological and functional properties. There is a vast community of philosophers and scientists for whom dualism and idealism are not plausible metaphysical options.[35] Why deny them the right to theorise with the

[34] Strictly speaking, Putnam (1967) puts the point in terms of functional states: '…the functional-state hypothesis is *not* incompatible with dualism! Although it goes without saying that the hypothesis is "mechanistic" in its inspirations, it is a slightly remarkable fact that a system consisting of a body and a "soul", if such things there be, can perfectly well be a Probabilistic Automaton' (436). He defines a Probabilistic Automaton such that its Total States and inputs are 'neither mental nor physical *per se*.'

[35] Chalmers (2003) attests to the existence of this community: 'It is often held that even though it is hard to see how materialism could be true, materialism *must* be true, since the alternatives are unacceptable' (268). Of course, his arguments aim to question the metaphysical presumption of this community.

concept of consciousness?

To be fair, Antony (2008) addresses this criticism later:

> …the sharpness of our pretheoretical concept **consciousness** now threatens any c-materialist view that employs a vague concept of consciousness (or what it calls 'consciousness'). Of course, it must be granted that c-materialism may be true, in which case a vague **consciousness** will eventually supplant our current sharp concept. But c-materialism has hardly begun to provide the kind of explanation of consciousness that biology provided for life. (261)

'C-materialism' refers to materialist theories that invoke complex neurophysiological or functional properties to explain consciousness. In this passage, Antony acknowledges that consciousness science may revise how we conceive of consciousness, such that borderline consciousness becomes conceivable. If our conception of consciousness changes in this way, a vague concept of consciousness will 'supplant' our current sharp concept.[36] He implies that the sharp concept of consciousness has not been supplanted already.

I think it crucial to assess why Antony is so confident on this last point. As the passage indicates, he draws an analogy between consciousness science and modern biology. Modern biology changed how we conceive of life, such that the sharp pre-modern concept of life was supplanted by a vague one. Antony makes two

[36] It is not clear if Antony interprets this kind of conceptual change as conceptual *replacement* or conceptual *development*. He distinguishes the 'crude, pretheoretical concept' from the '*correct* concept (or at least a more correct concept)' (253). But he also connects our '*current* concept' with 'more *correct versions*' of it (241), and our 'modern' concept with 'earlier versions' of it (253). I think he may well be agnostic between these interpretations. (This would fit his 'neutrality' on the metaphysics of concepts; see note 25 above.)

Earlier, in Antony (2006b), he explicitly allows for these two possibilities. The first possibility is 'not of conceptual development but rather of conceptual *switching*, a change of subject' (532). In the second possibility, 'our current sharp concept could develop into a vague concept while retaining its identity'. He adds, in a note, that 'we know almost nothing' about how we decide real cases (537).

epistemological claims about this history. First, to secure the conceptual change, pioneering biologists must bear the 'burden of proof' and develop 'theories of such scope and power that earlier views of life could no longer be sustained' (261). Second, before these theories were developed, the sharp concept itself offered 'good reason' to believe that non-vitalist theories which entailed borderline cases of life are false (261).

From this history, Antony deduces an epistemological requirement. Call it the *burden of proof condition.* It requires that, for a change from a sharp concept of F to a vague concept of F, scientists develop new theories with the vague concept that render the theories with the sharp concept unsustainable. He puts this burden of proof on today's materialists. They must develop an explanation of consciousness 'sufficiently persuasive to justify our rejecting central features of our current concept' (261). Until then, he concludes that the sharp concept of consciousness gives reason to believe that 'c-materialist theories that employ vague concepts of what they call "consciousness" are either false or not about consciousness at all' (262).

I doubt that Antony is right to set the burden of proof so high. His argument leads to an implausible picture of conceptual change in science. On one hand, suppose that the sharp concept of life gave pioneering biologists 'good reason' to believe that their non-vitalist theories of life are *false about life* because they entail that life has borderline cases. Why should these biologists have continued to develop non-vitalist theories of life until they became 'sufficiently persuasive' and 'so impressive' in scope and power that vitalist theories of life 'could no longer be sustained'? That picture requires the agents of conceptual change to be implausibly motivated.

On the other hand, suppose that the sharp concept gave these biologists good reason to believe that their non-vitalist theories are *not about life at all*. Should the biologists have changed their minds about the subject of their theories only when they became 'sufficiently persuasive' and 'so impressive' in scope and power that vitalist theories of life 'could no longer be sustained'? That picture makes conceptual change implausibly abrupt. It also implies that, until vitalist theories became unsustainable, these biologists could not intelligibly raise and discuss questions about borderline cases of life.

If Antony admits less abrupt conceptual changes to his picture, then he needs to explain how vitalist and non-vitalist biologists once had reason to believe that they were proposing viable theories about the same subject, even though they disagreed on the possibility of borderline cases. And he needs to explain why materialists today who allow for borderline consciousness lack the same reason to believe that they are using a vague concept to refer to consciousness, make intelligible judgements about borderline consciousness, and propose viable theories about consciousness. It may be true that materialist theories of consciousness are not 'sufficiently persuasive' today to make non-materialist theories unsustainable. But many materialists take themselves to have good reasons for committing to materialism *per se* (Papineau 2002, ch. 1; Stoljar 2015, §17). I think this metaphysical support already earns them the right to theorise with a vague concept of consciousness. Nothing in Antony's analogy proves otherwise.

For now, Antony's methodological approach does not strikes me as feasible. He takes the *a priori* psychological study of concepts more seriously than I countenance.

Consider, for example, how he deploys the common properties condition in debate. He admits that some elements of our conceptions are '*unconscious* or *implicit* in that we have no introspective access to them' (246). So his arguments can only draw on conscious elements:

> The elements of [consciousness] on which my arguments are based, accordingly, are only those to which we have conscious access. It will thus remain open that there are unconscious elements that entail that **consciousness** is vague. In the absence of evidence for such elements, however, it is rational to suppose that they do not exist. (246)

From Antony's perspective, after he has shown that conscious elements cannot help fulfil the common properties condition, the burden of proof is on those who believe that borderline consciousness is conceivable. They must provide evidence that the common properties condition can be fulfilled. Unless they can do so, he concludes that the concepts of conscious state and conscious creature are not vague. Our ordinary judgements about borderline conscious creatures must, therefore, be unintelligible. I note that, for Antony, these judgements do not constitute evidence for the existence of elements that help to fulfil the condition.

My inclination is to argue in the opposite direction. We know that most of the terms in our language, outside of mathematics, are vague. From my perspective, because we commonly judge that some creatures are borderline conscious, our analysis of the concepts of conscious state and conscious creature should not render these judgements unintelligible. If Antony's psychological requirements for understanding a vague concept implies that our judgements about borderline conscious creatures are unintelligible, then I am ready to infer a flaw either in his requirements or their

application to the concepts of conscious state and conscious creature. Why? We know so little about the psychology of vague concepts, including how they work in relation to their associated conceptions. In this circumstance, we should trust our ordinary judgements more than any rudimentary theory about vague concepts that attempts to overrule these judgements.

2.5 Conclusion

In this chapter, I analysed the concept of borderline consciousness in terms of phenomenal consciousness and borderline cases. My analysis explained how we use behavioural and neural criteria to apply this concept, and how these criteria are apt to produce borderline cases. It also emphasised our expectation that science will classify at least some borderline conscious creatures more definitely. This expectation underlies two points of my analysis: the provisional nature of our behavioural and neural criteria, and the need to investigate how scientists manage borderline cases.

Then I defended the concept of borderline consciousness against McGinn and Antony's challenges. My defence depended on the resources in my conceptual analysis. At the same time, it clarified the metaphysical and methodological commitments of my analysis. Unlike McGinn, I distinguished borderline cases of consciousness from degrees of consciousness. I left open the possibility that borderline consciousness arises from semantic indecision, irremediable ignorance, metaphysical indeterminacy, or context sensitivity. Unlike Antony, I denied that

accepting a vague concept of conscious state commits us to materialism. I also denied that materialists must accept that their current theories of consciousness are either false or not about consciousness. However, I emphasised that nothing in my analysis prevents us from conceding that debates on borderline consciousness are intelligible among materialists only.

Chapter 3

Borderline consciousness: two epistemological challenges

3.1 Introduction

In Chapter 2, I analysed the concept of borderline consciousness in terms of phenomenal consciousness and borderline cases. According to this analysis, our operational criteria for attributing phenomenal consciousness to others draw on their behavioural and neural similarities to conscious humans. These criteria are apt to produce borderline cases. However, we expect science to improve them, and so classify the borderline cases more precisely. I also clarified my analysis by considering two challenges from McGinn (1996) and Antony (2006b, 2008), which deny that borderline consciousness is conceivable.

Now let me distinguish two epistemological challenges that borderline consciousness poses to consciousness science. The first is more obvious. It is the *problem of classification*: how can scientists classify borderline conscious creatures more precisely? This problem arises because we expect science to discover a more precise borderline between conscious and non-conscious creatures. As I will show, scientists can address the challenge by investigating phenomenal consciousness as a natural kind. But those who do so face a second challenge, also produced by borderline consciousness. This is the *problem of multiple kinds*. When these scientists discover more than one overlapping kind in their samples of conscious creatures, how can they identify the kind to which all and only conscious creatures belong?

I shall examine both epistemological challenges, though my ultimate aim in this chapter is to clarify the nature and significance of the multiple kinds problem in consciousness science. First, I explain how scientists can solve the classification problem by building neural theories of phenomenal consciousness. Then I interpret this strategy according to some epistemological norms associated with inference to the best explanation (Block & Stalnaker 1999). I also interpret it, in more metaphysical terms, to be the investigation of phenomenal consciousness as a natural kind (Block 2007a, b; Shea & Bayne 2010). Second, through a simplified model, I explain why borderline consciousness leads scientists to discover more than one overlapping kind in their samples of conscious creatures. I clarify this problem by contrasting it with the phenomenon of multiple realisability in mental kinds. Third, I demonstrate how this problem produces three theoretical impasses on the neural structure of phenomenal consciousness (Irvine 2013), the development of foetal consciousness (Derbyshire & Raja 2011; Chin 2011), and the possibility of artificial consciousness (Prinz 2003, 2005).

3.2 The problem of classification

How can scientists classify borderline conscious creatures more precisely with respect to phenomenal consciousness? First, I shall distinguish two scientific strategies that appear in the philosophical literature.[1] They depend, respectively, on

[1] I learnt most from Allen and Trestman (2015), Dennett (1995, 1996), and Shea and Bayne (2010). The relevant literature has two sources. The first is debates about specific borderline conscious creatures. They usually appear in bioethics, philosophy of science, and science journals; I will be citing a representative sample. The second source is philosophical debates about animal consciousness and

behavioural analogies and neural theories. Second, I will interpret the second, more promising, strategy in both epistemological and metaphysical terms. Third, I will assess two doubts about this strategy. This analysis sets up the multiple kinds problem in the next section.

The first scientific strategy looks for *behavioural analogies* between borderline conscious creatures and conscious humans. Recent experiments on borderline conscious animals test for complex behavioural patterns that indicate 'motivational trade-offs' between preferences. Scientists find surprising behavioural similarities between some animals and conscious humans. Fishes are less likely to avoid electrical shocks if they can stay near a member of the same species (Dunlop *et al.* 2006; Braithwaite & Boulcott 2007; Braithwaite 2010). Hermit crabs sometimes leave their shells to avoid electrical shocks, but they are more likely to leave less-preferred shells (Elwood & Appel 2009; Appel & Elwood 2009; Gherardi 2009; Elwood 2011).

A few scientists cite these behavioural patterns as evidence that the fishes and crabs can experience pain. But others contest this inference: they interpret the behaviour to be merely the result of complex reflexes (Rose 2002; Chandroo *et al.* 2004; Sneddon 2009; Mason 2011; Rose *et al.* 2014; Brown 2014). I think this reliance on complex behavioural similarities faces two limitations. First, laboratory experiments usually test for codifiable and replicable behaviour, yet this is also the kind of behaviour most plausibly interpreted as reflexive and non-conscious. Second, as Allen and Trestman (2015) notes, critics can always 'exploit some *dis*analogy

animal minds, which sometimes address epistemological issues. See the surveys in Allen and Trestman (2014) and Lurz (2009), §2.

between animals and humans' to counter the appeal of any similarities (§1).

Other experiments test, instead, for intentional action. In normal humans, such behaviour indicates phenomenal consciousness (Block 1995b: 234). Farah (2008) reports on a 1996 review of vegetative-state patients in a hospital. Almost half of them were misdiagnosed: 'they were simply the result of insufficient sampling of patients' behavior' (12). More careful testing revealed behaviour that is consistent with what is diagnostically known as the 'minimally conscious state'. The misdiagnosed patients were able to follow simple commands and to interact intentionally with the environment.

In a set of intriguing experiments, vegetative-state patients are instructed to perform various imagining tasks while their brain activity is recorded (Laureys & Boly 2007, 2009). When their brain activity matches the activity normally associated with a task, it is interpreted by some scientists as intentional action that expresses the patients' understanding and following of an instruction. Their interpretation remains controversial: such brain activity is, at best, a limiting case of intentional action.[2] But the results may help to re-classify some vegetative-state patients. Here I note two in-principle limitations to this approach. First, it cannot work for creatures incapable of comprehending human instructions. Second, as Shea and Bayne (2010) point out, it cannot work for those 'incapable of volition at all' (11).

The second, more comprehensive, strategy is to build *neural theories* of phenomenal

[2] Block (2007a) argues that the patient's 'act of imagining' should be seen as 'no less an indication – though of course a fallible indication – of consciousness than an external behavioral act' (484). There remains doubt over whether this brain activity counts as genuinely intentional action. For more critical discussion of the experiments, see Bayne (2009a), 486-7, and Shea and Bayne (2010).

consciousness.[3] Allen and Trestman (2015) raise three questions that scientists should address in pursuing this strategy.

> One strategy for bringing consciousness into the scientific fold is to try to articulate a theoretical basis for connecting the observable characteristics of animals (behavioral or neurological) to consciousness. What effects should consciousness have on behavior? What capacities and dispositions should we expect a conscious creature to have that might be absent in a nonconscious creature? What neurophysiological structures and processes might realize the dynamics or information processing required for consciousness? (§5)

Their questions point to three related levels of theorising in consciousness science: the behaviour associated with phenomenal consciousness, the psychological capacities and dispositions responsible for that behaviour, and the neurophysiological structures and processes that realise information processing in those psychological capacities and dispositions. By building a theory across these levels, scientists can look for neural structures and processes that explain behaviour which we ordinarily associate with phenomenal consciousness. The aim is not to produce yet another analogy between borderline conscious creatures and conscious humans, based on neurophysiology rather than behaviour. Rather it is to discover the neural structure of phenomenal consciousness, and thereby re-classify borderline conscious creatures that possess this structure.[4]

Allen and Trestman suggest that this strategy depends on 'inference to the best explanation' (§4.3). But they do not explain its role. So here is my interpretation,

[3] See Farah (2002, 2005) for some early statements of the general strategy. In Farah (2008), she proposes a 'neuroscience approach' to the mental states of non-human animals with 'limited communicative abilities' (14). Bayne (2009b), 491, outlines a similar approach.

[4] Two caveats: (a) For simplicity, I will use 'neural structure' broadly to mean neural structures and processes, as well as neuro-functional structures and processes. (b) I do not thereby rule out the possibility that the structure of phenomenal consciousness is partly representational. Even the most ambitious representationalist defines qualia as a specific kind of representation 'where the kind can be specified in functionalist or other familiar material terms' (Lycan 2015, §2).

based on the epistemology of theoretical identity. Theoretical identities in science are justified by inference to the best explanation (Block & Stalnaker 1999; Block 2002b; McLaughlin 2003; Prinz 2003).[5] Suppose that scientists find a causal or constitutive relation between a neural structure in conscious creatures and their characteristic behaviour. This may be due to accidental correlations. It *just* happens that the neural structure is systematically correlated with phenomenal consciousness. And it *just* happens that the behaviour causally or constitutively related to this neural structure is also the behaviour operationally associated with phenomenal consciousness. Alternatively, there may be a common cause that brings about both the neural structure and phenomenal consciousness.

Unless scientists have evidence of a common cause, the best explanation is that this neural structure is the neural structure of phenomenal consciousness, in the sense that H_2O is the chemical structure of water. More simply, scientists might say: phenomenal consciousness *is* this neural structure, in the sense that water *is* H_2O. Once the identity between phenomenal consciousness and a neural structure is established, scientists can classify borderline conscious creatures more precisely. Only creatures with this structure count as subjects of experience.

Advancing a theoretical identity brings three explanatory benefits. First, the identity explains why the relevant neural structure is systematically correlated with phenomenal consciousness. Second, it explains why, in conscious creatures, the neural structure is causally or constitutively related to behaviour that we ordinarily

[5] On the general role of inference to the best explanation in science, see Boyd (1980); Godfrey-Smith (2003), 43-4; and Douven (2011), §1.2. Lipton (2004), §4, describes how inference to the best explanation works, though he does not mention theoretical identities.

associate with phenomenal consciousness. Third, the identity excludes – as empirically meaningless – questions about how the neural structure causes, produces, or gives rise to phenomenal consciousness. Likewise, the identity between water and H_2O excludes questions about how H_2O causes, produces, or gives rise to water. In this sense, identities are not open to further explanation.[6] Scientists accept that water is H_2O not because they have discovered a causal or mechanistic relation between water and H_2O, but because the identity between water and H_2O contributes to a better overall explanation of water, its behaviour, and its correlates.

Let me clarify the epistemological norms at work here. In this use of inference to the best explanation, the theoretical identity is a better explanation than its rivals because it is simpler in several dimensions, including parsimony.[7] Rather than construe the relevant neural structure and phenomenal consciousness as two distinct entities, scientists need only commit to one entity. They also avoid committing to accidental correlations, undiscovered common causes, or undiscovered causal or mechanistic relations between the neural structure and phenomenal consciousness. Such norms of simplicity seem to be decisive whenever inference to the best explanation is used to justify a theoretical identity. Block and Stalnaker (1999) note that their 'power and importance' are 'widely acknowledged even if no one has ever been able to formulate them precisely' (23).

Next I want to interpret this strategy in more metaphysical terms. Following Block

[6] The same point is made in Block and Stalnaker (1999), §6; Block (2002b), 411-2; Papineau (1993), §4.6; Papineau (2002), §5.2; and Prinz (2003), 116. An early statement is in Putnam (1967): see section V, 439-40, on the 'Methodological Considerations' that support identity claims.

[7] Simplicity is often said to be a theoretical virtue: 'other things being equal, simpler theories are better' (Baker 2013). Parsimony refers to ontological simplicity: it assesses the 'number of kinds of entities postulated by the theory'.

(2007a, b) and Shea and Bayne (2010), I interpret it to be the investigation of phenomenal consciousness as a natural kind. How is searching for a phenomenon's structure related to studying the phenomenon as a natural kind? This relation is mapped out in Figure 3.1 below. A phenomenon's structure is the set of structural properties that are causally or constitutively responsible for the superficial properties, or most of the superficial properties, associated with the phenomenon. In their model of natural kinds such as gold, water, and tigers, both Kripke (1980) and Putnam (1975a) interpret a phenomenon's structure to be its 'nature'.[8] By uncovering the hidden structure of a phenomenon, scientists delineate boundaries of a natural kind – the kind formed by all instances of the phenomenon. This structure constitutes what Kripke calls the 'essence' of the kind. It defines the kind more precisely than superficial properties because it is necessarily shared by all members of the kind.

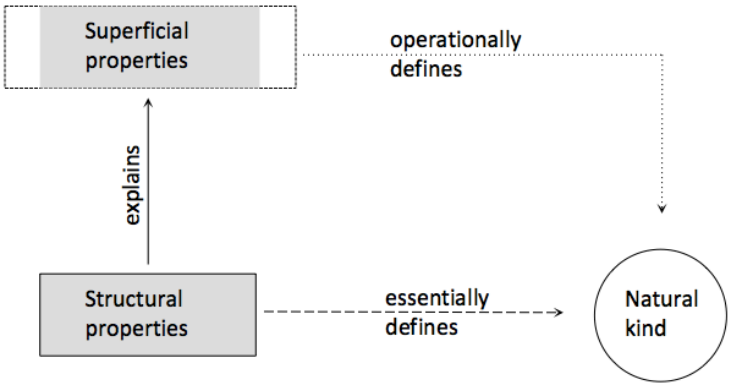


Figure 3.1: The natural kind strategy

[8] See Kripke (1980), 124-7 and 138; Putnam (1970), 140-1; and Putnam (1975), 233. I discuss their model and its possible modifications in Chapter 5.

Therefore, if scientists discover the neural structure of phenomenal consciousness, they will delineate boundaries of the natural kind formed by all conscious creatures. This neural structure is necessarily shared by all creatures that belong to the kind. It is also related, causally or constitutively, to the superficial properties that we use to operationally define consciousness. The joint discovery of neural structure and natural kind supports two scientific practices. First, by referring to the neural structure, scientists can revise our operational criteria for attributing consciousness. They can thereby trace a more precise borderline between conscious and non-conscious creatures, and re-classify at least some borderline conscious creatures. Second, by referring to the kind, scientists can make better inductive generalisations and causal explanations about conscious creatures.

This strategy belongs within what Shea and Bayne (2010) call the 'natural kind methodology'.[9] They compare a *future* theory of phenomenal consciousness to the current physico-chemical theory of water. So they recommend that scientists look for a natural kind in order to 'make better inferences about consciousness and its absence in vegetative state' (476). They also present this as a methodology 'for the science of consciousness more generally' since it promises a theory of phenomenal consciousness (459). Correspondingly, Allen and Trestman (2015) compare the *current* science of consciousness with the early science of chemical substances. We provisionally identify conscious creatures by their behaviour, just as we once

[9] Shea and Bayne explicitly adopt a broadened conception of natural kinds: 'any property that supports induction as a result of nomological principles or natural laws counts as a natural kind' (471).
They describe the general methodology as follows: 'Finding an apparent cluster of properties does not guarantee that there will be a natural property which explains the clustering (a natural kind property), but when the clustering is best explained by a natural kind property, we thereby have the means to go beyond our pre-theoretical ways of characterising the phenomenon through picking out the natural kind in new ways' (470).

identified samples of gold 'by contingent characteristics rather than its atomic essence' (§4.6). In both cases, 'putative samples must be identified by rough rules of thumb (or working definitions) rather than complete theories'.

Finally, I address two doubts about this natural kind strategy. First, what ensures that phenomenal consciousness is a natural kind? Nothing does. The risk that a putative kind is not a natural kind is built into the epistemology of natural kinds. Both Kripke and Putnam emphasise this epistemic possibility: for any putative species that scientists investigate, there may be no shared properties explaining the behaviour which we associate with the putative species.[10] So it is entirely possible that, when we group together creatures via their behaviour, we mistake members of distinct species as members of a natural kind. It is just as possible that, when we group together creatures via their behavioural and neural similarities to conscious humans, we are mistaken in treating all conscious creatures as a natural kind.

That said, this strategy can succeed even if all possible conscious creatures do not form a natural kind defined by a neural structure. It can lead to a viable neural theory of phenomenal consciousness so long as conscious humans and other mammals form a natural kind defined by a neural structure. Like Block (2007b), scientists who build neural theories of phenomenal consciousness need not assume that there is a hidden structure shared by all possible conscious creatures – 'including mammals, birds, octopi, conscious machines, and conscious extra-terrestrials' (534). They need only assume that 'there is a neural signature of

[10] See, for instance, Kripke (1980), 121: 'Since we have found out that tigers do indeed, as we suspected form a single kind, then something not of this kind is not a tiger. Of course, we may be mistaken in supposing that there is such a kind. In advance, we suppose that they probably do form a kind.'

consciousness in humans that is shared at least by other mammals with similar sensory systems'. Block adds that this assumption is 'shared by the field and looks promising so far' (534).

As Shea and Bayne (2010) demonstrate, no *a priori* argument proves that phenomenal consciousness is not a natural kind in this sense. It is true that the ordinary notion of consciousness picks out 'a number of different phenomena', which are unlikely to form a natural kind (476). They include phenomenal consciousness, access consciousness, and self-consciousness. But this does not affect the possibility that phenomenal consciousness as a determinable and its determinates are natural kinds. When scientists build theories of phenomenal consciousness, they usually specify that they are theorising about consciousness in the subjective or 'what it is like' sense.

Second, what ensures that the discovery of a natural kind will help to classify borderline conscious creatures more precisely? Shea and Bayne (2010) explain:

> If consciousness were a natural kind, it would be very surprising to discover that our access to it was limited to the pre-theoretical measures that we associate with it. In every other case, discovering a natural kind property allows us to go beyond our pre-theoretic measures. We see no good reason why consciousness cannot be investigated in the same way. (470)

They cite the discovery of the hepatitis C virus as the 'underlying pathology' responsible for the disease, which was initially identified through its symptoms. As they emphasise, doctors were already able to diagnose and treat the disease before this discovery. Now doctors can test for the presence of the virus to make a definitive diagnosis. If phenomenal consciousness is a natural kind, then scientists

can similarly improve our operational criteria for detecting it, and thereby improve our 'access' to it. This is why Shea and Bayne expect that investigating phenomenal consciousness as a natural kind can help to classify some vegetative-state patients more precisely.

Historically, the discovery of a phenomenon's structure has enabled scientists to refine their operational criteria for identifying the phenomenon. This, in turn, has enabled them to distinguish fakes. As Kripke (1980) reminds us, discovering gold's atomic structure helps to distinguish fools' gold from real gold, though both share many superficial properties (119). Allen and Trestman (2015) make the same point about gold: after scientific investigation, some putative samples 'turned out to be gold and some not' (§4.8).

But I do not find it obvious that scientists' success with false cases entails their success with borderline cases. Vague classifications are not misclassifications – no matter which philosophical theory one adopts about the nature of vagueness. So it is one thing for scientists to refine our operational criteria enough to identify creatures falsely classified as conscious and those falsely classified as not conscious. It is quite another thing to refine our criteria enough to classify borderline conscious creatures more precisely.[11] Thus far, the familiar philosophical illustrations about water and gold only show that the discovery of a natural kind helps with misclassifications. They leave open the possibility that borderline cases of consciousness pose a challenge to consciousness science that false cases do not.

[11] Shea and Bayne add that, if consciousness 'comes in degrees', then borderline cases 'would be unavoidably built into the basic ontology of the phenomenon' (477). But they do not clarify this link between degrees and borderline cases. They emphasise, instead, that the natural kind methodology is 'adept' at dealing with properties that come in degrees.

3.3 The problem of multiple kinds

Here arises the epistemological challenge that borderline consciousness poses to scientists who investigate phenomenal consciousness as a natural kind. When these scientists discover more than one overlapping kind in their samples of conscious creatures, how can they identify the kind to which all and only conscious creatures belong? First, I shall explain why borderline consciousness leads scientists to discover multiple kinds in their samples of conscious creatures. Second, I will clarify this problem of multiple kinds by contrasting it with the phenomenon of multiple realisability in mental kinds. My analysis in this section is made in relatively abstract terms; in the next section, I will apply it with concrete details to debates from consciousness science.

Let me use the simplified model in Figure 3.2 to explain the link between borderline consciousness and the multiple kinds problem. As our operational criteria for attributing phenomenal consciousness indicate, phenomenal consciousness is commonly associated with a set of behavioural and neural properties. We neither codify nor demarcate these properties well. Suppose that most of these properties are divided into two clusters, superficial cluster$_1$ and cluster$_2$. Creatures with all the superficial properties are classified as conscious, while those with none of the properties are classified as not conscious. Those with the properties from only one superficial cluster count as borderline conscious.

In this model, scientists find two hidden structures related to the superficial properties. First, each structure is causally or constitutively related to a different

superficial cluster. Hidden structure$_1$ explains superficial cluster$_1$, while hidden structure$_2$ explains superficial cluster$_2$. Second, each structure is in the conscious creatures from the scientists' samples. Third, each structure defines a different kind. Hidden structure$_1$ defines kind$_1$, while hidden structure$_2$ defines kind$_2$. Since the conscious creatures from the scientists' samples belong to both kinds, kind$_1$ and kind$_2$ must overlap. The borderline conscious creatures belong to one, but not both, of these kinds.

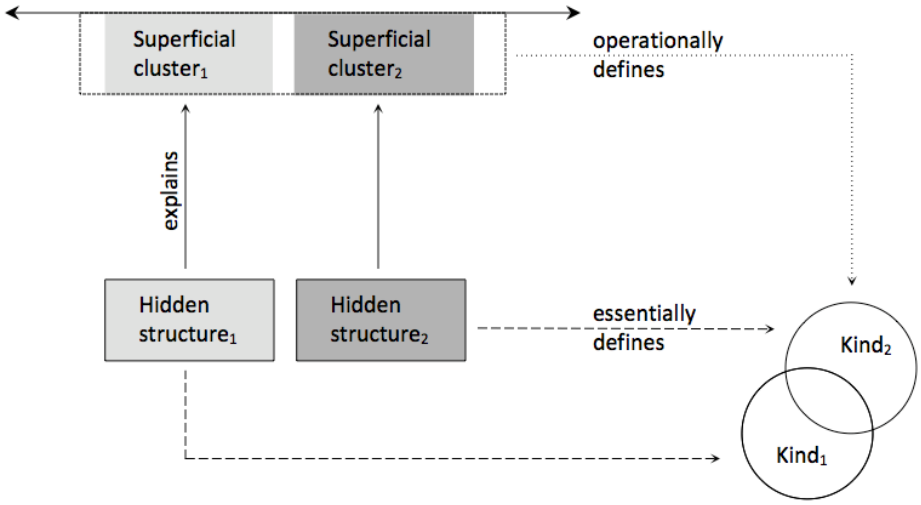


Figure 3.2: The multiple kinds problem

For a product of evolution such as phenomenal consciousness, the superficial clusters that actually exist are likely to be more numerous and complex. In turn, the hidden structures which scientists discover are likely to be more numerous and

complex.[12] I will mention some complexities later when I discuss the debates in consciousness science. But they do not affect the thrust of my epistemological challenge, which remains the same as that represented in Figure 3.2. Which one of the overlapping kinds discovered in the samples of conscious creatures is *the* kind to which all and only conscious creatures belong?

At this point, inference to the best explanation seems to offer no help – at least not through the epistemological norms that I have already cited. In the model from Figure 3.1, scientists face a set of *basic explananda*: the systematic correlation between a set of structural properties and phenomenal consciousness, as well as the causal or constitutive relation between the structural properties and the superficial properties associated with phenomenal consciousness. Any creature with these superficial properties is provisionally classified as conscious. Unless scientists have evidence of a common cause, the best explanation of these relations is a theoretical identity between phenomenal consciousness and the structural properties.

On the other hand, in the model from Figure 3.2, scientists face a set of *complex explananda*. Both hidden structure$_1$ and structure$_2$ are systematically correlated with phenomenal consciousness. Each hidden structure is causally or constitutively related to a different cluster of superficial properties associated with phenomenal consciousness. Any creature with all the superficial properties is provisionally classified as conscious. But a creature with the properties from only one cluster is

[12] Scientists may even find it difficult to distinguish and order the levels that define these structures. See Wimsatt (1976), especially the section on 'Complexity and the Incomparability of Levels'. He argues that this kind of 'interactional and descriptive complexity' is typical of 'efficiently organized complex functional systems' (254). It is an 'expected product' of evolution and other selection processes.

classified as borderline conscious. The epistemological norms that I have cited do not support any identity as the best explanation of these complex relations between phenomenal consciousness, hidden structures, and superficial properties.

Should scientists choose one of the hidden structures to be the structure of phenomenal consciousness? But there seems to be no reason to choose one structure over the other. I do not see how identifying phenomenal consciousness with one of the hidden structures, rather than the other, makes for a better explanation of the complex explananda. This move is *ad hoc*, re-classifying by fiat some borderline conscious creatures as conscious and others as not conscious. Should they choose the combination of both structures? Again, there seems to be no reason to do this. Identifying phenomenal consciousness with both structures does not make for a better explanation than identifying it with one of them. It is just as *ad hoc*, re-classifying by fiat all borderline conscious creatures as not conscious. Thus, the epistemological norms that I have cited do not pick out one of the overlapping kinds as the kind to which all and only conscious creatures belong. If scientists want to use inference to the best explanation here, they will need new norms to manage the multiple kinds.

I want to highlight some neglected remarks in Kripke (1980) and Putnam (1975a), which indicate that the multiple kinds problem crops up outside consciousness science. Kripke (1980) appeals to a concept's vagueness to account for the unexpected discovery of more than one kind. We generally assume that an 'initial sample' of something contains 'one uniform substance or kind' (136). But what happens when this assumption is misplaced?

> If the original sample has a small number of deviant items, they will be rejected as not really gold. If, on the other hand, the supposition that there is one uniform substance or kind in the initial sample proves more radically in error, reactions can vary: sometimes we may declare that there are two kinds of gold, sometimes we may drop the term 'gold'. (136)

Scientists have to decide somehow between these possibilities and others. Here Kripke brings up vagueness: 'To the extent that the notion "same kind" is vague, so is the original notion of gold.' He adds that, ordinarily, this vagueness 'doesn't matter in practice', though he does not explain the basis for his confidence. He also does not say what ought to happen in cases where the vagueness does matter to science.[13]

Putnam (1975a) also brings up the unexpected discovery of multiple kinds, though he does not connect it to borderline cases. He warns that our sample of something 'may have two or more hidden structures – or so many that "hidden structure" becomes irrelevant, and superficial characteristics become the decisive ones' (241). Putnam's example is jade. Nephrite and jadeite are two mineral species with distinct chemical structures, $Ca_2(Mg,Fe)_5Si_8O_{22}(OH)_2$ and $NaAlSi_2O_6$ respectively. Yet both species have come to be seen as types of jade.

Unfortunately, I do not think that this example helps scientists with my epistemological challenge. Nephrite and jadeite are not overlapping kinds. Most philosophers do not regard jade as a natural kind because it is associated with two unrelated chemical kinds (Block 2002b: 413; Hacking 2007b). As Hacking (2007b) observes, the term 'jade' and its translations now refer to a commercial or cultural

[13] In a footnote elsewhere, Kripke urges caution in handling vagueness: 'Logicians have not developed a logic of vagueness' (51). He relates vagueness to 'open texture', which allows for cases where 'the answer may be indeterminate' (50-1). But this claim offers us limited help. First, the open texture lies in the identity of a particular, not the boundary of a natural kind. Second, Kripke does not say what the indeterminacy implies.

kind that bears hardly any scientific interest. This kind does not contribute to any significant scientific explanations. Why did the Chinese begin to use '*yu*' to cover both nephrite and jadeite? Because its extension was 'determined by the workability of the substances, their polished appearance, and a deep cultural tradition' (272). The English eventually followed suit with the term 'jade' due to their 'business interests, not artistic or mineralogical ones' (273).

Next, I contrast the problem of multiple kinds with the phenomenon of multiple realisability. The multiple kinds problem, in its general form, is an epistemological challenge posed by borderline cases to scientists who investigate natural kinds. On the other hand, Putnam (1967) and Fodor (1974) are concerned with a metaphysical problem raised by the special sciences, in which a phenomenon is systematically associated with many distinct physical kinds. Putnam concentrates on multiple realisability in psychology, while Fodor extends it to economics. More recently, other philosophers debate the wider significance of multiple realisability.[14]

I focus on Putnam's arguments. Putnam (1967) notes how ambitious it is to claim that pain is identical to the same physical state across species:

> the physical-chemical state in question must be a possible state of a mammalian brain, a reptilian brain, a mollusc's brain (octopuses are mollusca, and certainly feel pain), etc… Even if such a state can be found, it must be nomologically certain that it will also be a state of the brain of any extra-terrestrial life that may be found that will be capable of feeling pain before we can even entertain the supposition that it may *be* pain. (436)

Indeed, the hypothesis promoted by the identity theorist is 'still more ambitious':

[14] Bickle (2013) surveys the philosophical debates on multiple realisability. Recent contributions to the literature include Sober (1999), Gillett (2003), Polger (2006), Aizawa and Gillett (2009), and Shapiro and Polger (2012).

every mental state is supposed to be a brain state. In Putnam's assessment, it is unlikely that, for each mental state, scientists can find a brain state common to every species capable of having that mental state. This would require the development of 'neurophysiological laws that are species-independent' (437).

It is more 'reasonable' to hope that scientists can find psychological laws that are species-independent (437). These laws will characterise 'the kind of functional organization that is necessary and sufficient for a given psychological state', such as pain or hunger. Each state of the functional organisation is realised by different physical states in mammals, reptiles, molluscs, and some extra-terrestrials. On the basis of these considerations, Putnam presents a hypothesis about the nature of mental states. It is 'more plausible' that mental states are functional states of an organism, rather than physical states of its brain (433). If mental kinds are functional kinds, then they can be realised by distinct physical kinds.

This appeal to multiple realisability has faced increasing criticism. Putnam (1988) himself disavows the functionalist hypothesis because he believes that mental states can be realised by different functional states. Some philosophers deny that multiple realisability poses a threat to the identity theory.[15] They argue that science permits the reduction of mental states to various species-specific brain states (Lewis 1969, 1994; Kim 1992). Other philosophers debate whether multiple realisability arises in a metaphysically significant way (Bechtel and Mundale 1999; Shapiro 2000; Shapiro and Polger 2012). Indeed, they disagree on how to distinguish metaphysically significant degrees or dimensions of multiple realisability (Gillett 2003; Polger and

[15] Their responses are outlined in Smart (2012), §5-6, and Bickle (2013), §2-3.

Shapiro 2008; Aizawa and Gillett 2009a; Bickle 2013, §3).

My aim is not to settle these debates. Rather it is to clarify the multiple kinds problem in consciousness science by distinguishing it from Putnam's problem. First, the two problems result from different kinds. In Putnam's problem, the relevant kinds are defined by physical structures which are related to the same superficial properties. These kinds cannot overlap. Each is defined by a physical structure specific to a species. Their *distinctness* is what prevents us from identifying a mental kind with a physical kind. On the other hand, in the multiple kinds problem, the kinds are defined by hidden structures which are related to different clusters of superficial properties. These structures need not be species-specific, or purely physical. Moreover, the kinds cannot be as distinct as those that motivate multiple realisability. Their *overlap* is assured because each contains the scientists' samples of conscious creatures.

Second, the problems require different solutions. Putnam's problem needs a *metaphysical* solution. We need to interpret the nature of mental states that are systematically associated with distinct physical states. So Putnam proposes his functionalist hypothesis as an alternative to the identity hypothesis. On the other hand, the multiple kinds problem needs an *epistemological* solution. Scientists need a strategy for managing the overlapping kinds in their samples of conscious creatures. Until this problem is solved, they cannot classify borderline conscious creatures more definitely.

Might Putnam's functionalist hypothesis nevertheless help with the multiple kinds

problem? After all, it concerns the nature of mental states, including phenomenal ones such as pain. I offer two reasons to be cautious. First, as I have noted, the functionalist hypothesis remains controversial among philosophers. Some argue that it is especially implausible when applied to phenomenal states (Block 1978 and 1980; Loar 1997: 615). Others argue that the neuroscientific investigation of mental states should not be pre-empted by contestable appeals to multiple realisability.[16] Second, Putnam's hypothesis rests on the assumption that mental states are shared by mammals, reptiles, molluscs, and some extra-terrestrials. This assumption cannot be granted by scientists who ask whether borderline conscious creatures such as frogs and octopuses are, in fact, conscious. Applying Putnam's hypothesis directly to phenomenal consciousness would beg the question. It remains possible, of course, that science will reveal the structure of phenomenal consciousness to be partly functional.

3.4 Three impasses in consciousness science

Now I will demonstrate how the multiple kinds problem produces three theoretical impasses on the nature of phenomenal consciousness. First, I describe a recent impasse between scientists who debate the neural structure of phenomenal consciousness. Then I clarify why the multiple kinds problem is so intractable in this debate. Second, I discuss an impasse between scientists who debate the development of foetal consciousness. I thereby emphasise that the multiple kinds

[16] See, for instance, Bechtel and Mundale (1999); Bechtel and McCauley (1999); Aizawa and Gillett (2009b, 2011); Bickle (2013), §2.3 and §3. Significantly, this consensus is formed by philosophers who disagree on both the extent of multiple realisability and its metaphysical significance.

problem hinders the more precise classification of borderline conscious creatures. Third, I explain why the multiple kinds problem will lead to an impasse on the possibility of artificial consciousness.

The neural structure of phenomenal consciousness

I begin with the scientific debate on the neural structure of phenomenal consciousness. Here I draw on Irvine (2013), who analyses this debate from 'a philosophy of science perspective'. She explains how an impasse arises between scientists who propose conflicting neural theories of consciousness. Following Shea and Bayne (2010), she interprets these theories to be the result of investigations into 'scientific kinds'. Since these kinds enable 'reliable predictions and broad generalisations', they are 'what a science is really about' (93). How can scientists discover the kind formed by all conscious creatures? They look for the neural structure of consciousness. According to Irvine, this structure is a 'commonly co-occurring' cluster of properties, whose co-occurrence is produced by underlying mechanisms.[17]

Unfortunately, as Irvine shows, multiple kinds show up in the scientists' search:

> Attempts to provide behavioural measures, neurophysiological measures, neural mechanisms, and neural correlates of phenomenal consciousness all fail because differing operationalisations split 'consciousness' into one of the many varied and fine-grained scientific kinds. These include different types of attention, different streams of information processing, decision-making and so on. (160)

[17] Here she assimilates the model in Shea and Bayne's natural kind methodology to the homeostatic property cluster model described in Boyd (1991, 1999).

With different operationalisations in mind, scientists propose different neural theories. Each operational test for consciousness focuses on a different cluster of behavioural properties associated with consciousness. Through investigation, scientists relate each cluster of behavioural properties to a different neural structure. Each neural structure defines a different kind. The result: the discovery of 'many varied and fine-grained scientific kinds', each with a claim to be *the* kind formed by all conscious creatures.

Irvine illustrates this multiple kinds problem with two dominant types of theories about phenomenal consciousness. The first type associates consciousness with the ability to integrate information and make it available across various psychological faculties. These faculties typically include attention, sensory perception, working memory, decision-making, and action. She cites the Global Workspace theory (Baars 1997; Shanahan & Baars 2007), the Global Neuronal Workspace theory (Dehaene & Naccache 2001; Dehaene *et al.* 2006) and the Informational Integration Theory (Tononi 2004, 2008).[18] These theories see attention, working memory, and reportability as 'essential features' of consciousness (82).

The second type of theories sees attention, working memory, and reportability as 'experimental confounds'. Consciousness is held to be distinct from attention and cognitive accessibility. Irvine cites the theories of phenomenal consciousness in Block (2005, 2007a) and Lamme (2004, 2006). For instance, Block (2007a) argues that 'phenomenal consciousness overflows cognitive accessibility' (481); we are conscious of more information than is available for our use in remembering,

[18] See Irvine (2013), 6-7.

reasoning, and reporting. Similarly, Lamme (2004) claims that 'phenomenal awareness' does not require attentional selection and processing by action- or memory-related areas of the brain.[19]

I use Figure 3.3 to illustrate how these two types of theories lead to overlapping kinds. How do scientists operationalise consciousness differently?[20] For scientists who see attention and cognitive accessibility as constitutive of consciousness, the behavioural tests require attentional behaviour and 'subjective' reports. In attentional blink tests, subjects have to report a quick succession of two different stimuli; in change blindness tests, they have to report a change between two alternating and nearly identical stimuli. For scientists who see attention and accessibility as not constitutive of consciousness, reportability is an 'experimental confound' rather than 'essential marker' (45). Thus they prefer objective tests of task performance. In these tests, subjects have to make 'forced-choice' responses, which indicate their sensitivity to stimuli that are not attended to and therefore not available for report.

[19] Their views are reiterated and defended in Block (2011) and Lamme (2010). See also Dretske (2004, 2007).
[20] Irvine (2013), §2 and 3, evaluates various subjective and objective measures of consciousness. Here I highlight aspects of her account that are related to the multiple kinds problem.

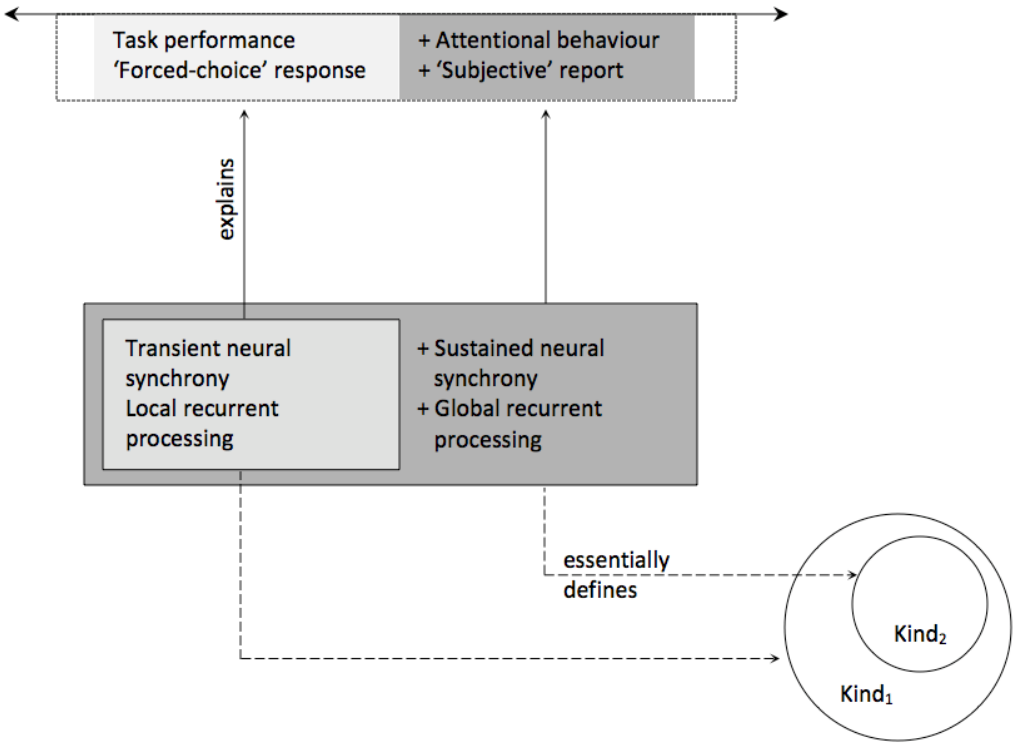


Figure 3.3: Multiple kinds in the phenomenal consciousness debate

With these tests, scientists can look for neural structures causally or constitutively related to behaviour that they operationally associate with consciousness. Here are their provisional discoveries, partly synthesised by Irvine: successful task performance and the ability to make appropriate 'forced-choice responses' are related to transient neural synchrony and local recurrent processing, while attentional behaviour and the ability to make appropriate subjective reports require sustained neural synchrony and global recurrent processing too.[21]

[21] For more details, see Irvine (2013), §5.3 on 'Neurophysiological Measures of Consciousness'. *Neural synchrony* occurs when neurons fire together in different parts of the brain. 'Transient' vs. 'sustained' refer to how well and how long the neurons fire together. *Recurrent processing* occurs when

Following these discoveries, global workspace theories focus on the more complex neural structure (Dehaene & Naccache 2001; Del Cul *et al.* 2007). On the other hand, theories that dissociate phenomenal consciousness from attention and cognitive accessibility imply that the less complex structure is sufficient for phenomenal consciousness (Lamme 2004, 2006; Block 2007a). As Figure 3.3 indicates, the more complex structure includes the less complex one. Each structure defines a different kind of creature. $Kind_{2,}$ which is defined by the more complex structure, is nested within $kind_1$, which is defined by the less complex one. Irvine does not mention the overlapping nature of these kinds; but it should be obvious that conscious creatures in the scientists' samples belong to both $kind_1$ and $kind_{2.}$

Which kind consists of all conscious creatures? Irvine (2013) highlights an epistemological difficulty here. Scientists who disagree on the appropriate behavioural markers of consciousness will differ on its neural structures.

> [N]europhysiological measures are simply measures of whatever phenomena are operationalised by a behavioural measure…The dependence of neurophysiological work on behavioural operationalisations of consciousness means that they do not provide independent evidence. (85)

Therefore, scientists are stymied by a 'lack of convergence' on one cluster of commonly co-occurring properties and mechanisms (86). Each type of theory focuses on a different cluster of co-occurring properties and mechanisms. When scientists attend to these different explananda, they produce conflicting hypotheses

---

information is processed in both 'forward' and 'backward' directions, between early and later areas of the brain. 'Local' vs. 'global' indicate how widespread the recurrent processing is in the brain.

I shall use these as proxies for the other structures, processes, and mechanisms that are said to constitute consciousness.

about the neural structure of phenomenal consciousness.[22] Irvine emphasises that this problem is not transient, but 'chronic' (151): it appears repeatedly in scientific debates about the nature of consciousness.[23]

According to Irvine, scientists disagree on the best way to operationalise phenomenal consciousness because of their 'conflicting intuitions and pre-theoretical commitments' (53). For instance, operational tests for the above theories are based on conflicting intuitions and commitments about the role of attention and cognitive accessibility. They reflect 'two very different conceptions of where along a chain of processing consciousness occurs, and thus how it should be measured' (52).[24] Irvine does not explain why there is room for these conflicting intuitions and commitments. But she clarifies why the intuitions and commitments are 'untestable'. If the correctness of an operational test varies according to the commitments held, then none of the tests can be used to evaluate the commitments without begging the question.

I want to extend this analysis in two ways. First, Irvine does not draw the link between borderline consciousness and the scientists' discovery of multiple kinds. But it is borderline consciousness which creates room for scientists to have conflicting intuitions and commitments about the role of attention and cognitive accessibility. According to our operational criteria for attributing phenomenal consciousness,

[22] Some scientists who favour global workspace theories deny that science at this stage can investigate the possible dissociation between phenomenal consciousness and cognitive accessibility. See Irvine (2013), 108, who quotes Dehaene et al. (2006) and Kouider et al. (2007).

[23] See also Irvine (2013), 79: 'Unless it is accepted that a particular kind of behavioural response is a good marker of consciousness, the accompanying neurophysiological marker will also be a questionable marker of consciousness. This problematic relationship between behavioural and neurophysiological markers of consciousness is repeatedly found'.

[24] For two opposing views on the role of attention, see Koch and Tsuchiya (2007, 2012), and De Brigard and Prinz (2010). I return to this debate in Chapter 4.

creatures who are capable of forced-choice responses, attentional behaviour, and subjective reports are conscious of the stimuli. Those who are capable of only forced-choice responses count are neither definitely conscious nor definitely not conscious of the stimuli. Without their borderline consciousness, scientists would not have as much room to deploy different operational tests for phenomenal consciousness. And they would not discover so 'many varied and fine-grained kinds', each of which contains their samples of conscious creatures.[25]

Second, Irvine does not explain why inference to the best explanation cannot address this 'lack of convergence'. Instead, she focuses on why operational tests cannot do so. Let me say why the epistemological norms that I cited earlier do not help, even if scientists discover more fine-grained behavioural responses and their neural correlates.[26]

Compare the two hypotheses represented in Figure 3.4. In the first hypothesis, scientists identify phenomenal consciousness with the most complex structure (neural structure$_{1+2+3}$). To explain the full range of behavioural responses (behavioural responses$_{1+2+3}$), they must refer to the other neural structures nested in this most complex structure. But they can demarcate the same structures, and thereby explain the same range of behavioural responses, even if they adopt the second hypothesis and identify phenomenal consciousness with a less complex structure (neural structure$_1$). What matters in accounting for the behavioural

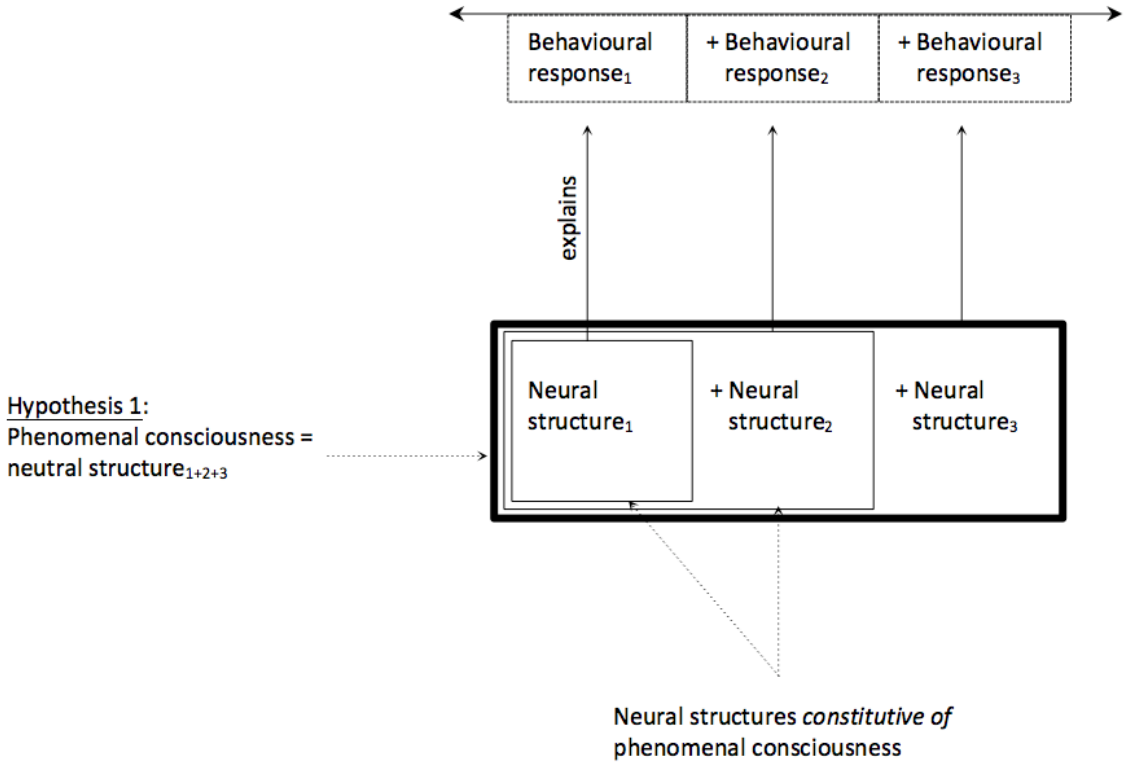
[25] Indeed, the relevant kinds are even more fine-grained than our discussion suggests. Irvine argues that even 'reportability and related notions of "access" are incredibly vague and fail to pick out a single target phenomenon' (117). Rather they refer to a 'vast range of phenomena that are produced by a range of task-specific mechanistic components'.

[26] I analyse an example in §4.3, in which Block (2007a, 2008) and Prinz (2003, 2005) disagree on the role of attention in phenomenal consciousness.

responses *per se* are the demarcations between the neural structures, not the identity between phenomenal consciousness and any of the neural structures.

I emphasise that, whichever hypothesis they choose, scientists will be committed to the same number of neural structures in order to account for the total explananda. If they choose neural structure$_{1+2+3}$ to be phenomenal consciousness, then they must construe neural structure$_1$ and structure$_2$ as structures *constitutive of* phenomenal consciousness. If they choose neural structure$_1$ to be phenomenal consciousness, then they must construe neural structure$_{1+2}$ and structure$_{1+2+3}$ as structures *constructed from* phenomenal consciousness. So neither hypothesis leads to an explanation that is more parsimonious.

What of the simplicity gained through unification? The first hypothesis unifies the less complex structures by interpreting them as constituents of phenomenal consciousness. The second hypothesis unifies the more complex structures by interpreting them as products of consciousness. Both hypotheses emphasise different dimensions of unification, but exhibit similar degrees of unification. Neither hypothesis seems to offer a more unified account of the total explananda. There appears to be no reason to pick either the kind defined by neural structure$_{1+2+3}$ or the kind defined by neural structure$_1$ as the kind formed by all conscious creatures. This is the intractability of the multiple kinds problem.



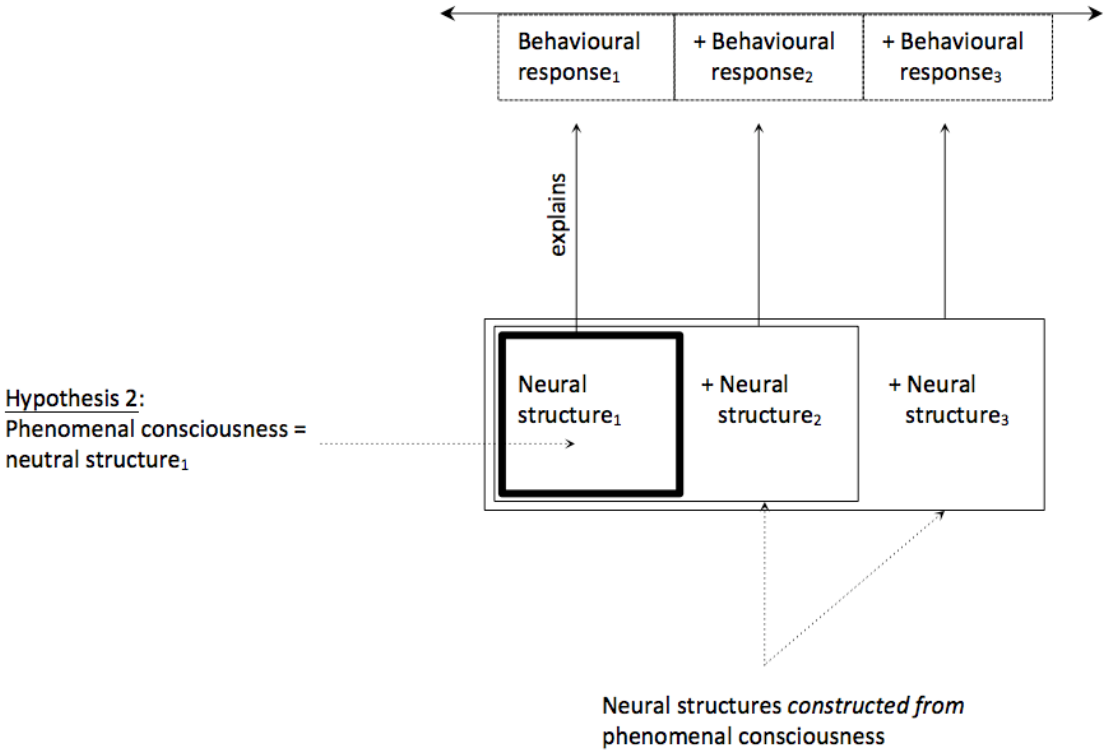


Figure 3.4: Two hypotheses on the nature of phenomenal consciousness

The development of foetal consciousness

In the debate on foetal consciousness, scientists seek to classify late-stage human foetuses more precisely with respect to phenomenal consciousness. Their investigation centres on this problem: when do the foetuses become conscious enough to feel pain?[27] Glover and Fisk (1999) offer a standard definition of the problem: 'Pain is a subjective experience. The foetus cannot tell us what it is feeling, and there is no objective method for the direct measurement of pain' (881). They thereby distinguish pain from nociception. Nociception is the body's physiological reaction to noxious stimuli. It includes neural transmissions from stimulated receptors to the spinal cord, which lead to reflex reactions and further transmissions up to the brain. Everyone agrees that some foetuses are capable of nociception. What is controversial is when these foetuses start to experience the unpleasant sensations that adults normally experience as a result of nociception.

According to our operational criteria for attributing phenomenal consciousness, late-stage human foetuses are classified as borderline conscious. From about 14 weeks onwards, foetuses and pre-term neonates develop behavioural responses to noxious stimuli which resemble those in infants and adults. They start by moving away from touch (Glover and Fisk 1996). By 26 weeks, neonates also respond to noxious stimuli with measurable withdrawal movements (Andrews and Fitzgerald 1994). At 28-30 weeks, neonates respond to heel lancing with facial movements similar to those in adults suffering pain (Craig *et al*. 1993). As the foetuses mature, they are capable of

[27] For philosophical analyses of this debate, see Benatar and Benatar (2001), Derbyshire (2001), and Chin (2007), ch. 2. Derbyshire and Raja (2011) and Chin (2011), §3, assess recent developments. There are reviews of the scientific literature in Lee *et al.* (2005), Anand (2006), Derbyshire (2006), Lowery *et al.* (2007), Templeton *et al.* (2010), and Lagercrantz (2014).

increasingly complex facial movements, including those that would indicate 'pain' or 'distress' in pre-term infants (Reissland *et al.* 2013).

To advance their debate, scientists increasingly appeal to hypotheses about the 'neural basis' of phenomenal consciousness in humans (Chin 2011, §3.1). They assume that phenomenal consciousness requires neural activity in the cortex.[28] But they propose different cortical structures to be sufficient for foetal consciousness. Some scientists point to indirect connections between the thalamus and cortex, which route through the cortical subplate (Anand 2006; Lowery *et al*. 2007). These appear in foetuses at 16-18 weeks. Other scientists emphasise direct thalamo-cortical connections that form at 23-24 weeks (Glover and Fisk 1999). In a multidisciplinary review, Lee *et al.* (2005) argue that 'conscious recognition or awareness' is based on constant and synchronous electrical activity in the thalamo-cortical connections. In tests of pre-term neonates, such activity is detected only at 29-30 weeks.[29]

These structures are nested in the following sense: at each stage of development, new structures are added to structures from earlier stages. For synchronous electrical activity to be possible, the thalamo-cortical connections must already be in place. Direct thalamo-cortical connections supplement indirect ones developed earlier, which are in turn based on subplate-cortical connections. Although the

[28] This assumption is widely shared: see Lloyd-Thomas and Fitzgerald (1996); Glover and Fisk (1996); Smith *et al.* (2000); Lee *et al.* (2005); Mellor *et al.* (2005); Lowery *et al*. (2007).
McCullagh (1997) offers some early dissent: 'Two questions — whether the cortex is normally involved in the appreciation of pain and whether it is necessary for this — are regularly conflated' (302). Merker (2007) presents a systematic challenge to this assumption. In commentary, Anand (2007) and Devor (2007) support his challenge and apply it to the case of pain perception.

[29] Many scientists cite data about pre-term neonates to make approximate inferences about foetuses of the same age. Mellor *et al.* (2005) challenge this use of the pre-term neonate as a model of the foetus; but see Chin (2007), 59-61, which shows that their own inferences rest on the same model.

subplate itself disappears after 32 weeks when the cortical plate matures into the cerebral cortex, functional subplate neurons remain in the adult cortex and play vital roles in cognitive processing.

I use Figure 3.5 to represent how foetuses develop a spectrum of increasingly complex behaviour: behavioural response$_1$ at about 14 weeks; behavioural responses$_{1+2}$ at about 26 weeks; behavioural responses$_{1+2+3}$ at about 28 weeks; and so on. These behavioural responses are causally or constitutively related to a set of nested neural structures. The first structure consists primarily of indirect thalamo-cortical connections; it explains behavioural response$_1$. The second structure, which adds direct thalamo-cortical connections to the indirect ones, explains behavioural responses$_{1+2}$. And the third structure, which requires a specific form of electrical activity in the thalamo-cortical connections, explains behavioural responses$_{1+2+3}$.

Which of these neural structures is the basis of phenomenal consciousness in humans? There seems to be no reason to choose one structure over another. Each neutral structure explains only a cluster of behaviour associated with conscious humans who are in pain. But each cluster of behaviour — behavioural response$_1$, or behavioural responses$_{1+2}$, or behavioural responses$_{1+2+3}$ — only suffices to indicate borderline consciousness in the foetuses. When different groups of scientists choose different neural structures as the basis of phenomenal consciousness, they arrive at an impasse on foetal consciousness.

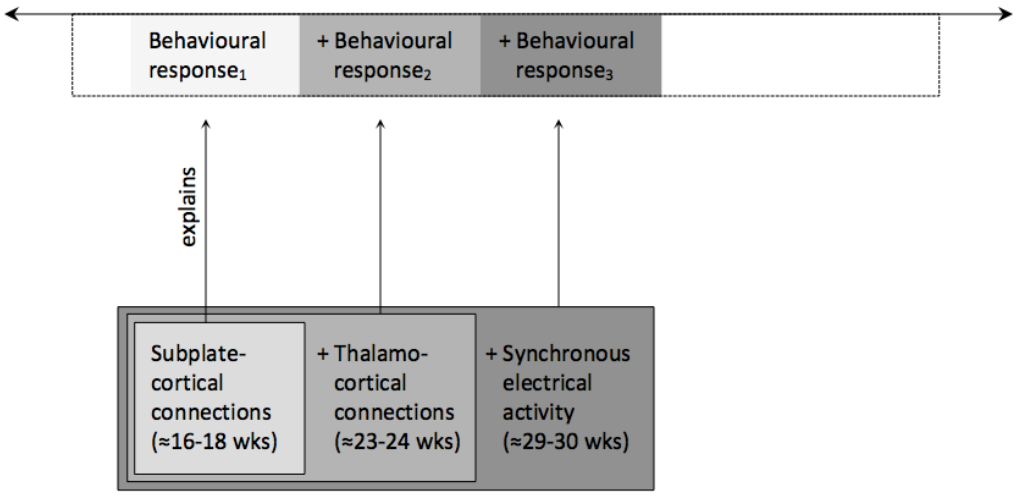


Figure 3.5: Multiple structures in the foetal pain debate

As Figure 3.6 shows, each neural structure defines a different kind. The first structure defines $kind_1$, consisting of humans from the age of about 16 weeks and above. The second structure defines the smaller $kind_2$, consisting of humans from about 23 weeks and above; the third defines the yet smaller $kind_3$, consisting of humans from about 28 weeks and above. Because the neural structures are nested, the kinds make up a nested hierarchy. Thus $kind_2$ is contained within $kind_1$; $kind_3$ within $kind_2$. Conscious humans in the scientists' samples belong to all three kinds. So which is the kind to which all and only conscious humans belong? Unless scientists can solve this multiple kinds problem, they cannot determine which human foetuses between 16 to 30 weeks should be re-classified as conscious enough to feel pain.

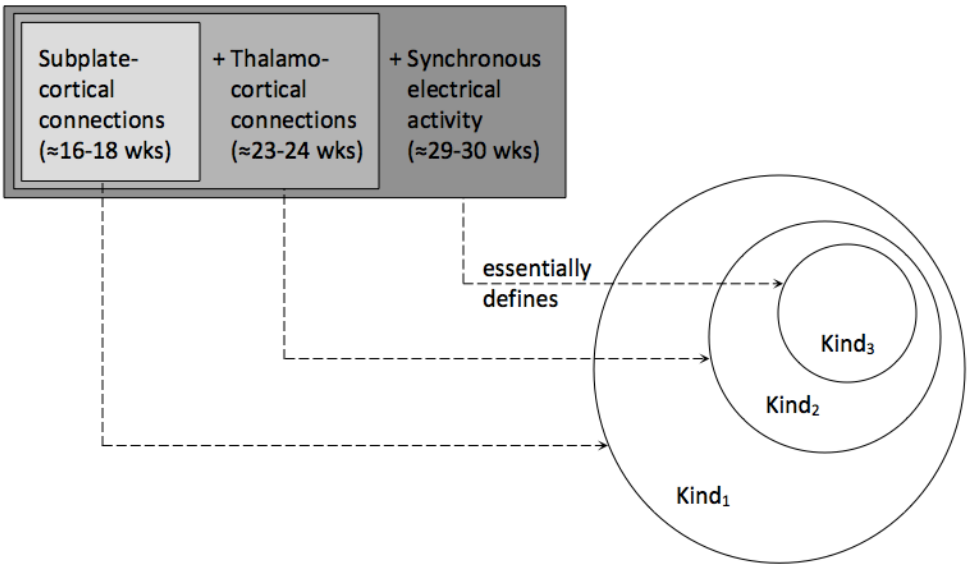


Figure 3.6: Multiple kinds in the foetal pain debate

Similar impasses arise when scientists cite neuroscientific data to support their attributions of phenomenal consciousness to non-human animals at the borderline (Seth *et al.* 2005; Edelman *et al.* 2005; Baars 2005; Edelman & Seth 2009).[30] Some argue more specifically over whether fishes can feel fear and pain (Rose 2002, 2007; Chandroo *et al.* 2004; Braithwaite & Boulcott 2007; Braithwaite 2010); others argue over the possibility that octopuses and other invertebrates are conscious (Mather 2008, §2; Elwood 2011; Godfrey-Smith 2013). The prevalence and persistence of these impasses suggest that simply combining data across species will not help to classify borderline conscious creatures more precisely. Indeed, in each of these impasses, scientists already appeal to data drawn from other species to support their side of the debate.

[30] The scientific literature is vast: see, for instance, Panksepp (2005); Watt (2005); Wynne (2005); Merker (2007); Cabanac *et al.* (2009); Feinberg and Mallatt (2013).

The possibility of artificial consciousness

I end my survey by considering an impasse that has yet to arise in consciousness science.[31] Prinz (2003) asks whether we can, in principle, 'create consciousness out of inorganic stuff' (112). This is 'roughly equivalent' to the question of whether we are conscious in virtue of the brain's computational properties rather than its biological ones. Why? If biological properties turn out to be constitutive of phenomenal consciousness, then a machine without these properties cannot, in principle, be conscious. According to Prinz, this question is bound to stymie scientists who investigate phenomenal consciousness as a natural kind: 'It simply isn't the case that scientific investigations into the nature of consciousness will make questions of machine consciousness disappear' (117).

He traces this impasse to the following 'problem of levels'. Suppose that scientists build their best possible theory of human consciousness. This theory contains a description of the systems that constitute phenomenal consciousness in humans. Suppose further these systems include working memory, perception, and attention. Prinz argues:

> The real difficulty stems from the fact that we can describe the key systems involved in consciousness at varying degrees of abstraction...In each case, it is far from clear how we should determine which levels matter. (120)

For instance, working memory can be analysed in at least three levels – in terms of

[31] Here I focus on the philosophical analysis in Prinz (2003, 2005). Block (2002b) and Papineau (2002), ch. 7, present similar impasses on artificial consciousness. McLaughlin (2003) and Hohwy (2004) respond to the arguments in Block (2002b); in Chapter 4, I assess the arguments in Papineau (2002).

its psychological profile, algorithm, and implementation.[32] At the psychological level, working memory appears to be a 'temporary storage faculty with some executive capacities' (120). At the algorithmic level, these capacities are rendered as a set of 'rules, representations, operations, and so forth'. Finally, at the implementation level, these rules and representations are realised biologically in 'physical events in the nervous system'.

Are the processes at the biological level constitutive of phenomenal consciousness? To test this claim, scientists might experiment on their samples of conscious humans. Their method, following a standard scientific heuristic, is to look for what Prinz calls 'difference-makers' (121).[33] It involves changing the processes at the implementation level, while keeping constant the processes at the other two levels. If this intervention makes a difference to consciousness in the samples, then processes at that level may be constitutive of consciousness.

Suppose that it is technically possible to substitute silicon chips for the neurons in an experimental subject's nervous system. And suppose that it is nomologically possible to do this while keeping constant the subject's psychological profiles and algorithms. Here arises the in-principle difficulty with this experiment. The altered subject will become a 'functional duplicate of a normal person with a brain', except that he is made of silicon (123). So he will behave exactly as a normal conscious human does –

[32] Prinz adapts these levels from the model of visual perception in Marr (1982). He emphasises that each level is multi-tiered, with 'scores of levels to choose from'. Moreover, there is 'no established method for individuating levels' (120) and 'no reason to expect perfectly convergent levels when comparing inorganic and organic devices' (121). Like him, I put aside these concerns to focus on the in-principle problem of determining which levels matter.

[33] See also Irvine (2013), §6.3.1, who elaborates the rationale for this heuristic. It is part of a methodological framework for demarcating the mechanisms underlying a phenomenon. She draws on the account of neuroscientific practice in Craver (2009), which explains 'how background conditions are separated from constitutive components of a system'.

reporting pain, showing signs of anger, apparently 'seeing sunsets and smelling roses'. By design, the artificial duplicate will meet the behavioural criterion that we ordinarily use to classify humans as conscious. This implies, however, that scientists do not have a genuine test to dissociate the contribution of biological structure from that of functional structure. In Prinz's diagnosis: 'The measures of experience are behavioural, so the obvious test for the phenomenal impact of matter is going to be utterly uninformative' (123).

The result is another theoretical impasse about phenomenal consciousness. To test for the role of biology, scientists need to rely on the artificial duplicate's behaviour, including his reports of conscious experiences. Yet they cannot take his behaviour at face value without begging the question. Here is how Prinz analyses the scientists' quandary:

> With inorganic brains, we must base conclusions about phenomenology on behaviour, and behaviour is not sufficient evidence. To assume otherwise would beg the question against those who say biology is crucial. But we cannot assume that biology is crucial, because that would beg question against the other side. The problem is that we cannot decide between these options. (125)

As he emphasises, this is a 'serious epistemological problem' that arises when scientists seek to discover, via experiment, if biology matters to consciousness (130). It does not draw support from a 'radical form of scepticism'.

Prinz's analysis emphasises the significance of this epistemological problem. If scientists cannot discover whether phenomenal consciousness can be realised in a non-biological machine, then they cannot discover the nature of phenomenal

consciousness. Although scientists are ‘very close to understanding the material basis of consciousness’, they will ‘never get all the way there’ (111). In this sense, he acknowledges that the study of artificial consciousness cannot be neglected in the science of phenomenal consciousness.

Moreover, Prinz (2003) suggests that a similar epistemological problem arises when scientists study ‘other possible creatures’ which share the psychological profile of conscious humans, but differ from us subtly at the algorithmic level (123). For instance, their working memory may have different storage capacities and executive functions. Prinz (2005) goes a step further, stating that the same problem arises for creatures which partly resemble us in psychological profile, but are ‘quite different’ at the neural level. His examples are octopuses, pigeons, bees, and slugs. However, he does not show how the difference-maker problem applies to them. He says only that he sees ‘no obvious way’ to discover ‘how similar a mechanism must be to those found in our own brains in order to support conscious experiences’ (391).

I stress that this is another situation in which borderline consciousness leads scientists to discover more than one overlapping kind in their samples of conscious creatures. According to our operational criteria for attributing phenomenal consciousness, artificial duplicates count as borderline conscious. They meet the behavioural criterion, but not the neurological one. These duplicates give rise to Prinz’s ‘problem of levels’, which is a special case of the multiple kinds problem.

Let me use Figure 3.7 to clarify the connection beween these problems. In the thought experiment, scientists will discover two hidden structures that are systematically related to the behaviour which we associate with conscious humans. The first is a functional structure shared by conscious humans and artificial duplicates. The second is a biological structure found in conscious humans; this biological structure is one way to realise the functional structure. Since the biological structure realises the functional structure, the kind defined by the former is nested within the kind defined by the latter. The functional structure defines $kind_1$, which includes conscious humans and artificial duplicates. The biological structure defines $kind_2$, which excludes the duplicates. I note that Prinz's thought experiment cannot show that $kind_1$ and $kind_2$ are, in fact, natural kinds; we do not know yet if the artificial duplicate is nomologically possible.[34]

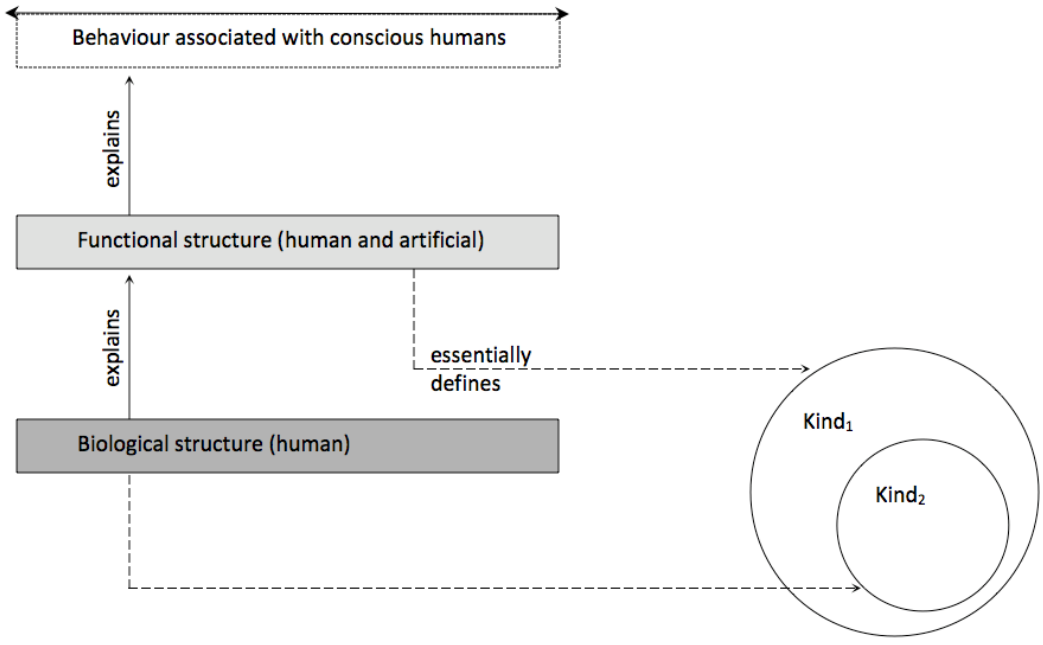


Figure 3.7: Multiple kinds in the artificial consciousness debate

[34] I shall come back to this point in Chapter 6.

Suppose that scientists will discover both kinds. Can inference to the best explanation help to address their predicament? Here is the set of complex explananda at stake: Both the functional and biological structures are systematically correlated with phenomenal consciousness. Both structures are also systematically related to the same behaviour. The functional structure is constitutively related to the behaviour associated with conscious humans; the biological structure is related to the same behaviour because it implements the functional structure. In addition, creatures that possess the functional structure but not the biological one are classified as borderline conscious.

There seems to be no reason to identify phenomenal consciousness with one structure rather than another. If scientists focus on the systematic relation between the functional structure and the behaviour associated with conscious humans, then inference to the best explanation supports an identity between phenomenal consciousness and the functional structure. But this move is *ad hoc*, re-classifying by fiat artificial duplicates as conscious. On the other hand, if scientists focus on the equally systematic relation between the biological structure and the behaviour associated with conscious humans, then inference to the best explanation supports an identity between phenomenal consciousness and the biological structure. Yet this is just as *ad hoc*, re-classifying by fiat the same duplicates as not conscious.

Whether the scientists identify phenomenal consciousness with the functional or biological structure, they must refer to both structures to account for the total explananda. If they identify phenomenal consciousness with the functional structure, then they have to use the biological structure to explain how this functional

structure is realised differently in conscious humans and their artificial duplicates. If they identify phenomenal consciousness with the biological structure, then they must use the functional structure to explain why the artificial duplicates share the same behaviour as conscious humans even though they do not share the humans' biology.

Again, neither hypothesis offers a more unified explanation of the total explananda. The first hypothesis interprets the biological structure as only one realisation of phenomenal consciousness, while the second interprets phenomenal consciousness as only one realisation of the functional structure. I conclude that the epistemological norms currently associated with theoretical identities are as inept in this impasse as in the other two.

3.5 Conclusion

In this chapter, I distinguished two epistemological challenges that borderline consciousness poses to consciousness science: the problem of classification and the problem of multiple kinds. According to my analysis, the more promising strategy for re-classifying borderline conscious creatures is to build neural theories of phenomenal consciousness. My analysis also highlighted some epistemological norms and metaphysical assumptions behind this strategy. First, I clarified the epistemological norms that govern inference to the best explanation when it is used to justify theoretical identities. Second, I interpreted this strategy to be the investigation of phenomenal consciousness as a natural kind.

Next, I turned to the epistemological challenge that confronts scientists who investigate phenomenal consciousness as a natural kind. I explained why borderline consciousness leads them to discover more than one overlapping kind in their samples of conscious creatures. Then I demonstrated how this multiple kinds problem produces three theoretical impasses on phenomenal consciousness, foetal consciousness, and artificial consciousness. I emphasised the intractability of these impasses by showing that the epistemological norms cited earlier do not suffice to solve the multiple kinds problem.

Chapter 4

Three paths to phenomenal pessimism

4.1 Introduction

In Chapter 3, I explained how borderline consciousness poses an epistemological challenge to scientists who investigate phenomenal consciousness as a natural kind. It leads these scientists to discover more than one overlapping kind in their samples of conscious creatures. So how can they identify the kind to which all and only conscious creatures belong? I demonstrated that this multiple kinds problem produces three theoretical impasses on the nature of phenomenal consciousness.

My analysis highlights two aspects of this multiple kinds problem. The first aspect is its surprising *ubiquity*. It affects scientists who investigate the neural structure of phenomenal consciousness, the development of foetal consciousness, and the possibility of artificial consciousness. Indeed, as Kripke suggests, a generalised version of the problem afflicts other scientists who use the natural kind methodology to investigate phenomena with borderline cases. The second aspect is the problem's seeming *intractability*. It is not obvious how scientists can solve it by inference to the best explanation if they follow the epistemological norms currently used to justify theoretical identities.

Here I shall explain and evaluate three philosophical responses to the multiple kinds problem in consciousness science. In different ways, Irvine (2013), Prinz (2003, 2005), and Papineau (2002, 2003) argue that the problem cannot be solved. First, I

address Irvine's recommendation that scientists eliminate the concept of consciousness from their research. Her view is based on methodological grounds. Second, I assess Prinz's view that scientists are irremediably ignorant of the nature of phenomenal consciousness, even though they can discover a lot about human consciousness. His view is partly supported by an empirical theory of human consciousness. Third, I examine Papineau's view that the concept of consciousness-as-such suffers from referential indeterminacy. His view is based on arguments on the special nature of phenomenal concepts.

These three philosophers are pessimistic, in different degrees, about the science of phenomenal consciousness. After assessing their arguments, I conclude that this phenomenal pessimism is not warranted. I also claim that their arguments furnish us with three clues on how to address the multiple kinds problem. The most important clue points to the possibility that scientists can solve this problem by an empirically constrained form of stipulation.

4.2 Irvine on methodological eliminativism

Irvine (2013) urges scientists to take an eliminativist stance on consciousness. By this, she means that the concept of consciousness is '*not* a viable scientific concept' (11). It should be eliminated from scientific theorising. She clarifies this claim in two ways. First, she limits her claim to the concept of consciousness, rather than the phenomenon of consciousness:

> …the claim is not that consciousness or experience does not exist (whatever that might mean), but that 'consciousness' is not a viable scientific concept. That is, there can be no scientific measures, theories or mechanisms of consciousness because 'consciousness' is just not a concept that is amenable to scientific methodology. (165)

I note her puzzlement, even vexation, expressed in 'whatever that might mean'. She implies that the concept of consciousness is so problematic that there is no clear meaning to the claim that consciousness as a phenomenon does, or does not, exist.[1]

Second, Irvine specifies that the viability of scientific concepts be assessed in terms of their 'utility' (11). This requires an assessment of these concepts in three dimensions: 'whether they identify scientific kinds, their epistemic role, and their pragmatic value'. Her perspective is avowedly from the philosophy of science. She aims to evaluate consciousness science 'according to the standard practices found elsewhere in science, particularly as formulated in philosophy of psychology, biology and neuroscience' (8). This promises a 'more directly naturalistic' approach to the limits of consciousness science (9).

Irvine's arguments for eliminativism are based on methodological considerations. She evaluates 'the science of consciousness as it stands' – focusing on 'whether it is able to function as a science, and what the status of "consciousness" is given current empirical evidence' (165). She thereby distinguishes her eliminativism from 'other radical or eliminativist positions related to consciousness' (164). For instance, Wilkes (1984, 1988) relies on linguistic considerations rather than methodological ones.

[1] The point is made murkier when she adds that 'all materialists should presumably feel the pull of arguments that "consciousness" just is not a *phenomenon* that has a place in science' (167). Earlier she evaluates whether '"consciousness" and its subtypes are appropriate target *phenomena* for scientific research' (154). However, in a footnote at the beginning, she specifies that '"consciousness" will be used throughout to denote the *concept* of consciousness' (3). (All italics are mine.)

According to Irvine, Wilkes argues that the concept of consciousness 'has not had a stable referent over time, does not have a stable referent over languages, and has many different meanings even in western languages' (164). Thus, its current use is too 'recent and culturally relative' to pick out a scientific kind.

Dennett (1988, 1991) serves as a more complicated, though illuminating, contrast. On one hand, unlike Irvine, he is happy for scientists to use the concept of consciousness. Indeed, he draws on cognitive science to build the Multiple Drafts theory of consciousness, which identifies consciousness with what he calls 'cerebral celebrity'.[2] He is committed to the 'reality' of conscious experience: conscious states have properties which define their 'experiential content' (1988: 619). On the other hand, Dennett denies that these properties include qualia. As I noted in §2.2, he focuses on a specific conception of qualia that attributes to them a combination of these scientifically problematic properties: ineffable, intrinsic, private, and directly or immediately apprehensible in consciousness.

According to Irvine, Dennett is 'trying to give' the concept of consciousness 'a very different meaning' (147). This includes rejecting 'most of the questions and intuitions associated with consciousness' (165). She does not say how he rejects them, so I will briefly explain the basis of his eliminativism. Instead of methodological considerations applied to scientific debates, Dennett (1988) provides a series of thought experiments as 'intuition pumps':

> I want to shift the burden of proof, so that anyone who wants to appeal to

[2] See Dennett (1993), 929: 'Consciousness is cerebral celebrity – nothing more and nothing less. Those contents are conscious that persevere, that monopolize resources long enough to achieve certain typical and "symptomatic" effects--on memory, on the control of behavior, and so forth.'

> private, subjective properties have to prove first that in so doing they are *not* making a mistake…I want to make it just as uncomfortable for anyone to talk of qualia – or 'raw feels' or 'phenomenal properties' or 'subjective and intrinsic properties' or 'the qualitative character' of experience – with the standard presumption that they, and everyone else, knows what on earth they are talking about. (620)

He intends to show that these theoretical concepts of qualia are too vague to be scientifically useful. Worse, the pre-theoretical concept is 'so thoroughly confused': any useful concept will be 'so radically unlike' this ancestor that it is 'tactically' better to eliminate all talk of qualia from science. For Dennett, whether we should eliminate an entity or identify it with something less problematic is 'often best seen not so much as a doctrinal issue but as a tactical issue' (639, note 2).[3] As I show below, Irvine agrees with him that eliminativism is partly pragmatic in nature.

So what are Irvine's arguments for eliminating all talk of consciousness from science? At the heart of her arguments is the discovery of multiple kinds. As I noted in §3.4, her investigation into the dominant theories of phenomenal consciousness reveals 'no convergence' on a cluster of commonly co-occurring properties and mechanisms (156). Instead, each type of theory identifies a different cluster of properties and mechanisms— 'the product of the precise behavioural operationalisation of consciousness that is used in an experimental paradigm' (157). In effect, scientists have discovered multiple kinds in their samples of conscious creatures. To reiterate Irvine's worry: the 'differing operationalisations split "consciousness" into one of many varied and fine-grained scientific kinds' (160).

[3] For critical discussion, see Block (1995b). Block concedes the truth that 'the difference between eliminativism and reductionist realism is often purely tactical' (237), but he denies that this applies to the theory of consciousness proposed in Dennett (1991).

How does this methodological predicament support eliminativism? Irvine interprets the predicament as part of the failure of 'standard and indispensable' methods used in science to discover a natural kind — including what Shea and Bayne (2010) call the 'natural kind methodology'. From her perspective, this failure in consciousness science 'strongly suggests' that the concept of consciousness 'can be safely and productively eliminated from the scientific domain' (157).

However, as Irvine recognises, this inference faces two related objections. Both centre on the fact that her assessment – however thorough and nuanced – is of science 'as it stands'. The first objection counters that the failure of standard methods may be temporary. Perhaps it results from consciousness science 'currently being an immature science, so not yet possessing appropriate experimental methods or theoretical frameworks' (151). The second objection posits a neglected or new method for tackling the multiple kinds problem: 'It could be argued that we may simply be missing appropriate methods, and that they will be developed through further research' (155).

Against the first objection, Irvine can point to convincing evidence that the multiple kinds problem is robust and recurrent. First, as her examples show, the problem persists despite attempts to 'integrate a wider range of measures and experimental paradigms' (156). Current theories of consciousness already draw on a range of behavioural and neural data. Second, the problem only worsens as the range of behavioural and neural data widens. Each significant operational test or hypothesis about the neural basis of consciousness leads scientists to a new cluster of properties and mechanisms.

I do not think that Irvine has a compelling response to the second objection. What follows is my reconstruction of her arguments.[4] She claims that the failure of 'standard scientific methods', together with methods designed for consciousness science, 'suggests that there is a serious problem in consciousness science' (155). Here she may be shifting the burden of proof onto those who propose that a neglected or new method will address the multiple kinds problem. But, to establish her eliminativism, Irvine must explain why the default response to a 'serious' methodological predicament in consciousness science is to eliminate the concept of consciousness – rather than to look harder for a solution or come to terms with our ignorance. Might she be relying on a verificationist assumption that unanswerable questions are framed by meaningless concepts?

The motivations – both rational and non-rational – that bring scientists to eliminativism are notoriously murky.[5] Despite Irvine's attention to the 'standard practices found elsewhere in science', she does not cite any example in which scientists eliminate a concept because they confront the multiple kinds problem.[6] She even acknowledges that, in at least one example, scientists appear to have

[4] I draw together those aspects of her analysis that are related to the multiple kinds problem. Some aspects appear among what she calls the 'epistemological' and 'pragmatic' factors for eliminativism (§9.3.3-9.3.4); her evaluation of these factors pivots on how they cope with the discovery of multiple kinds.

[5] See, for instance, the considerations surveyed in Stich (1998), Chapter 1. These include normative 'principles of reasoning', 'personalities of the people involved', as well as 'social and political factors in the relevant scientific community or in the wider society in which the scientific community is embedded' (67).

For a subtle analysis of eliminativism in the context of inter-level and intra-level scientific reductions, see Wimsatt (1976). In the section entitled 'On the Eliminability of Levels and Generalized First-Person Perspectives', he uses this analysis to defend an anti-eliminativist stance on consciousness.

[6] She mentions two philosophers who propose eliminating concepts in cognitive science. However, both proposals have not been adopted by cognitive scientists and other philosophers. Machery (2009) argues that the concept of *concept* should be eliminated because it refers to three kinds of information, which can be dissociated. Griffiths (1997, 2004) makes a similar argument about the concept of *emotion*, which refers to many specific 'categories of psychological state and process'.

managed the multiple kinds discovered in their samples. Their putative solution is to develop a more complex, context-sensitive, classification.

This example comes from biological classification (158). In practice, scientists now associate the concept of species with a set of multiple kinds. According to Ereshefsky (1992, 1998), these kinds include biospecies, ecospecies, and phylospecies.[7] The first kind defines a species primarily as a lineage maintained by interbreeding, the second as a lineage maintained by ecological forces, and the third as a basal monophyletic lineage. Scientists tend to use the concept of species to refer specifically to one of these kinds. Which kind? That depends on the biological context in which the concept is used.

Even when the concept is not used to refer to one of these kinds, it remains useful. Brigandt (2003) explains why:

> Due to the overlap and reinforcement of the mechanisms generating species, the extension of different species definitions shows a good deal of overlap. The general species concept figures in theoretical generalizations and explanations across different branches of biology. (1313)

According to Brigandt, this creates two theoretical roles for the general concept. First, in 'many biological theories and branches', scientists can 'live with' the concept without further specifying the kind to which they refer (1313). Second, even in biological theories for which scientists need a more specific definition of species, the

[7] As Irvine notes, whether this pluralism should lead to eliminativism about the concept of species is a matter of controversy among some philosophers. She contrasts the pro-eliminativist view in Ereshefsky (1992) with the anti-eliminativist view in Brigandt (2003). The former has since moderated his view: Ereshefsky (2010a) suggests that we should retain the term 'species' and be realist about the individual taxa that biologists call 'species'. But he maintains that we should be sceptical of species as a biological category. I will examine biological classification in more detail in Chapter 5.

concept ‘sets the standards’ for what counts as an adequate definition. Both roles generate a ‘unifying effect’ among the different species definitions.

Can’t the same method be used in consciousness science? Suppose that scientists associate the concept of phenomenal consciousness with a family of related phenomena. Each phenomenon is a sub-type of experience related to one of the multiple kinds. Depending on their context, scientists can use the concept of phenomenal consciousness to refer to one sub-type of experience. Here is why Irvine rejects this more complex, context-sensitive, classification:

> While it is possible to label these different phenomena as subtypes of consciousness, …having two sets of labels of the same phenomena, one compris[ing] the vocabulary of perceptual and cognitive abilities, and one with the word ‘consciousness’ added, is of little practical or explanatory benefit. (154)

She repeatedly points out that the phenomena highlighted by current theories of consciousness can already be described with other scientific concepts, such as ‘different kinds of perception, different types of responses, different types of attention, different types of neural processing, and so on’ (88). Irvine claims that these descriptions are ‘perfectly adequate’. Similarly, putative mechanisms of consciousness are ‘easily characterised as (parts of) mechanisms of well-known cognitive and neural phenomena’ (117). They often ‘pick out scientific kinds already investigated in cognitive neuroscience’; in consciousness science, they end up ‘re-labelled in a confusing and unproductive way’ (118).

Much rests on Irvine’s judgement that this new classification is of ‘little practical or explanatory benefit’. How does she support it? First, she assesses its *explanatory*

benefits. According to Irvine, it is 'not a productive theoretical move' to classify well-known phenomena 'under the banner of consciousness research' (90). Their 'stark differences' stop them from supporting generalisations (159). They often have 'very little in common with each other' — not enough for scientists to identify 'a group of related properties that are easily associated with the personal level phenomenon of consciousness' (160).

Second, Irvine turns to its *practical costs*. She argues that the current labels used in consciousness science are 'unclear and leave out vital information' about the differences between various neural structures, which are better conveyed by the vocabulary of cognitive science (89). If scientists continue to use a single concept to refer to multiple kinds, this 'may lead to confusion' and debates in which 'all sides talk past each other' (162-3).

I am not persuaded by her arguments. First, at this stage in the development of consciousness science, it seems premature to assess the explanatory benefits of a new classification. Irvine is too severe when she claims that various phenomena are 'unceremoniously lumped together in consciousness science and expected to form a coherent, and scientifically interesting, whole' (117). The different neural structures already have something significant 'in common with each other'. As I emphasised in §3.4, they are systematically related to behaviour which we associate with consciousness at the 'personal level'. Beyond this, I do not see why she demands that properties highlighted by the more context-sensitive classification be 'easily associated' with consciousness at the 'personal level'. It is not even clear how she assesses this ease of association.

Other explanatory benefits may accrue later. With a classification that has sub-types of phenomenal consciousness, scientists may go on to discover more complex structures that involve the multiple kinds. Moreover, some of these kinds overlap because they are nested in each other. Brigandt's analysis of species suggests that, if there is an overlap between the kinds associated with a phenomenon, then at least some scientists can productively build theories of the phenomenon without further specifying a kind.

Second, the practical costs mentioned by Irvine can be managed by the new classification. The point of this classification – with its sub-types of phenomenal consciousness – is to demarcate differences between the relevant neural structures, while recognising that these structures are all systematically related to behaviour that we associate with consciousness. Irvine acknowledges that a basic taxonomy already helps scientists to disambiguate phenomenal and access consciousness, as well as creature and state consciousness (163).[8] But she denies that a 'more detailed taxonomy' will help with phenomenal consciousness itself: such a taxonomy will be 'a vast and complex one that simply relabels the phenomena identified in the cognitive sciences'. To her, this suggests that the concept of consciousness is 'inherently ambiguous' and 'simply unnecessary' to describe the phenomena.

I deny that her final inference is warranted. Why should the concept of consciousness be *inherently* ambiguous just because it is *currently* ambiguous? A 'vast and complex' classification may be exactly what scientists need to reduce the current ambiguity. Indeed, as she concedes, one job of scientific theorising is to

[8] For elaboration, see Irvine's earlier analysis of 'Taxonomies of Consciousness' in her §1.3.

supply, and then to refine, classifications. At this point, Irvine's claim that the new classification 'simply relabels' phenomena better described in operational and cognitive terms, and her conclusion that the concept of consciousness is 'simply unnecessary', only begs the question. Both depend on her tenuous claim that this new classification brings little explanatory benefit.

Interestingly, Irvine provides a plausible historical explanation for why the neural structures related to consciousness are, currently, better described by other concepts.

> The science of consciousness…has not been around for very long, so is not well entrenched. Consciousness was studied by a few psychologists and psychophysicists starting at the end of the nineteenth century, but behaviourism largely prevented much research being carried out on this phenomenon through the first half of the twentieth century. (161)

Even the rise of cognitivism in the 1960s did not rehabilitate consciousness as an 'appropriate' subject for many scientists (4). Instead, neural structures were defined and investigated by different research programmes, such as those on perception, learning, and attention. Only since the 1990s has consciousness science brought together some of these phenomena under a common research agenda — albeit with a classification that remains incomplete.[9]

Whether this classification ought to be improved or eliminated is an open question. As I have argued, Irvine's case for methodological eliminativism is unproven. In her preface, Irvine admits that reactions to her eliminativist stance seem to 'depend on

[9] In their pioneering article, Crick and Koch (1990) note that most work in cognitive science and the neurosciences then do not refer to consciousness, even though 'many would regard consciousness as the major puzzle confronting the neural view of the mind' (263). They attribute this 'remarkable' fact to the 'legacy of behaviorism' and the perceived lack of 'any useful way of approaching the problem'.

the background assumptions that are brought to the table' (v). Some regard her claims as 'fundamentally misguided', while others view them as 'entirely sensible and rather unsurprising'. For those who reject her stance, she hopes that the argument 'at least points to methodological problems in consciousness science that deserve more attention'.[10] She also hopes to spur 'more use of philosophy of science' in current debates about consciousness.

I heed two lessons from Irvine's arguments. First, she confirms the intractability of the multiple kinds problem in consciousness science. Why hasn't the impasse over phenomenal consciousness been resolved? It is not because scientists lack behavioural and neural data for framing hypotheses about the nature of phenomenal consciousness. Indeed, more data seem to exacerbate the multiple kinds problem: locating new clusters of properties and mechanisms leads to more rival kinds.

Second, Irvine indirectly points us to a neglected resource for addressing the multiple kinds problem. She mentions that biologists develop more a context-sensitive classification in order to manage the multiple kinds in their samples. Of course, her arguments mean to convince us that the science of consciousness is more problematic than the science of species – so problematic that this putative solution for species cannot work for consciousness. But I have shown that those

[10] This is reiterated in Irvine's conclusion, when she assesses her philosophical arguments: 'Even if these are not accepted as definitive reasons to eliminate "consciousness" from scientific discourse, they should at least illustrate the methodological problems that consciousness science currently faces, and highlight that they are more serious than currently acknowledged' (172).

My analysis concurs with the assessment in Kozuch (2013): 'While Irvine probably does not succeed in showing "there can be no science of consciousness"…, she does point out areas in consciousness research that do not yet seem to live up to standards usually at work in science' (2).

arguments are not compelling. This suggests that we ought to look more closely at how biologists refine their classification of species. In this respect at least, Irvine is right: we need more philosophy of science.

4.3 Prinz on moderate mysterianism

Prinz (2003) responds to the multiple kinds problem as it arises in the debate on artificial consciousness. When scientists build a theory of phenomenal consciousness, they face a choice between at least two levels: the levels of functional structure and biological structure. Theories built at different levels centre on different kinds of creatures. A kind defined by functional structure is distinct from one defined by biological structure. The former kind allows for the possibility of artificial consciousness, while the latter rules it out.

As I noted in §3.4, Prinz argues that scientific investigation into the nature of phenomenal consciousness cannot help in this debate: 'artificial experience eludes us because the basis of experience eludes us' (116). In particular, scientists cannot tell whether any of the human brain's biological properties is constitutive of consciousness. Prinz's cases rests primarily on what he calls the 'difference-maker problem' (121). This refers to the failure of experimental tests to dissociate the roles of functional and biological structures in human consciousness. It is an in-principle difficulty, which arises even if we suppose that scientists can replace a human subject's brain cells with silicon chips without changing his psychological profile.

How does Prinz respond to this epistemological challenge? Unlike Irvine, he does not suggest that scientists eliminate the concept of consciousness. Rather he recommends a 'sceptical position' on artificial consciousness:

> The problem isn't that it would be impossible to create a conscious computer. The problem is that we cannot know whether it is possible. There are principled reasons for thinking that we wouldn't ever be able to confirm that allegedly conscious computers were conscious. The proper stance on computational consciousness is agnosticism. (111)

This agnosticism reflects what Prinz believes to be our irremediable ignorance of biology's role in consciousness. It is a variant of mysterianism about consciousness. However, Prinz's mysterianism is not wholly defeatist. He argues that scientists can still formulate 'concrete empirically grounded theories of consciousness' (111). They bring us 'very close' to understanding the basis of consciousness, though we will never 'get all the way there'. In this respect, his mysterianism is more 'level-headed' than others (112).

To clarify this position, Prinz contrasts his moderate mysterianism with that defended in McGinn (1991, 1999). McGinn assumes that the mind-body problem needs to be solved by an explanation of how the brain causes, produces, or gives rise to phenomenal consciousness:

> How is it possible for conscious states to depend upon brain states? How can technicolour phenomenology arise from soggy grey matter?...How could the aggregation of millions of individually insentient neurons generate subjective awareness? (1989: 349)[11]

[11] See also McGinn (1989), 353: 'There just *has* to be some explanation for how brains subserve minds…Brain states cause conscious states, we know, and this causal nexus must proceed through necessary connections of some kind—the kind that would make the nexus intelligible if they were understood.'

According to him, such an explanation is not forthcoming because human beings suffer from a cognitive deficit. In principle, they cannot detect the physical properties of the brain that generate phenomenal consciousness.

Prinz rejects both claims. First, he agrees with Block and Papineau that the mind-body problem is to be solved by establishing an identity between some of the brain's properties and phenomenal consciousness. This identity is not itself something to be explained in terms of a causal or mechanistic relation between the brain's properties and phenomenal consciousness. It does not make sense to ask how one half of an identity causes, produces, or gives rise to its other half. Rather we accept the identity if it offers us 'explanatory buying power' over the systematic relations associated with phenomenal consciousness (116).[12]

Second, Prinz denies that there are any 'intractably unobservable properties' in the human brain. His mysterianism is not based on our inability to detect some of the brain's properties. Rather it arises from our inability to discover if the brain's biological properties are constitutive of phenomenal consciousness. Where McGinn posits a cognitive deficit, Prinz identifies an epistemological difficulty:

> The method of difference-makers seems to be the only way to find out what levels matter and, when applied to the level of neurons (or lower), we simply cannot tell whether the transformation makes a difference. There is no way to rule out a biological contribution to consciousness. (130)

This limitation lies, ultimately, in our 'ability to measure consciousness'. Because we measure changes in consciousness through behaviour, our experimental tests cannot

[12] See the references cited in my note 6 from Chapter 3. McLaughlin (2003) makes the same criticism of McGinn's challenge.

dissociate the role of functional properties from that of biological properties which realise them.

In Prinz's thought experiment, the artificial duplicate behaves exactly as conscious humans do. He automatically passes any experimental tests for consciousness. If we insist that his behaviour expresses consciousness, we would only beg the question against those who believe that some of the brain's biological properties are constitutive of phenomenal consciousness.

What about the positive, 'level-headed', aspect of Prinz's mysterianism? This is yet another point of difference from McGinn. According to Prinz, we can 'make considerable progress' in consciousness science (116). By developing a theory of how phenomenal consciousness arises in human beings, we can already 'learn a lot about the material basis of consciousness', though 'not everything'. Indeed, Prinz goes on to propose what he takes to be the 'best current empirically informed' theory of human consciousness (117).

I shall sketch Prinz's Attended Intermediate-level Representations theory because it contributes to my evaluation of his mysterianism.[13] Prinz emphasises that humans experience the world through our senses 'from a particular vantage point' (118). So, in his theory, the contents of human consciousness are both perceptual and perspectival. The constitutive conditions of our consciousness include attention, access to working memory, rational deliberation, and deliberate control of action. At the psychological level, consciousness arises in humans when we attend to

[13] I draw on Section III in Prinz (2003) and the update in Prinz (2005). The most detailed account of his AIR theory is in Prinz (2012).

phenomena in such a way that perceptual states become available for remembering, reasoning, and responding.

This theory straddles the algorithmic and implementation levels too. It offers a hybrid 'neurofunctional' account of human consciousness: consciousness is identified with 'a functional role that is implemented by mechanisms specifiable in the language of computational neuroscience' (2005: 390). At the algorithmic level, consciousness relates to the processing of intermediate-level representations in our perceptual systems. These representations are 'vantage-point specific and coherent' (2003: 119); they are distinct from lower-level representations which are too local to be coherent, and higher-level representations which are too abstract to preserve perspective. At the implementation level, consciousness normally involves a neural circuit of 'perceptual centres in temporal cortex, attentional centres in parietal cortex, and working memory centres in frontal cortex' (119).

How does Prinz interpret this theory? To him, it describes the 'physical basis of consciousness in us' (130). This basis consists of the 'material conditions that make consciousness possible in us' (131). However, the theory does not list the constitutive conditions of phenomenal consciousness; we will 'never get all the way' to the nature of consciousness.[14] Instead, the result is something more curious:

> In sum, we can identify properties that are necessary for consciousness (e.g., attention and perception) and properties that are sufficient (the biological substrates of attention and perception in us), but none that are definitely both necessary and sufficient. This leaves us with a degree of mystery… (131)

[14] See also Prinz (2005), 392: 'We will be able to discover conditions that are sufficient for consciousness, but we will never have a perfectly settled list of the conditions necessary for consciousness.'

In particular, we can never know if our artificial duplicates have experience or not. Although they possess properties that we think are necessary for phenomenal consciousness, they lack properties that we think are sufficient.

Let me translate Prinz's result into more familiar terms. Creatures that lack the necessary properties (at the level of functional structure) are not conscious, while creatures that possess the sufficient properties (at the level of biological structure) are conscious. Artificial duplicates that lack the sufficient properties but possess the necessary ones are neither definitely conscious nor definitely not conscious; they are borderline conscious. Since his theory does not say which properties are both necessary and sufficient, he has no way to classify the artificial duplicates more definitely. That is consistent with Prinz's mysterianism.

Is this modest mysterianism defensible? Prinz denies that it involves any contradiction. He compares it to ordinary situations in which we 'know necessary properties and sufficient properties without knowing necessary and sufficient properties' (111). In Prinz's example, a person knows that drinking at least one teaspoon of alcohol is necessary for intoxication, and that drinking two litres is sufficient for intoxication. Yet he can know this without knowing the quantity that is both necessary and sufficient. This comparison suggests that Prinz's curious result about phenomenal consciousness is neither contradictory nor unintelligible. But it does not support his conclusion that this result is the best we can hope for. Significantly, he fails to mention any scientific situation in which we are content with this anomalous state of knowledge about a natural phenomenon.

I shall evaluate the negative aspect of Prinz's position, then its positive aspect. To support his mysterianism about artificial consciousness, Prinz cites the in-principle failure of experimental tests to tell us whether biological properties are constitutive of phenomenal consciousness. So his mysterian conclusion stands only if testing dissociations is the sole method for solving the multiple kinds problem. What, in turn, supports this assumption? Prinz never says. In the above passage, he claims only that the method of difference-makers '*seems* to be the only way to find out what levels matter' (130, with my italics).

I do not think that this assumption is warranted. Prinz does not explain why inference to the best explanation cannot solve the multiple kinds problem. Elsewhere, when he discusses the epistemological problem of other minds, Prinz offers a clue to why he slights inference to the best explanation as a way to justify attributions of consciousness:

> I think this strategy works for mentality in general, but not necessarily for consciousness. We do not know enough about the causal role of consciousness to know what behaviours can be explained without it. (127)

But this assessment is not backed by any argument. Earlier, when I discussed the theoretical impasse on artificial consciousness, I proved only that the epistemological norms currently associated with theoretical identities cannot help with the multiple kinds problem.

Prinz does not rule out a neglected or new method for solving the multiple kinds problem. I note that, unlike McGinn, he concedes that the epistemological difficulty in the impasse on artificial consciousness 'has little to do with human cognitive

limitations' (116). So he cannot appeal, generally, to any cognitive limitation to rule out a neglected or new method. Moreover, unlike Irvine, Prinz does not see that biologists who investigate that nature of species have coped with the discovery of multiple kinds. So he has not tried to show that their method will fail for scientists who investigate the possibility of artificial consciousness.

Perhaps Prinz means to shift the burden of proof onto those who place their hopes on a neglected or new method. He would not be wrong to issue this challenge. Those who believe – as I do – that the multiple kinds problem can be solved need to explain how scientists can manage the multiple kinds which they discover in their sample of conscious creatures. And those who believe – as I do – that inference to the best explanation can help in this epistemological predicament need to specify the relevant norms. However, accepting the challenge only implies this: we *do not* know that artificial consciousness is possible, and we *do not* know how we can know that artificial consciousness is possible. In my judgement, this is quite far from justifying Prinz's pessimistic conclusion that we 'cannot know' that artificial consciousness is possible, and that we 'wouldn't ever be able' to discover that computers are conscious.

In its positive aspect, Prinz's position on consciousness science is less pessimistic than Irvine's. It therefore appears to be more defensible. Unlike Irvine, he does not recommend that scientists eliminate the concept of consciousness and disavow all talk of consciousness. Instead, he believes that science can make 'considerable progress' toward a theory of human consciousness. By outlining the Attended Intermediate-level Representation theory, he seeks to convince us that 'enormous

progress' has already been made (2005: 392).

Yet I want to cast some doubt on how stable his moderate position is. As I have demonstrated, the multiple kinds problem arises in both the debate on the neural structure of phenomenal consciousness and the debate on the possibility of artificial consciousness. So Prinz faces an *ad hominem* challenge. If he is committed to agnosticism about the levels which are constitutive of consciousness, why isn't he also committed to agnosticism about the structures which are constitutive of consciousness in humans?

Consider the role of attention in Prinz's theory. Why does he believe that attention is necessary for human consciousness?

> In characterizing the psychological profile level, we were implicitly using something like the method of difference-makers but applied at a single level…Attention seems, in anecdote and experiment, necessary for consciousness. (122)

Prinz (2003, 2005) cites the inattentional blindness experiments conducted by Mack and Rock (1998). When the subjects' attention is occupied by a task, many of them do not notice stimuli displayed briefly and unexpectedly in a central spot of their visual field. They claim not to be conscious of the stimuli. According to Mack and Rock, this suggests that conscious perception requires attention.

In §3.4, I noted the views of theorists who contest this interpretation of the experimental data about attention. They argue that such experiments test for reportability rather than consciousness. According to these theorists, consciousness ought not to be conflated with attention and reportability. In their rival

interpretation, the subjects in Mack and Rock's experiments may consciously perceive the unexpected object (Block 2001, 2007a, 2008; Dretske 2004, 2007). However, without due attention, what they perceive is not sufficiently conceptualised and made accessible for reporting. As Block (2008) points out, it is controversial whether the subjects are suffering from 'inattentional blindness or inattentional inaccessibility' (296).

Both sides in this impasse operationalise consciousness differently. So neither side can apply the 'method of difference-makers' without begging the question against the other side. In his arguments, Prinz focuses on the failure of experimental tests to tell us if biological properties make a constitutive contribution to consciousness. Here we face the failure of experimental tests to show us if attention makes a constitutive contribution. Again, this epistemological difficulty reflects a limitation in our ability to measure consciousness. Subjective reports of experience are only possible if the relevant content is attended to and made accessible for reporting. Therefore, as Block (2008) emphasises, subjective reports cannot test whether attention and accessibility are constitutive of consciousness.

Prinz does concede 'some wiggle room' to opponents of his theory:

> A sceptic can claim that the phenomenology remains constant during inattention, but loses access to the cognitive centres that are involved in reporting. This scepticism has a kind of stubborn intractability that might ground a mysterian conclusion in its own right. (123)

If this conclusion stands, then scientists have poor prospects for developing a theory of the 'physical basis of consciousness in us'. Prinz would have to concede the

positive aspect of his position. His mysterianism would not really be more 'level-headed' than McGinn's. In practical terms, it might not be so different from Irvine's eliminativism. Therefore, to maintain the viability of his modest mysterianism, Prinz must deny that the multiple kinds problem reported by Irvine is as serious as the one which he finds in the debate on artificial consciousness.

How does he do so? Prinz argues that there is no intuitive or theoretical reason to believe that consciousness is possible without attention and deliberation: 'While we cannot rule out consciousness in creatures whose behaviour is driven by radically different kinds of mechanisms, we have no reason to think such creatures are conscious' (122). First, Prinz compares this possibility to the idea that paramecia, clams, and insects may be conscious. Because they cannot deliberate, these creatures respond to the environment through immediate reflexes. Because they cannot selectively attend to the environment, they rely on limited perceptual systems tuned to highly specific inputs. He finds 'little compulsion' to attribute consciousness to such creatures (122). Later, he construes this move as resisting 'weird sceptical intuitions' (126).[15]

Second, Prinz claims that his theory of human consciousness offers the better overall explanation of anecdotal and experimental data about attention. His theory takes 'seriously' the first-person reports in the inattentional blindness experiments, where subjects claim not to see the unattended stimulus (124). It also accounts for an anecdote from Block (1995), in which a person suddenly realises that a drill has been

[15] Prinz's opponents are thereby likened to sceptics who take seriously the possibility that neurons in a Petri dish have 'feels' (126). In contrast, his proposal that biological properties may be necessary to consciousness is 'not an example of wild, sceptical standard-raising' (127). Thus he need not adopt a 'radical form of scepticism' (130).

digging nearby while he was engaged in intense conversation. According to Prinz, this person is conscious of the drill throughout his conversation because he has been partially attending to it. In contrast, attention is 'more completely consumed' during the inattentional blindness experiments (124). Finally, his theory avoids postulating *ad hoc* that subjects in the inattentional blindness experiments have forgotten what they once consciously perceived.

I do not find these arguments compelling. First, not everyone shares Prinz's intuition about creatures which lack attention and deliberation. He may be right about paramecia, clams, and insects. However, as I noted in §3.4, there remain active scientific debates on consciousness in human foetuses and hermit crabs. These creatures also rely on more limited perceptual systems and reflexes to navigate their environment. Yet their behaviour and neurophysiology is similar enough to conscious humans that our operational criteria classify them as borderline conscious. So theorists who doubt that attention and availability for report are constitutive of consciousness need not be drawing on what Prinz calls 'weird sceptical intuitions'.

Second, other theorists interpret the anecdotal and experimental data about attention differently. How do they thereby deny that attention is constitutive of consciousness? They cite other first-person reports in which subjects claim that they are aware of more than they can attend to or report on (Pashler 1998: 9). They point to experimental data which suggest that 'phenomenal consciousness overflows cognitive accessibility' (Block 2007a: 481; 2008). Such theorists can agree with Prinz that shifts of attention explain the contrast between Block's anecdote and the inattentional blindness experiments. But, in their rival interpretation, these shifts

only affect whether subjects can report their consciousness of visual stimuli; attention does not constitute this consciousness. Moreover, they need not suppose that subjects in the inattentional blindness experiments have forgotten what they consciously perceived.[16] Indeed, their interpretation implies otherwise: what the subjects consciously perceived has yet to enter working memory.

My aim is not to settle this continuing dispute on the role of attention.[17] Instead, I mean to show that Prinz's theory of phenomenal consciousness in humans is not untroubled by the multiple kinds problem. He has to explain how he can resist a mysterian conclusion in the debate on phenomenal consciousness – even while he promotes a mysterian conclusion when the same problem crops up in the debate on artificial consciousness. Until he does so, Prinz's moderate mysterianism is not a stable position. I draw a lesson: we need a more unified approach to the multiple kinds problem, which attends to its ramifications in the debates on phenomenal consciousness, borderline consciousness, and artificial consciousness.

4.4 Papineau on referential indeterminacy

According to Papineau (2003), the multiple kinds problem poses a challenge

[16] Prinz (2005) appears to concede this point. Here he presents his interpretation more equivocally: 'It is possible that these subjects consciously perceived the gorilla and simply didn't categorize it; but it is equally possible that they didn't perceive it consciously at all' (386-7).

He adds that his interpretation is 'most consistent' with experiments on neglect, where subjects report no conscious perception of stimuli that they cannot attend to. But those theorists who reject his view of inattentional blindness will similarly want to re-interpret data from the neglect experiments.

[17] De Brigard and Prinz (2010) assess the scientific evidence on both sides, and add more cautiously that 'the evidence alleging dissociations between consciousness and attention is not decisive' (51). See also Koch and Tsuchiya (2007, 2012). I think that their impasse reflects the intractability which I analysed in §3.4 (Figure §3.4).

throughout consciousness science.[18] Like Prinz, Papineau concedes that this science can reveal 'many interesting and indeed surprising things' about human consciousness (208). However, it faces a problem when it looks for the 'material essence' of a phenomenal property:

> Yet science can only take us so far. It may be able to rule out certain candidates as the referents of phenomenal concepts, often surprisingly so. But the trouble is that, even after science has narrowed things down in this way, there will always remain a plurality of [material properties] M in play for any given phenomenal [property] Ø, and no further means of deciding which is the real referent of the phenomenal concept. (208-9)

For Papineau, 'material properties' refer to (a) strictly physical properties, which are the first-order properties studied by the physical sciences, and (b) properties that supervene on strictly physical properties. As this passage indicates, he believes that scientists will find 'too many candidates' for every phenomenal property – and, therefore, too many kinds of creatures associated with each phenomenal property.

I shall focus on Papineau's discussion of consciousness-as-such. He construes this as a determinable phenomenal state: its determinates are more specific phenomenal states, such as seeing something red, feeling pain, and hearing middle C. To wonder if a creature is conscious-as-such is his way of asking if it is 'conscious at all'.[19] How do scientists investigate consciousness-as-such? In Papineau's account of 'standard research methodology', scientists look for a material property of the brain that is correlated with human reports of phenomenal consciousness. This property must be

[18] The analysis in Papineau (2003) is largely condensed from Papineau (2002), ch. 7. I will also draw on points in Papineau (1993, 1996), where he first proposes that the concept of consciousness is vague. In that earlier analysis, he does not clearly distinguish the vagueness of phenomenal concepts from that of psychological ones; see note 12 in Papineau (2002), 197.
[19] The concept of consciousness-as-such is analysed in Papineau (2002), §7.5.

'common to all those cases where humans report they are phenomenally conscious, and absent whenever they deny this' (218).[20]

He argues that scientists will 'inevitably find a variety of material candidates which fit perfectly' (218). First, aside from strictly physical properties, there are bound to be higher-level structural properties. These may be defined 'physiologically, neuronally, and computationally' – so long as they capture some level of causal organisation realised by the strictly physical properties. Second, there will be broadly defined properties in addition to narrowly defined ones. The former include 'causal or historical relations to your environment', while the latter are limited to 'current matters inside your skin' (213). Third, there will be properties about which the subject has formed a judgement, as well as those which involve 'only lower-order attentional properties' (218). So which is *the* property that defines the kind formed by all conscious creatures?

Papineau (2002) recognises that this problem about the 'material essence' of phenomenal consciousness is tied to the status of borderline conscious creatures – that is, 'whether other animals, or future computer robots, or possible extraterrestrials, have experiences like phenomenal pain' (177). He highlights the case of a 'silicon doppelgänger'. It resembles Prinz's artificial duplicate since it shares

[20] I see two problems with this account. (a) The material property should also be causally or constitutively related to human behaviour that we ordinarily associate with consciousness. (b) Although scientists rely on human subjects to report their conscious states, they do not normally ask these subjects to affirm or deny that they are (or were?) conscious. Besides first-person reports, they also use third-person attributions of consciousness to normal human subjects. I put aside these problems since they do not affect Papineau's core arguments.

The same problems recur in Papineau (2002), 187, when he describes the 'basic strategy' of 'standard research into phenomenal consciousness'. Here, in note 5, he claims that third-person attributions of phenomenal states are 'posterior to empirical research into phenomenal concepts, not a starting point' (184). Yet these attributions do serve as defeasible evidence in this research – even if Papineau is right that they are not 'on a par with subjects' self-reports as evidence'.

all the higher-level structural properties of a conscious human, but is composed of silicon rather than carbon. If consciousness is strictly physical in nature, then this doppelgänger is not conscious. If consciousness is computational in nature, then this doppelgänger is conscious.

He claims that standard research methodology is 'impotent' when faced with this dilemma (2003: 208). Why? According to Papineau (2002), here is how 'normal scientific research' works:

> When scientists seek to uncover the nature of some natural kind, like water or temperature, say, they will typically begin with some direct observational judgements that certain things are water, certain things are hotter than others, and so on. And then they will seek to construct a theory which will identify further scientifically interesting properties which are common to these observationally identified samples. (182-3)

So, in general, scientists start to theorise about the nature of a phenomenon by making observations about obvious cases of the phenomenon. In consciousness science, first-person reports by human subjects serve as 'observation reports' about their own experience. They supply an 'initial sample' or 'observational data-base' of cases to spark scientific theorising.[21]

During this theorising, scientists will stumble on multiple kinds in their samples of conscious humans. In response, the 'obvious strategy' is to expand their samples, in search of 'further dissociative data' (189). To dissociate the roles of strictly physical properties and higher-level structural properties, scientists can focus on a silicon doppelgänger – a creature that has the same structural properties as a conscious

[21] These terms are used, respectively, in Papineau (2002), 184, and Papineau (2003), 207.

human but lacks the strictly physical ones. Yet Papineau argues that this strategy is bound to fail: scientists have no way to test if the doppelgänger is conscious. After all, the 'canonical method' for telling whether a subject is in a conscious state is to rely on his report (2003: 210). We already know that the doppelgänger behaves exactly as a conscious human does, which includes making the same reports. It thereby passes every behavioural test for detecting conscious states. However, we cannot take its reports 'at face value' without begging the question of whether it is in a conscious state (211).

The result is an epistemological limitation in consciousness science. Scientists cannot 'pinpoint' a property as the 'material essence' of any conscious state, including consciousness-as-such (219). How do we make sense of this 'in principle' limitation? Papineau contrasts two interpretations. In the first interpretation, we suffer from *irremediable ignorance* about the nature of phenomenal consciousness. Scientists will never discover whether silicon doppelgängers and other borderline conscious creatures are conscious. At most, they can identify a set of material properties that are correlated with phenomenal consciousness in humans.

The second interpretation, which Papineau favours, blames *referential indeterminacy*. We are not committed to any facts about phenomenal consciousness that 'lie beyond our epistemological access' (215). Rather our phenomenal concepts – including the concept of consciousness-as-such – are vague in a special sense. They suffer from this referential indeterminacy: 'it is indeterminate what *kind* of property they refer to', whether the property be strictly physical or structural, narrowly or

broadly defined, etc.[22] According to Papineau (2002), this means that there are no 'definite facts of the matter' about the applicability of phenomenal concepts to borderline cases (178). That, in turn, explains why scientists cannot give 'definite answers' about phenomenal consciousness in non-human creatures such as octopuses, advanced computer robots, and Proxima Centaurians.

Let me reconstruct Papineau's arguments for adopting this interpretation.[23] He begins by equalising the burden of proof on both interpretations. On one hand, he casts some doubt on the diagnosis of irremediable ignorance:

> There would indeed be something puzzling if…there were definite facts about consciousness that no amount of empirical research could possibly uncover. If there were such definite but inaccessible facts, then surely there ought to be some explanation of why we can't find out about them…It is not as if the facts of consciousness are too far away, or too small for our instruments, or anything like that. (2002: 197)

Papineau does not clarify how much significance he attaches to this doubt. Admittedly, it offers no positive support for referential indeterminacy. I think it shows, however, that positing irremediable ignorance is not without cost. We would still need to explain why the world contains these special material facts that scientists can never discover – even after they have discovered all other material facts. Compared to referential indeterminacy, irremediable ignorance is not obviously a simpler explanation for the epistemological limitation faced by scientists.

[22] Papineau (2003) contrasts this dimension of vagueness with another in which there is 'no sharp cut-off point' and indeterminacy about 'where to divide some continuum' (215). In Papineau (1993), §4.8, he argues that the latter dimension also applies to the concept of consciousness: any structural complexity is 'likely to come in degrees, with no clear cut-off point beyond which you definitely qualify as conscious, and before which you don't' (124).

[23] Here I draw together arguments from Papineau (2002, 2003). His arguments are critically analysed in Bermudez (2004) and Antony (2006a, 2006b). I learnt from their analyses, though my focus is different. I will examine some arguments that they neglect, and emphasise the connections between Papineau's arguments.

On the other hand, Papineau acknowledges the logical distance from an epistemological limitation to referential indeterminacy. His claim of referential indeterminacy in phenomenal concepts would be 'uncomfortably ad hoc' if it were motivated only by the fact that 'we can't find the answers' to the questions posed about phenomenal consciousness (2003: 215). He cannot infer directly from our inability to answer a question to the indefiniteness of the question, and then to the indeterminacy of concepts used to raise the question. Such leaps are open to a charge of verificationism: they suggest that 'our concepts cannot lay claim to matters beyond our epistemological access'.

Instead of this verificationist inference, Papineau cites 'independent reasons' for favouring referential indeterminacy, which are based on the 'special constitution' of phenomenal concepts (2002: 178). Here is a summary of these reasons:

> I take phenomenal concepts to be basic unstructured terms, whose referential powers depend on the information they (are designed to) causally track (Papineau 2002, ch. 4). At bottom, I take this to be a matter of tracking the experiences of human beings. But experiences are material, if they are anything, and, as we have seen, there are various different material properties, physical and structural, broad and narrow, which will serve to make the same classifications among normal human beings. So there is no reason to suppose that there is anything in the semantic constitution of phenomenal concepts to focus them precisely on one rather than another of these humanly coextensive material properties… (2003: 216)

I do not find Papineau's reasons compelling. But before I can evaluate them, I need to explicate them in relation to his account of phenomenal concepts. What are the relevant aspects of these concepts? And how do they support his conclusion that these concepts have indeterminate reference?

In his summary, Papineau refers obliquely to the quotational and causal/teleological aspects of phenomenal concepts. First, the quotational aspect explains why he claims that phenomenal concepts are 'basic' in structure. According to Papineau (2003), these concepts are not connected *a priori* to material concepts (206). They take the demonstrative form 'that experience: ——'. We fill this gap with a current experience or its recreation in our imagination, and thereby attain a concept referring to that specific type of experience. In this sense, an exemplar of that type of experience is 'quoted' within the concept.[24] Suppose that someone visually imagines a red surface. When he uses the operator 'that experience: ——' to quote his imaginative state, he forms a phenomenal concept referring to the experience of seeing something as red.

Papineau (2002) extends this 'quotational model' from determinate phenomenal states to the determinable state of consciousness-as-such. He speculates that we either introspectively classify some state as conscious-as-such or imaginatively recreate the state of consciousness-as-such. We might do so by 'thinking first about some determinate mode of phenomenal consciousness, and then ignoring its special features' (186). Eventually we can quote the state of consciousness-as-such to form the concept of consciousness-as-such.[25]

[24] Papineau argues that this quotational aspect produces an 'illusion of distinctness', which drives us towards dualism (217). In contrast to phenomenal concepts, material concepts that refer to experiences do not quote the experiences themselves. So we feel that they 'leave out' the experiences, and infer that the phenomenal concepts must refer to some distinct non-material properties. For details of this diagnosis, see Papineau (2002), §6.5.

[25] In subsequent work, he rejects this proposal: it threatens circularity when we ask about the concept of experience used to form the operator 'that experience: ——'. See Papineau (2006), 121. Here Papineau also modifies the quotational model. He no longer treats phenomenal concepts as demonstratives, but keeps the 'crucial feature' that 'phenomenal references to an experience will

Second, the causal/teleological aspect explains how the 'referential powers' of phenomenal concepts are determined. Papineau (2003) refers us to his analysis in Papineau (2002). In that earlier work, he appeals generally to causal or teleosemantic theories of representation. Causal theories, in their basic form, propose that a concept refers to 'that entity which normally causes classificatory uses of that concept' (113). Teleosemantic theories suggest that a concept refers to the entity 'which it is the biological function of the concept to track'. Papineau prefers the latter because biological malfunctions promise to explain better why concepts can sometimes be used to misrepresent the world.[26]

Papineau assumes that one of these theories can explain the referential powers of phenomenal concepts. So these concepts refer to experiences either because they are causally related to experiences, or because it is their biological function to track experiences. He suggests that, for both the operator 'the experience: ——' and any specific psychological state that fills its gap, there will viable stories about their contribution to the causal relations or biological functions of the resulting phenomenal concept (117). He does not supply more detail about these stories: 'I shall simply proceed on the assumption that they are available.' And that is all he says to support his above claim that, 'at bottom', phenomenal concepts are used for 'tracking' experiences in humans.

Next, let me evaluate how far these aspects of phenomenal concepts support the claim that they have indeterminate reference. Here is how Papineau (2003) deploys

---

deploy an instance of that experience' (123). I put aside this modification: I do not see that it affects his arguments for referential indeterminacy.

[26] See also his defence of teleosemantics in Papineau (1993), ch. 3.

the quotational aspect of the concept of consciousness-as-such.

> We somehow construct this concept by taking exemplars of human conscious states, and use the resulting concept to classify humans depending on whether or not they are in some such conscious state or not. There is no reason to suppose the concept so constituted is precisely enough focused to discriminate between the physical and structural properties shared by conscious humans, or between the broad and narrow properties, or between Higher-Order and attentional properties. (219)

I do not see that this argument offers a positive reason to believe in the concept's referential indeterminacy. At most, it shows that the quotational structure of this concept allows for referential indeterminacy.[27] Papineau is right in this sense: nothing that we know so far about the operator 'that experience: ——' guarantees that this concept will refer determinately to one material property. When we quote our own experiences as exemplars, we are not aware of this quotation picking out just one kind of material property shared by these experiences.

What about the causal/teleological aspect? Papineau (2002) makes the following terse argument:

> any causal or teleosemantic account will leave it indeterminate exactly which of the correlated material candidates any given phenomenal concept refers to. For all the correlated material candidates will figure equivalently in the characteristic causes or biological functions of the relevant phenomenal judgements, and so causal or teleosemantic considerations will fail to pick out one material candidate rather than another as the referent. (198)

[27] I think the weakness of his argument accounts for his locution: 'There is no reason to suppose…' This phrase also appears at the end of the summary from Papineau (2003). In fact, it appears repeatedly in Papineau (2002), ch. 7; see the instances at 200, 215, and 229.

Perhaps, by using this phrase, he means to imply that the burden of proof is on someone who believes that phenomenal concepts have determinate reference. But this implication would be dialectically problematic. Papineau has yet to offer any argument that shifts the burden of proof in this direction.

But, crucially, Papineau supplies no evidence for his claim that all the correlated material candidates 'figure equivalently' in the characteristic causes or biological functions associated with the concept of consciousness-as-such. For two reasons, I doubt that this claim could make a good argument for the concept's referential indeterminacy. First, to establish this claim, Papineau would have to define the concept's characteristic causes or biological functions. But he can hardly define these without begging the question about the concept's referential powers. At this stage of semantic theorising, any hypothesis about a concept's semantic mechanisms must draw on commitments about its semantic values.

Second, even if Papineau's claim is true, one might take it to cast doubt on his application of causal and teleosemantic theories of reference to phenomenal concepts. In fact, as Antony (2006a) points out, causal and teleosemantic mechanisms are 'notorious for generating too many semantic values' for concepts (279). So their inability to specify determinate referents for phenomenal concepts may not even reflect anything distinctive about such concepts. Papineau (2002) concedes that this topic raises 'both points of detail and general issues in the theory of reference', which he has to put aside (117).[28] In my judgement, our grasp of semantic mechanisms is too speculative to warrant Papineau's conclusion about the referential indeterminacy of all phenomenal concepts.

[28] This concession does not jibe with the strength of his later claim: 'If we refer back to our earlier analysis of phenomenal concepts, we can see that there is nothing in their semantic workings that could possibly ensure that they refer to one rather than another of the material properties which are characteristically present when normal humans report that they are phenomenally seeing something red or are in pain' (2002, 198).

Compare the weaker claim made in Papineau (1996) about the concept of consciousness: 'it seems to me important to consider whether the absence of any suitable physical property is not at least *consistent* with everything else we know about consciousness' (695).

To support his conclusion, Papineau does invoke the following scenario. Suppose that we use one of the correlated material candidates to classify normal humans. Whichever of the material candidates we choose, we would divide normal humans into the same two groups. Why? Because these material candidates are either all present or all absent in normal humans. From this fact about classification, Papineau makes a dubious inference: 'Since the same categorization of human beings would result in any case, we can conclude that phenomenal concepts refer indeterminately to any of those material properties' (199). If we glance back at the summary from Papineau (2003), this consideration is cited as though it is decisive.

Yet I do not think that this consideration has much force. For it is not obvious that the reference of phenomenal concepts is constrained by the classifications which we can use them to make. It is not true, in general, that the reference of concepts is constrained in this way.[29] Moreover, Papineau appears to be relying on a stronger, even less plausible, assumption about phenomenal concepts: that their reference is constrained by the classifications which we *currently,* or *commonly*, use them to make. Elsewhere he disavows the verificationist claim that concepts 'cannot lay claim to matters beyond our epistemological access'. But here he seems to assume – without argument – that phenomenal concepts cannot lay claim to uses beyond our current, or common, classifications.

Might an analogy help at this point? Papineau (2002) invites us to think of phenomenal concepts as 'crude tools' (199). These concepts have 'no theoretical

[29] A trivial example: the concept *being tall* does not refer indeterminately to having above-average height and having above-average femur length, although these material properties are either both present or both absent in normal humans.

articulation' relating them to just one kind of material property. According to Papineau's analogy, it is not their job to track material properties so precisely. Rather they 'do their job adequately' if they 'enable us to respond to the packages of co-occurring material properties associated with experiences'. More specifically, the 'task' for the concept of consciousness-as-such is 'simply' to classify humans into those with 'the kind of cerebral states that our phenomenal concepts enable us to recognise first-personally' and those without such states (215). It fulfils this task 'effectively enough' so long as the package of co-occurring material properties correlates with human reports of phenomena consciousness.

I think this analogy would illuminate the role of phenomenal concepts in our conceptual toolbox if we are already convinced that they have indeterminate reference. But the analogy offers us no independent reason to believe so. It is premised on the claim that phenomenal concepts lack the necessary theoretical articulation. It thereby begs the question against anyone who believes that phenomenal concepts refer to natural kinds. If the concept of consciousness-as-such refers to a natural kind, then we should defer to scientists to work out its theoretical articulation, rather than assume that it lacks any articulation in one way or another.

So far, my verdict is that Papineau's case for referential indeterminacy in the concept of consciousness-as-such is not proven. His arguments are, as he promises, based on independent claims about the semantic workings of phenomenal concepts. But they show, at most, that what we now know about phenomenal concepts is compatible with their referential indeterminacy. His arguments are variously weakened by: a speculative appeal to semantic mechanisms, an unexplained restriction of our

phenomenal concepts to their current classificatory uses, and a question-begging analogy. Papineau's claims about the semantic mechanisms, classificatory uses, and crude tasks of phenomenal concepts fit well together. They produce a picture of what the concepts might look like with indeterminacy. But these claims do not add up to compelling arguments for indeterminacy.

Now I return to a more fundamental premise of Papineau's case. I want to cast doubt on his claim to have found an 'in principle' epistemological limitation in consciousness science. This is the starting point shared by both interpretations, whether irremediable ignorance or referential indeterminacy. By challenging it, I will point us toward a more promising direction.

Why does Papineau think that science is 'impotent' when scientists discovers more than one kind in their samples of conscious humans? Because, like Prinz, he assumes that testing for dissociations is the only scientific method for solving the multiple kinds problem. Repeatedly Papineau emphasises the search for a 'test', 'test creature', or 'test case' to help pick out one property from a package of co-occurring material properties (2002: 191, 192, 195; 2003: 213). This, in turn, leads to his disappointment that experimental tests on silicon doppelgängers cannot work. These tests cannot dissociate the contribution of strictly physical properties and higher-level structural properties to phenomenal consciousness. They thereby fail to solve the multiple kinds problem in consciousness science.

From this failure, Papineau infers a more sweeping epistemological limitation.[30] Papineau (2002) claims that science 'can provide far fewer answers' about the material referents of phenomenal concepts 'than many people suppose' (176). Their mistake is to assume that consciousness science 'can proceed just like other kinds of scientific research'. Papineau (2003) concludes that science itself is 'impotent' when confronted by the 'plurality of such human material correlates for any conscious property' (219). As a result, consciousness science is 'fated to disappoint its more extreme enthusiasts'.

My challenge arises at this point. Papineau's epistemological pessimism rests on an implicit assumption that testing for dissociations is the only method available to scientists. But why is this assumption warranted? Testing dissociations may be, as Papineau claims, the 'obvious strategy'. But that does not make it the only method available to 'standard empirical research', 'empirical investigations', and 'empirical science' (2002: 177; 2003: 214). Like Prinz, Papineau does not explain why inference to the best explanation cannot solve the multiple kinds problem. He also does not offer any argument to rule out a neglected or new method for solving it. Admittedly, it is difficult to prove a negative – that is, to prove that no other scientific methods are available. However, I think it reasonable that, before we assent to Papineau's claim that consciousness science is impotent, we should examine how scientists in other fields cope with the multiple kinds problem.[31]

[30] Papineau's titles indicate the scope of his claim. Papineau (2002), ch. 7, is titled 'Prospects for the scientific study of phenomenal consciousness', while Papineau (2003) asks 'Could there be a science of consciousness?'

[31] In a recent commentary on Block (2007a), Papineau (2007) reconsiders his pessimism: persuaded by Block's 'richly textured use of the empirical data', he is now 'less pessimistic about the possibility

Intriguingly, Papineau brings up one method: stipulation. He sometimes suggests that we can 'refine' a concept that has indeterminate reference. To refine this concept, we stipulate a referent from among the candidates associated with the concept. But what are the philosophical grounds for this kind of conceptual refinement? Does this refinement have a role in science? I shall reconstruct, then evaluate, Papineau's answers to these questions. Let me enter a caveat: I am not sure that I fully understand him, since his remarks on conceptual refinement are brief and scattered throughout his arguments.

This is how Papineau (2002) brings up the possibility of conceptual refinement:

> 'Bald' is a vague term, and different refinements of the term may issue in different verdicts on whether my friend is exactly as bald as I am. As we use the term, there need be no fact of the matter as to whether we are exactly equally bald. (201)

Here he seems to be going beyond a claim that the term 'bald' has indeterminate reference. I take him to be suggesting that its indeterminacy can be reduced by refinement, which will draw a more precise boundary for the term. Papineau extends this reasoning to phenomenal concepts.[32] Consider whether a silicon doppelgänger can see something red. He proposes that '[d]ifferent ways of refining the term "seeing something red" will issue in different verdicts'. Therefore, 'our actual unrefined uses' of this term do not fix whether the doppelgänger can have a visual experience of seeing something as red.

In Papineau (2003), he acknowledges that the analogy with baldness is 'less than

---

that further empirical data may cast even more light on the boundaries of phenomenal consciousness' (521). I note that Block uses inference to the best explanation to interpret his data.

[32] I put aside the fact that his reasoning shifts between the vagueness of concepts and that of terms.

perfect' since phenomenal concepts do not lack a 'sharp cut-off point'. So, instead, he compares phenomenal concepts to the traditional Eskimo term for whale oil.[33] This term's distinctive vagueness is exposed when a petroleum product is introduced as a substitute for natural whale oil. For we see that the term refers indeterminately – either to a 'biologically or chemically identified type' or to 'anything with the requisite appearance and use' (215). Again, 'nothing in previous Eskimo linguistic practice or usage' fixes its reference.

Here, according to Papineau, 'it seems clear that nothing is at issue but a matter of conceptual refinement' (216). *Conceptual refinement*, in his sense, involves 'conceptual decision': we can stipulate whether the Eskimo term refers to one kind or another. I note that this conceptual refinement depends on *empirical discovery*. As Papineau's analysis makes clear, we are not arbitrarily choosing boundaries for a term that lacks a sharp cut-off point. Rather we are choosing one kind over another to be the referent of a term; at least some of these kinds are discovered through empirical investigation. Our stipulation is, thereby, empirically constrained. Papineau believes that, despite appearances, his analysis also applies to the referential indeterminacy of phenomenal concepts. If we resist the analogy, it is only because 'we find it so natural to think about conscious experiences dualistically'.[34]

Let me highlight two problems with Papineau's analysis. First, he does not explain *why* referential indeterminacy should be reduced by empirically constrained

[33] He borrows this comparison from Block (2002b), 422. According to Block, referential indeterminacy arises in this case because 'it may be indeterminate whether the Eskimo term is a natural kind term or not'.

[34] Papineau connects this claim with his earlier diagnosis, in which phenomenal concepts drive us towards dualism by producing an 'illusion of distinctness'. (See note 24 above.) The illusion is especially strong with the concept of consciousness-as-such: 'Surely a light is on, or it is not' (219).

stipulation. Their connection just 'seems clear' to him, at least for the Eskimo term. His earlier analysis of the term 'bald' appears to fit the theory that vagueness arises from semantic indecision. As I noted in §2.2, this theory claims that the precise boundaries of vague terms are yet to be fixed, and that we can stipulate a more precise boundary in order to classify borderline cases more definitely. So it may be that Papineau is drawing on a controversial theory of vagueness for support.

I can cite some indirect evidence that he is doing so. Papineau explicitly disavows two other philosophical theories of vagueness. Without argument, he rejects epistemicism: 'not even God, who knows everything, would be able to tell whether or not the doppelgänger with [our higher-level structural properties] is in pain, for roughly the same reasons that God would be unable to tell whether I am bald or not' (215). And he seems to accept that metaphysical indeterminacy about consciousness-as-such is counter-intuitive: whether a creature is conscious at all does not seem like a fact that could be indeterminate. So his interpretation does not claim that 'how it is for the doppelgänger itself' is vague; rather any vagueness 'lies in our concept' (219).

Second, Papineau does not clarify *when* referential indeterminacy should be reduced by stipulation. Does it apply in both non-theoretical and theoretical contexts? Even if he is right that stipulation is a plausible response to indeterminacy in non-theoretical terms, such as the Eskimo term for whale oil, it is not clear that stipulation is a viable method for reducing indeterminacy in theoretical terms.

I think Papineau evinces ambivalence on this point. On one hand, some of his

remarks suggest that stipulation can be used to reduce indeterminacy only in practical, rather than theoretical, contexts. For instance, Papineau (2002) claims that, if we encounter a silicon doppelgänger, then 'we would quickly come to regard it as conscious, and treat it accordingly' (230). But he clarifies that this conceptual decision would only be appropriate with respect to psychological concepts of consciousness, not the phenomenal concept of consciousness-as-such. The former concepts, which are not about experiences, play a dominant role in 'practical life, as opposed to theoretical reflection'. In Papineau (1993), he suggests that we would, 'in practice', count certain borderline conscious creatures as conscious '*because* we regard them as objects of moral concern' (128-9). This decision would be driven by a moral purpose: 'the moral conclusion would give us a motive for refining the indeterminate notion of consciousness in such a way' (129).[35]

Moreover, Papineau (2002) stresses his 'pessimistic conclusions' about the science of phenomenal consciousness (229). As I noted earlier, he makes much of this epistemological limitation: science is impotent when confronted with the many material properties correlated with consciousness in humans, and it is 'unable to discriminate' between these properties (229). If Papineau were to believe that stipulation is a viable method for reducing indeterminacy in theoretical contexts, then it would be a mystery why he exhibits such pessimism. It would also be difficult

[35] This point is reiterated in Papineau (1996), in reply to Tye (1996): 'it seems to me very dangerous to let moral decisions like this await a resolution of this supposed further issue about consciousness' (695). Papineau (2002) also claims that, 'in practice', we adopt another 'decision procedure' for identifying which human states count as conscious: these are states that 'can be thought about in first-person ways' (126).

Putnam (1964) seems to endorse a moral basis of stipulation: 'If we are to make a decision, it seems preferable to me to extend our concept so that robots *are* conscious – for "discrimination" based on the "softness" or "hardness" of the body parts of a synthetic "organism" seems as silly as discriminatory treatment of humans on the basis of skin color' (407).

to make sense of his insistence that consciousness science has limited prospects because it cannot proceed 'just like other kinds of scientific research'.

On the other hand, Papineau (2003) seems to grant stipulation a role in theoretical contexts. He claims that 'concepts defined in terms of theoretical roles can be taken either way' (217). According to Papineau, such concepts include the phenomenal concepts used in consciousness science. What does it mean to take a concept either way? He suggests that discoveries of referential indeterminacy be seen as 'opportunities for terminological decision'. Furthermore, he implies that this decision can be made through 'terminological fiat alone' (218). Unfortunately, he does not cite any evidence to support this role for stipulation.

Why does Papineau claim that some concepts used in science 'can be taken either way'? I take him to be relying on his model of theory-dependent terms (Papineau 1996). A term is called 'theory-dependent' in his sense when some theoretical assumptions contribute to its definition. The main point of Papineau's model is to demonstrate that a theoretical term can have determinate reference even if its definition is imprecise.[36] The arguments that make this point are not of interest here. Rather I want to examine how his model allows for special cases in which imprecise definitions lead to indeterminate reference. Papineau distinguishes two types of these cases. The first type involves a term that refers indeterminately to more than one property, so it is directly relevant to his claims about consciousness-as-such. In the second type of case, it is indeterminate whether a term refers at all.

[36] Papineau's model is part of the philosophical literature on the semantics of theoretical terms. This literature focuses on how scientific theory contributes to the meaning of these terms; it is surveyed in Andreas (2013). See also Field (1973) and Kitcher (1978); I will discuss their semantic models in the next chapter, when I apply them to species terms.

Here are his examples of the first type:

> …modern microbiology tells us that various kinds of chunks of DNA satisfy the undisputed criteria for "gene", and that further assumptions are needed to narrow the referent down. Similarly, relativity shows that both rest mass and relativistic mass satisfy the original Newtonian definition of mass as proportional to amount of matter, and that further criteria are required to render the referent of "mass" unique. (18)

For Papineau, the 'obvious remedy' for this referential indeterminacy is to refine the definitions of the terms through stipulation. He doubts that, during this refinement, there is 'anything which makes it right to go one way rather than another'. Indeed, he claims that the history of science reveals no 'obvious principle' by which scientists refine their terms so as reduce referential indeterminacy (19). Scientists tend to stipulate a referent based on sociological factors, 'rather than any substantial semantic or empirical facts'. In particular, he cites the balance of 'conservative and radical tendencies among theoretical innovators'.

To support this claim about stipulation, Papineau contrasts what happened to the terms 'caloric' and 'electricity'. According to him, these terms present 'structurally similar' cases (19). Initially scientists believed that both terms refer to fluids flowing between bodies. Then they discovered that, in each case, the appearance of flow is produced by kinetics. Yet the scientists involved drew opposing conclusions about the reference of these terms: caloric does not exist, while electricity does. This suggests, to Papineau, that scientists stipulate when they need to refine a theoretical term with indeterminate reference: 'there will generally be no objective reason to refine it in one direction rather than another' (19).

This argument for stipulation is weak. I find Papineau's examination of the caloric and electricity cases too cursory to be compelling. Here I raise three questions about these cases. First, did stipulation really play a role in them? Without looking at more historical details, we cannot confirm that scientists had 'no objective reason' to decide each case. After all, the structural similarities cited by Papineau do not exclude significant theoretical differences. Second, even if scientists did stipulate in these cases, what served as the basis of their stipulation? Papineau says only that it is 'not at all implausible' that the scientists stipulated according to their conservative and radical tendencies.

My third question brings us back to the multiple kinds problem. Are Papineau's claims about the terms 'caloric' and 'electricity' generalisable to terms that may refer to more than one property? I do not think that we can take this for granted. As Papineau admits, the caloric and electricity cases are different from those that involve multiple kinds: Lavoisier and Franklin risked finding *no* referent at all for their theoretical terms, rather than finding *too many*. Yet Papineau assumes that his analysis can also make sense of what happened to the terms 'gene' and 'mass'. According to him, before stipulation, the term 'mass' referred indeterminately to rest mass and relativistic mass. Upon the discovery of relativity, scientists had to spell out 'further criteria' in order to stipulate a unique referent. Since he cites no historical details about this case, the role of stipulation is hardly demonstrated. Indeed, it is controversial among other philosophers whether the term 'mass' ever

had indeterminate reference.[37]

4.5 Conclusion

In this chapter, I explained and evaluated three recent responses to the multiple kinds problem in consciousness science. My evaluation concluded that the arguments made by Irvine, Prinz, and Papineau do not warrant their pessimism about the science of phenomenal consciousness. I challenged Irvine's premature judgement that the explanatory benefits of a new classification in consciousness science will outweigh its practical costs. Then I objected to Prinz's assumption that testing for dissociations is the only method available for solving the multiple kinds problem. I also questioned if he can, consistently, evade a mysterian conclusion in the debate on human consciousness while he endorses a mysterian conclusion in the debate on artificial consciousness. Finally, I showed that Papineau's arguments on phenomenal concepts are too speculative about their semantic mechanisms and too restrictive about their potential uses.

However, in their arguments, I found three clues on how to address the multiple kinds problem. I intend to follow these clues in the rest of this thesis. First, we should explore how biologists refine their classification of species. Second, we need a unified approach to the multiple kinds problem, which attends to its ramifications in the debates on phenomenal consciousness, borderline consciousness, and artificial consciousness. Third, we should examine whether scientists can solve the

[37] For contrasting views, see Field (1973), Earman and Fine (1977), and Kitcher (1978), 546.

multiple kinds problem by empirically constrained stipulation. Some of Papineau's arguments direct us to this possibility: when scientists discover more than one overlapping kind in their samples of a putative kind, can they legitimately stipulate that their natural-kind concept refers to one of these kinds?

Chapter 5

Stipulation in the species problem

5.1 Introduction

In Chapter 4, I evaluated three recent philosophical responses to the multiple kinds problem in consciousness science. My evaluation challenged Irvine's methodological arguments for eliminating the concept of consciousness, Prinz's epistemological arguments for accepting irremediable ignorance of the nature of phenomenal consciousness, and Papineau's conceptual arguments for attributing referential indeterminacy to the concept of consciousness-as-such. I rejected their pessimism about consciousness science. Then, following Papineau's lead, I raised the possibility that scientists can solve the multiple kinds problem by stipulating that their natural-kind concept refers to one of the kinds discovered.

To examine this possibility, I shall now turn to biological classification. My aim is to show that biologists can solve the multiple kinds problem by empirically constrained stipulation. First, I explain why the model of natural kinds in Kripke (1980) and Putnam (1975) creates a presumption against stipulation, rather than a prohibition. In this model, stipulation brings the risks of metaphysical caprice, epistemological complacency, and semantic confusion. Second, I explain why the 'species problem' among biologists means that they face the multiple kinds problem too. Then I evaluate some arguments in LaPorte (2004, 2010) that support stipulating the reference of species concepts. Third, I propose a new epistemological argument for stipulation. My argument is based on biological practices described by biologists (Cracraft 2000; Coyne & Orr 2004; de Queiroz 2007) and philosophers of biology

(Stanford 1995; Ereshefsky 1992, 2010). Fourth, I make sense of stipulation from the metaphysical and semantic perspectives. To do so, I explain how stipulation enables biologists to explore different evolutionary structures related to a species. And I clarify how stipulation affects the reference of the species term.

My discussion of stipulation will focus on the Kripke-Putnam model of natural kinds.[1] Here are three reasons why. First, as I showed in Chapter 3, this model is used in philosophical debates on consciousness science. For instance, Block and his interlocutors cite it when they assess the hypothesis that phenomenal consciousness is a natural kind (Block 1995a: 236 in original; Church 1995: 426; Block & Stalnaker 1999: 3ff.; Block 2007a: 481; Block 2007b: 534). Shea and Bayne (2010) modify the model in their 'natural kind methodology' for studying the vegetative state (§4); Allen and Trestman (2014) use it to analyse scientific theories of animal consciousness (§4.8 and §5). Second, the Kripke-Putnam model has provoked philosophical debates on the use of stipulation in biological classification. I will assess critiques of the model by LaPorte (2004, 2010).

Third, as I show in the next section, the model integrates Kripke and Putnam's insights on natural kinds from the metaphysical, epistemological, and semantic perspectives. It is, in this sense, more comprehensive than other models that focus

[1] The philosophical literature on natural kinds is vast and expanding. I learnt most from the surveys in Hacking (1991, 2007), Koslicki (2008), and Bird and Tobin (2015). Alternative models of natural kinds appear in four main sources: (a) the *empiricist tradition*, whose analyses of scientific groups and kinds precede Kripke and Putnam (Mill 1874; Quine 1969); (b) *analytic metaphysics*, including appeals to natural kinds in metaphysical theory (Lewis 1983, 1984; Hall 2010a, 2010b) and analyses of their ontological role (Ellis 2001, 2008; Hawley & Bird 2011); (c) *philosophy of science*, which has developed more flexible models of natural kinds (Boyd 1988, 1999a, 1999b, 2010; Slater 2014) and applied them to the special sciences (Shea & Bayne 2010; Magnus 2012; Khalidi 2013; MacLeod & Reydon 2013); and (d) *philosophy of language*, which continues to debate Kripke and Putnam's views on the semantics of natural-kind terms (Stanford & Kitcher 2000; Schwartz 2002; LaPorte 2004, 2010; Beebee & Sabbarton-Leary eds. 2010, ch. 2-5).

on just one or two of these perspectives. This makes it ideal for analysing the metaphysical, epistemological, and semantic impact of stipulation. Moreover, the model is central to some recent debates on natural kinds, which occur in between metaphysics, philosophy of science, and philosophy of language (Beebee & Sabbarton-Leary 2010).[2]

5.2 The presumption against stipulation

Before I look at biological classification, let me explain how the Kripke-Putnam model of natural kinds creates a presumption against the use of stipulation in scientific classification.[3] I shall analyse the metaphysical, epistemological, and semantic principles that Kripke (1980) and Putnam (1975) build into their model.[4] My aim is to show that these principles do not prohibit the use of stipulation. Rather the principles imply that its use brings the risks of metaphysical caprice, epistemological complacency, and semantic confusion. These risks account for the presumption against stipulation. They must be addressed by any philosophical account of natural kinds that allows for stipulation.

[2] LaPorte (2004) reports that Kripke and Putnam's ideas form the basis of a 'celebrated philosophical tradition' that is 'typically taken for granted in high-profile philosophical discussions from a wide range of diverse philosophical areas' (1). The Kripke-Putnam model is so dominant that rival models are often built by modifying it. For instance, Boyd (1991, 1999, 2013) relaxes some assumptions made by Kripke and Putnam, while Ellis (2001, 2008) adds new constraints. I believe that my arguments for stipulation can work with some of these models, though I will not try to prove so here.

[3] Cf. Rescher (2001), 19: 'A presumption…is to be given *some* credit, even if not a totally unqualified endorsement. Its epistemic standing is not rock-solid, to be sure, but it is enough to call for a sufficient effort for its dislodging. Our commitment to a presumption may not be absolute, but it is nevertheless a commitment all the same.'

[4] In describing the Kripke-Putnam model, I tend to emphasise the similarities between Kripke and Putnam's views. I cite only the differences that are relevant to stipulation. Their differences are more systematically analysed in Putnam (1990) and Hacking (2007b). Hacking highlights their overlapping interests: 'Kripke was doing semantics and the theory of modalities, while Putnam was doing semantics and the philosophy of the natural sciences' (4).

The Kripke-Putnam model of natural kinds is based on what Kripke (1980) and Putnam (1975) say about theoretical identities established by science. These identities are expressed in statements such as 'water is H_2O', 'gold is the element with the atomic number 79', 'lightning is an electrical discharge', and 'heat is molecular motion'. 'Water', 'gold', 'lightning', and 'heat' are used here as terms for natural kinds and natural phenomena.[5] Both Kripke and Putnam agree on the necessary *a posteriori* status of these identities. They argue that, if water is H_2O, then it is necessarily H_2O; there is no possible world in which water is not H_2O. This identity between water and H_2O has been discovered by scientific investigation. It is, therefore, known to us *a posteriori*. It is not knowable *a priori* through our linguistic analysis of the term 'water' or our metaphysical intuition about the essence of water.

To develop this analysis into a model of natural kinds, Kripke and Putnam address two related questions: What makes a natural kind natural? And what makes for the essence of a natural kind?[6] The Kripke-Putnam model considers a kind to be natural if it reflects the world's basic structures. Kripke and Putnam distinguish between the superficial and structural properties associated with a kind. The superficial properties are the 'identifying marks' by which we initially identify the kind (Kripke

[5] I follow the literature on natural kinds in assimilating both countable and non-countable phenomena (Bird and Tobin 2015; Kripke 1980, 134). Strictly speaking, Kripke (1980) does not include light and heat as natural kinds, though he stresses the semantic similarities between natural-phenomena terms, such as 'light' and 'heat', and natural-kind terms (134), as well as the metaphysical similarities between natural phenomena and natural kinds (138). Putnam (1973) notes the kinship between what he calls physical-magnitude terms, such as 'temperature' and electricity', and natural-kind terms.

[6] Neither Kripke (1980) nor Putnam (1975) expends theoretical effort on a third metaphysical question, which focuses on the ontological status of natural kinds: What type of entities are natural kinds? Are they, for instance, reducible to sets or universals? See Bird and Tobin (2015), §1, for the distinction between these three questions. Hawley and Bird (2011) also separate what they call the naturalness question from the kindness question.

1980, 119); they belong to an 'operational definition', with which we can fallibly find instances of the kind (Putnam 1975, 232). The structural properties, on the other hand, are causally or constitutively related to the superficial properties. For instance, the superficial properties of gold include its yellow colour, shine, malleability, and ductility; its structural properties include being composed only of the element with the atomic number 79.

In the Kripke-Putnam model, natural kinds are individuated via their structural properties.[7] Kripke (1980) says:

> In general, science attempts, by investigating basic structural traits, to find the nature, and thus the essence (in the philosophical sense) of the kind. The case of natural phenomena is similar; such theoretical identifications as 'heat is molecular motion' are *necessary,* though not *a priori*. (138)

Thus, the structural properties associated with a kind are the essential properties that distinguish the kind from another. They are the properties that belong to every possible instance of the kind. For instance, the molecular structure H_2O is part of the essence of water. This implies that every possible sample of water is composed of H_2O molecules. Similarly, Putnam (1975) argues that a substance is not gold just because it has the 'superficial characteristics' of gold (235). Rather it must have 'the same general *hidden* structure (the same "essence", so to speak) as any normal piece of local gold.' This essence of gold includes the structural property of being composed only of the element with the atomic number 79. No possible sample of

[7] Neither Kripke nor Putnam define structural properties strictly: when they present the general model, they speak of 'basic structural traits' and 'hidden structure'. Their model can be expanded so that essential properties can be extrinsic, dispositional, etc. (See note 18 below.)

gold is composed of other elements.[8]

Because the essences of natural kinds are determined by the world's basic structures, these kinds are useful in science. In particular, they ground scientific explanations and inductions about the world. While Kripke does not make this point, Putnam mentions the role of natural kinds in explanations. In Putnam (1970), he describes natural kinds as 'classes of things that we regard as of explanatory importance; classes whose normal distinguishing characteristics are "held together" or even explained by deep-lying mechanisms' (139). Later, in Putnam (1990), he highlights the significance of individuating substances via their 'subvisible structure': 'the subvisible structure explains why different substances obey different laws' (60). Moreover, scientists can justifiably infer from the premise that all observed instances of a kind behave in a particular way to the conclusion that all instances of the kind will do so.[9]

I can now explain why stipulation brings the risk of *metaphysical caprice*. In the

[8] Later Putnam (1990) modifies the scope of this claim to cover only possible worlds which obey the same physical laws of nature. He thereby interprets the essence of a natural kind in terms of physical necessity, rather than metaphysical necessity: 'I now think that the question, "What is the necessary and sufficient condition for being water *in all possible worlds?"* makes no sense at all. And this means that I now reject "metaphysical necessity"' (70, original italics). I put aside this modification in Putnam's view since it does not affect my arguments on stipulation.

[9] The role of kinds in induction is analysed by Quine (1969) and Kornblith (1993). Quine (1969) notes that mankind has developed 'modified systems of kinds, hence modified similarity standards for scientific purposes' (276). Through scientific theorising, it has 'regrouped things into new kinds which prove to lend themselves to many inductions better than the old' (276). Although he disapproves of talk about essences, his model of theoretical kinds is an important pre-cursor to Kripke and Putnam's model of natural kinds. Why? Quine conceives of theoretical kinds as 'functional groupings in nature': the members of each kind bear similarities, which are grounded in mechanisms and structures discovered by the mature sciences (279). So, in his model too, theoretical kinds reflect the world's structures.

Before Quine, Mill (1874) also includes kinds in his account of induction. But the connection between Mill and the Kripke-Putnam model is more tenuous: his own model of 'real Kinds' excludes kinds whose structural properties are lawfully related to superficial ones. See Hacking (1991), 119-121; Magnus (2014).

Kripke-Putnam model, the essences of natural kinds are supposed to be determined by the world's basic structures. More specifically, a natural kind is individuated by the structural properties that explain its superficial properties. But if scientists stipulate a kind, how will the kind reflect the relevant structures rather than the scientists' whim? Here, I worry, is where the caprice of scientists disrupts the classifications in science. Insofar as a stipulated kind fails to reflect the world's structures, it will be less useful in explaining the world.

Next, I consider the epistemology of natural kinds. In the Kripke-Putnam model, the essences of natural kinds are discovered through scientific investigation. That is why Kripke classifies theoretical identities as necessary *a posteriori* truths, rather than *a priori* ones. Putnam agrees, and acknowledges the revolutionary significance of Kripke's insistence that 'a (metaphysically) necessary truth could fail to be *a priori*' (1975: 233). This epistemological aspect of their model seems to prohibit any stipulation of kinds. As LaPorte (2004) notes, its implication is now part of 'received wisdom' about the Kripke-Putnam model: 'Scientists' conclusions about essences are discoveries, not stipulations' (63).

However, I want to draw attention to an aspect of the Kripke-Putnam model that is often neglected or unmentioned. In my interpretation, this aspect of their model allows for some stipulation in scientific classification. Kripke (1980) describes what might happen if the multiple kinds problem arises in the case of gold.

> If…the supposition that there is one uniform substance or kind in the initial sample proves more radically in error, reactions can vary: sometimes we may declare that there are two kinds of gold, sometimes we may drop the term 'gold'. (These possibilities are not supposed to be exhaustive.) (136)

I think his claim that reactions to the multiple kinds problem 'can vary' is significant. Kripke is implicitly conceding that, in this predicament, the world's structures are insufficient to determine the essence of a kind. Instead we have to decide between various possibilities, which he has yet to specify exhaustively.

Putnam (1975) is more explicit about the room for variation. Indeed, he claims that it already exists in the case of water. According to his account, to be water is 'to bear the relation $same_L$' to the samples that we point to (238). Here is how Putnam defines the relation $same_L$: '*x* bears the relation $same_L$ to *y* just in case (1) *x* and *y* are both liquids, and (2) *x* and *y* agree in important physical properties.' He acknowledges that importance is an 'interest-relative notion' (239). Normally, the important properties of a substance are those which are '*structurally* important': they specify what the substance is made of, and how they are arranged to produce its 'superficial characteristics'. So, normally, the most important property of what we call 'water' is its molecular structure H_2O.

But Putnam acknowledges that, even in normal circumstances, what we count as structurally important can vary:

> …it may or may not be important that there are impurities; thus, in one context 'water' may mean *chemically pure water*, while in another it may mean the stuff in Lake Michigan…Again, normally it is important that water is in the liquid state; but sometimes it is unimportant, and one may refer to a single H_2O molecule as water, or to water vapour as water ('water in the air'). (239, original italics)

For Putnam, the world's structures are not sufficient to determine the essence of what we call 'water'. In his examples, we have to decide between these possible

kinds: a liquid consisting only of H_2O molecules, or one consisting mostly of H_2O molecules; a liquid consisting of H_2O molecules, or a collection of H_2O molecules in either liquid or non-liquid state; a collection of H_2O molecules or simply the molecular structure H_2O. And the scientists' decisions will depend on their interests.

When Putnam (1990) discusses the multiple kinds problem, he brings up interest relativity again. He compares two samples of iron: the first with only one isotope, the second with a naturally occurring distribution of isotopes. Do they belong to a natural kind? Or are they samples of two different substances? His answer: 'Well, it may depend on our interests. (This is the sort of talk Kripke hates!)' (68). According to Putnam, there will be 'some component of interest relativity here, and, perhaps, some drawing of arbitrary lines'. He does not explain the relation between interest relativity and arbitrariness, though he notes that the arbitrariness in defining a kind is 'infinitesimal' compared to that in distinguishing a particular.

So both Kripke and Putnam suggest that scientists sometimes need to stipulate the reference of a natural-kind term. Yet neither offers much evidence. Kripke mentions some possibilities, while Putnam claims that stipulation may depend on interests and involve arbitrariness. What are the conditions under which stipulation can be used? And what is the correct basis for stipulation under those conditions? Unless these questions are answered, I fear that their proposal risks *epistemological complacency.* It assumes that stipulation can solve the multiple kinds problem. And it fails to define enough constraints on stipulation.

Finally, I turn to the semantics of natural-kind terms. Both Kripke (1980) and Putnam

(1975) deny that a natural-kind term is synonymous with a description that identifies the instances of a kind.[10] Instead, they propose a causal-historical model of reference. Suppose that a speaker (or a group of speakers) intends for a term to refer to a natural kind. The reference of this natural-kind term is fixed when he uses it to baptise an initial sample of the kind, or when he picks out the kind by description. The term is then passed along a social and historical chain of communication to other speakers. To refer to the kind, other speakers need not be in direct contact with it.[11] They also need not have beliefs that suffice to identify instances of the kind; indeed, they may well have false beliefs about the kind.

For my purpose, the most significant implication of this model is that the reference of a natural-kind term remains stable even when scientists revise our beliefs to reflect their discoveries. Kripke (1980) adds that the term's meaning also does not change:

> …on the present view, scientific discoveries of species essence do not constitute a 'change of meaning'; the possibility of such discoveries was part of the original enterprise. We need not even assume that the biologist's denial that whales are fish shows his 'concept of fishhood' to be different from that of the layman; he simply corrects the layman, discovering that 'whales are mammals, not fish' is a necessary truth. (138)

According to Kripke, a scientific discovery about a natural kind leads only to an epistemological correction, as our old beliefs about the kind are replaced by new

[10] In this respect, they see a resemblance between natural-kind terms and proper names. Kripke (1980) notes: 'my argument implicitly concludes that certain general terms, those for natural kinds, have a greater kinship with proper names than is generally realized' (134). Putnam (1973) acknowledges his 'heavy indebtedness' to Kripke's ideas about proper names.

[11] For simplicity's sake, I will assume that a natural-kind term refers to a kind. This assumption is contested in the literature on the semantics of natural-kind terms: see Schwartz (2002), Salmon (2003), LaPorte (2004), 38-43, and Koslicki (2008), section IV. Besson (2010) and Martí and Martínez-Fernández (2010) defend the assumption. Those who reject it can recast my arguments in terms of the extensions of natural-kind terms.

ones. It does not imply any change in the extension of the natural-kind term. So the kind to which we refer with the term 'fish' never included whales. This is in contrast with the view that both Kripke and Putnam reject. According to that view, the term 'fish' is synonymous, at a given time, with a description that we use at the time to characterise the kind. Before the discoveries of marine biology, the term 'fish' would have referred to a kind that includes whales; later, the term came to refer to another kind that excludes whales.

Putnam (1973) highlights one advantage of their model. It supports the assumption that scientists can talk about the same entities even though they are committed to different theories:

> ...Bohr would have been referring to electrons when he used the word 'electron', notwithstanding the fact that some of his beliefs about electrons were mistaken, and *we* are referring to those same particles notwithstanding the fact that some of our beliefs – even beliefs included in our scientific 'definition' of the term 'electron' – may very likely turn out to be equally mistaken. (197)

As Putnam (1975) points out, this assumption informs our scientific practice. It is 'beyond question' that scientists talk 'as if later theories in a mature science were, in general, *better* descriptions of the *same* entities that earlier theories referred to' (237). The causal-historical model of reference explains why scientists are right to do so; it thereby explains how scientists can communicate their results with each other despite different theoretical commitments.

I note that, in this model, stipulation brings the risk of *semantic confusion*. If scientists stipulate the essence to be associated with a natural-kind term, they are

likely to shift the reference of the term. Two dimensions of semantic confusion may arise. The first is synchronic: might stipulation sow confusion between scientists who continue to use the term? The second is diachronic: might stipulation abruptly change the subject of the scientists' classification? To address this risk, a defence of stipulation must explain how scientists can minimise terminological confusion among themselves and avoid abruptly changing their subject.

Both Kripke and Putnam concede that their model needs to be modified to fit non-ideal situations.[12] Putnam (1975) includes one modification that may help with multiple kinds: he suggests that natural-kind terms 'typically possess a number of senses' (238). The term 'water', for instance, has a family of related senses. In its 'predominant' sense, the term refers to liquids consisting of H_2O molecules. In other senses, the term refers to just one of the other kinds that are consistent with our ordinary samples of water. According to Putnam, some natural-kind terms may have a 'deviant' sense, which still bears 'a definite relation to the core sense'. His example is the sense in which the term 'tiger' can refer to creatures that look just like tigers 'but have a silicon-based chemistry instead of a carbon-based chemistry' (239). He contrasts this with the 'extremely deviant' sense in which 'literally *anything* with the superficial characteristics' of a lemon counts as a lemon.

I conclude that none of the metaphysical, epistemological, and semantic principles in the Kripke-Putnam model prohibits stipulation in scientific classification. In fact, both Kripke and Putnam suppose that scientists can solve the multiple kinds problem by stipulation. However, as my arguments show, their model establishes a certain

[12] See, for instance, Kripke (1980), 139, note 70, and 162; Putnam (1973), 206, question 1; Putnam (1975), 238-41 on 'Other senses'.

presumption against stipulation: its use carries the risks of metaphysical caprice, epistemological complacency, and semantic confusion.

More positively, I see these three risks as setting a set of standards to be met by an account of stipulation in scientific classification. First, we need evidence from the epistemological perspective to support a *role* for stipulation. This evidence should clarify the conditions under which stipulation can solve the multiple kinds problem, and the basis for stipulation under those conditions. Second, we need to examine the *result* of stipulation from the metaphysical and semantic perspectives. This examination should clarify how a stipulated kind reflects the world's structures, and thereby grounds scientific inductions and explanations about the world. It should also clarify how stipulation affects the reference of the kind term, and yet avoids causing terminological confusion among scientists or abruptly changing their subject.

5.3 LaPorte on the species problem

Now I shall examine some recent arguments in LaPorte (2004, 2010) that support stipulation in biological classification.[13] LaPorte contends that 'biologists' conclusions, or future conclusions, about the essences of kinds recognized before current systematic theory are in general not discovered to be true' (92). Instead they

[13] LaPorte (2004, 2010) argues for stipulation in both biological and chemical classifications. His arguments about chemical classification are criticised in the new literature on chemical kinds: see, for instance, Bird (2007, 2010), Hendry (2010, 2012), and Massimi (2012). In contrast, his arguments about species essences have been neglected. They are briefly mentioned in Dupré (2004), Wilson (2004), and Richards (2010), 192. Bird (2007), section IV, criticises LaPorte's arguments on higher taxa, but is more sanguine about those on species: 'I shall not dwell on these points, which though well taken, do not seem to me to be especially problematic' (302).

are, or will be, stipulated to be true. His contention is based on the claim that biologists can legitimately stipulate the reference for species concepts. I focus on his arguments about species such as tigers *(Panthera tigris)* and Baltimore orioles *(Icterus galbula),* which were identified before biologists developed any of the modern species definitions. First, I explain why the 'species problem' means that biologists face the multiple kinds problem too. Second, I evaluate LaPorte's arguments for stipulation. They are not yet compelling: in particular, they illustrate new dimensions of epistemological complacency. In the next section, I will offer a new epistemological argument that avoids these complications.

What is the 'species problem' in biology?[14] This is the longstanding controversy over Species Concepts, in which both biologists and philosophers of biology promote conflicting definitions of what a species is. In the biological and philosophical literature, the definitions are commonly called 'species concepts'. To avoid confusion, I use 'Species Concepts' to refer to these *definitions*, and 'species concepts' to refer to the *concepts* of individual species, such as the concept *Panthera tigris* and the concept *Icterus galbula*. Mayden (1997) identifies over twenty Species Concepts. According to Coyne and Orr (2004), at least nine are 'serious competitors' in biology. They centre on either interbreeding, genetic or phenotypic similarity, ecological niche, evolutionary tendency, or phylogenetic history. Ereshefsky (2010c) emphasises that these Species Concepts in the controversy are 'not fringe or crank

[14] The literature on this problem straddles biology and philosophy of biology. For a sample of scientific analyses, see Claridge, Dawah, and Wilson (1997), Wheeler and Meier (2000), and Coyne and Orr (2004). Cracraft (2000) is an excellent guide to these analyses. For philosophical surveys, see Ereshefsky (2010b, 2010c) and Richards (2010). I also learnt from the early and more speculative analyses by Dupré (1981) and Kitcher (1984, 1989). Richards (2010) argues that the 'species problem' goes back to pre-Darwinian times: Darwin himself was confronted by 'a multiplicity of species concepts, based on similarity, fertility, sterility, geographic location and geologic placement and descent' (75).

concepts, but concepts proposed and investigated by prominent biologists' (261).

LaPorte's examples are two of the most prominent: the Biological Species Concept (BSC) and the Phylogenetic Species Concept (PSC). The BSC defines species as 'groups of interbreeding natural populations that are reproductively isolated from other such groups', while the PSC defines them as the 'smallest diagnosable cluster of individual organisms within which there is a parental pattern of ancestry and descent'.[15] In his interpretation, each definition reflects a hypothesis about species essence. So, according to the BSC, every possible population of a particular species has the ability to interbreed with other populations from the species. According to the PSC, every possible population of a particular species belongs to the same line of descent and shares a certain set of diagnostic traits.

Before I examine LaPorte's arguments for stipulation, let me clarify three related assumptions in his interpretation of the species problem. These allow him to apply the Kripke-Putnam model to species. First, like Kripke and Putnam, LaPorte treats species as kinds. Some philosophers of biology insist that species are individuals with parts, rather than kinds with members (Ghiselin 1974, 1997; Hull 1976, 1978, 1987; Ereshefsky 2010b, §2.2). Among other reasons, they require that kinds be unrestricted in spatio-temporal terms and ruled by laws of nature. Species do not meet these requirements: different generations of a particular species are spatio-temporally connected, and no laws of nature govern a particular species. But LaPorte

[15] LaPorte is quoting Mayr (1969) and Cracraft (1983) respectively. Cracraft's definition is actually one of several phylogenetic Species Concepts; see Coyne and Allen (2004), 281-8.

challenges the basis for such stringent metaphysical requirements on kinds.[16] He also offers an irenic proposal. Even if the organisms from a species constitute an individual, our talk about the species 'could *also* be satisfactorily interpreted as talk about a kind' (15). To implement his proposal, we interpret being part of the species-individual as being essential to the corresponding species-kind.[17]

Second, LaPorte attributes essences to species. But these essences are different from those mentioned in Kripke (1980) and Putnam (1975). The original Kripke-Putnam model draws on scientific folklore of the 1970s. As Hacking (2007a) notes, this folklore held that 'molecular biology would discover in the DNA of a species the necessary and sufficient conditions for being of that species' (233). Thus Kripke (1980) suggests that the essence of tigers includes their 'internal structure' (120); Putnam (1975) points to the 'genetic code' of lemons (240), as well as the 'carbon-based chemistry' of tigers (239).[18] In contrast, LaPorte notes that biologists today 'generally place organisms into taxa on the basis of shared ancestry' (64). This suggests to him that the essence of a species may include the 'hidden historical

[16] For instance, LaPortes notes in I.2: 'Historically delimited species can be kinds. There is evidently no reason there cannot be historical kinds as well as nonhistorical ones. A historical kind would simply be one whose membership conditions involve members' having some causal connection to an independently specified item – for example, the beginning of a lineage' (11). Later he argues in I.5: '…if there can be no laws about individuals, and if there are no laws about species, it does not follow that species are individuals. Species could still be lawless kinds' (14). Moreover, LaPorte follows other philosophers of biology, such as Kitcher (1984), in arguing for the possibility of biological laws about particular species.

See also Koslicki (2008), Section III, which includes other challenges against the species-as-individual thesis.

[17] Similar pragmatic proposals are made in Kitcher (1984, 1987); Dupré (1993), 58; Boyd (1999a), 162-4; and Okasha (2002), 193-4. Ereshefsky (2010c), §2.3, objects to such proposals because they conflate 'two distinct ways in which scientists classify': 'Parts of an individual *must* be appropriately causally connected. Members of a kind *must* be similar' (678). However, he fails to explain why shared ancestry cannot be a similarity.

[18] A caveat: both Kripke and Putnam do not, in principle, exclude non-intrinsic properties from the essences of natural kinds. Putnam (1975) notes that some diseases, such as tuberculosis, have a 'common hidden structure in the sense of an etiology' (241). More recently, Kripke (2013) concedes that 'evolutionary ancestry may also be relevant' to species classification (46).

bonds' between populations from that species.

Third, LaPorte assumes that Species Concepts aim to describe species essences.[19] He offers some indirect support for this assumption. For instance, in their cladistic definitions of taxa, some biologists refer to properties that belong to a taxon in both actual and counter-factual circumstances (2004: 11; 2010: 114-5). They thereby imply that these properties belong to the taxon in all possible worlds – or at least all the possible worlds that matter to their theories. In the Kripke-Putnam model, such properties are part of the taxon's essence. LaPorte (2010) adds that biologists have good reason to 'care about what a taxon's essence is': without identifying this essence, they 'can talk past each other' and end up in unproductive classification disputes (118-9).

I note that some biologists explicitly share LaPorte's assumption. In its latest entry on Species Concepts, the *Encyclopedia of Biodiversity* argues that the Biological Species Concept 'reverts to a new kind of essentialism, where evolutionary maintenance via interbreeding is the underlying reality, or essence of species' (Mallet 2007, 5). Its development has opened 'the Pandora's box of alternative essences, deemed more important by other biologists'.

Some philosophers deny that Species Concepts which refer only to relational properties describe species essences. For instance, Devitt (2008) argues that

[19] I thank Genoveva Martí for urging me to clarify this assumption during a conference discussion. Pedroso (2012) criticises LaPorte's appeal to biological practice to support biological essentialism. He leaves open the possibility that a non-essentialist alternative can 'account for the explanatory role of biological taxa without relying on essences' (189). I note that those who adopt this alternative model of species still face the multiple kinds problem: LaPorte's arguments on stipulation can be recast to fit their model.

relational properties cannot adequately explain the traits associated with a species: 'An adequate explanation must appeal to intrinsic properties of the organisms' (231).[20] For now, let me observe that LaPorte does not, in principle, exclude intrinsic properties from the essence of a species. He acknowledges that concepts such as the Phylogenetic Species Concept may be developed to refer to both historical bonds and 'genetic structure' (177). So those who agree with Devitt can apply LaPorte's arguments only to Species Concepts that refer, partly or potentially, to intrinsic properties. Later I will try to vindicate LaPorte's assumption by explaining why some of these concepts need not refer to intrinsic properties.

Following LaPorte's interpretation, the current controversy over Species Concepts means that biologists face the multiple kinds problem too. They inevitably find more than one kind in the samples of a putative species. For each species, the Biological Species Concept and the Phylogenetic Species Concept pick out different kinds. Both kinds include our samples of the species. In LaPorte's example, the initial sample of *Panthera tigris* consists only of Bengalese tigers. According to the BSC, the species *Panthera tigris* includes both Bengalese and Sumatran tigers as subspecies. Bengalese tigers can breed with Sumatran ones, even though the latter are uniquely adapted to survive in the dense jungle. On the other hand, the PSC excludes Sumatran tigers from *Panthera tigris*: they are 'distinct enough' from Bengalese tigers to form a 'diagnosable cluster', so they count as a different species (72). Therefore, Sumatran tigers belong to only one of two kinds that biologists discover in our samples of *Panthera tigris.* So which of these kinds, if any, is *Panthera tigris*?

[20] For some critical discussion of Devitt's argument, see Ereshefsky (2010c), §3.

This multiple kinds problem has not been solved despite widespread advances in biology. Indeed, it seems to be getting worse. As biologists discover more about biodiversity, they develop more Species Concepts. Here is how LaPorte responds to the problem:

> The recent literature offers dozens of professional conceptions of what a species is, and there does not seem to be any fact of the matter about which ought to be kept. It seems likely enough that one or another "species concept", as biologists call them, will prevail over others in biological discourse, but not that it will be discovered to be the true concept. This is a problem for the view that scientists' conclusions about the essences of particular species are discoveries. (71)

In this passage, LaPorte acknowledges the current disagreement among biologists on how to define a species. But he does not ask for more data to settle their disagreement. Rather he suggests that there is no 'fact of the matter' about which concept 'ought to be kept'. According to LaPorte's view, biologists can legitimately choose between the Species Concepts, and thereby stipulate the reference of the concept *Panthera tigris*. LaPorte draws two implications from this view.[21] First, if one Species Concept prevails over the others, scientists will not have discovered it to be 'the true concept'. Second, their conclusion about the essence of *Panthera tigris* depends on their stipulation that the concept *Panthera tigris* refers to the kind picked out by their chosen Species Concept.

I note that LaPorte never claims that there is no fact of the matter about which Species Concept *is* correct or true of the world. So, in my interpretation, he is not

[21] As LaPorte (2010) clarifies, it does not imply that species essences are 'constructed or invented' (117). Rather biologists choose between essences that are determined by the world's structures.

committed to any metaphysical indeterminacy in the nature of species.[22] Other formulations in LaPorte (2004) also emphasise the practical aspect of his stance: 'neither offers the only acceptable use for "species"' (74); there is 'more than one legitimate way to settle the use of "species"' (74); 'there is no fact of the matter that one proposal for delimiting species is the right one, and others wrong' (76). Of course, the fact that biologists disagree on what counts as a species does not license them to stipulate what does. As LaPorte acknowledges, 'the presence of competing views in the scientific community is not itself reason to say that an answer is up for grabs' (73). Correspondingly, the fact that more than one kind fits what we know about *Panthera tigris* offers no reason to believe that either kind can legitimately be the referent of the concept *Panthera tigris.* To assume otherwise would be – in my terms – epistemologically complacent.

So how does LaPorte support his view? He cites the distinctive character of the biologists' disagreement. According to him, this disagreement reflects classificatory interests that are diverse and difficult to manage. It is also rooted in semantic vagueness. I shall reconstruct these two related arguments in LaPorte (2004), before I evaluate them. Although LaPorte explains the scientific background clearly, he tends not to dwell on crucial steps in his arguments.

LaPorte's main argument draws on the *diverse classificatory interests* in biology. These include empirical and non-empirical interests. He does not elaborate on the empirical interests at stake, though he requires that the competing Species Concepts

[22] My interpretation is supported by another piece of evidence: in his notes 19 and 20, LaPorte explicitly denies that he is committed to any philosophical account of vagueness (193). I will come back to this point.

'measure up to the empirical world' (73). He also recognises that both the Biological Species Concept and the Phylogenetic Species Concept classify populations into 'natural and scientifically interesting' groups.[23] Since these concepts participate in different explanations, they serve different empirical interests. So the choice between them is 'not completely arbitrary'. However, LaPorte emphasises the dominant role of non-empirical interests in determining how biologists classify populations.[24] He cites, as examples, the biologists' vision of their field, the concept's ease of use, and its tidiness.

How do the Biological Species Concept and the Phylogenetic Species Concept compare with respect to these non-empirical interests? Using the BSC requires less laboratory work than using the PSC. Thus the BSC is favoured by field workers, as well as biologists 'who see classification as a tool for the wider scientific community and even for interested lay speakers' (74). On the other hand, the PSC is preferred by biologists who want classification to 'convey much more information for specialists'. The BSC is easier to use than the PSC because it simplifies classification. It places interbreeding populations, such as the Bengalese and Sumatran tigers, into a polytypic species. On the other hand, the PSC is tidier than the BSC because it is more comprehensive. Unlike the BSC, the PSC is able to classify asexual organisms into species. Therefore, as LaPorte notes, the BSC and the PSC accrue different 'costs and benefits' (74).

The result is an epistemological stalemate between proponents of different species

[23] For LaPorte, a natural group is one with 'explanatory value'(20). In the right circumstances, it is also 'useful for prediction and control'.

[24] This is affirmed later in his chapter: 'the considerations that would have to be taken into account to determine the acceptable use of "species" would be largely non-empirical' (75).

concepts. According to LaPorte, they often claim that 'this or that concept is the "correct" one' (73). Yet none of the concepts 'seems to trump its competition in such a way that it emerges as the one concept that properly resolves the nature of "species".' So neither the BSC nor the PSC appears to be the best candidate for defining the nature of species, and for identifying the essences of particular species. On the basis of this stalemate, LaPorte deduces that there is room for stipulation:

> Neither could ever be discovered to offer the only acceptable use for 'species', because neither *offers* the only acceptable use for 'species'. There is more than one legitimate way to settle the use for 'species', so the eventual standardization of the use of that term in one direction or another will not have been discovered to be correct. (74)

His conclusion is not that some other concept – besides the BSC and the PSC – offers the only acceptable use for the term 'species'. Instead he believes that biologists have 'more than one legitimate way' to define what a species is. To resolve their stalemate, they can legitimately stipulate 'in one direction or another'. LaPorte does not say more to explain his deduction. Its plausibility appears to rest on his claim that the stalemate between biologists arises 'largely' from differences in the non-empirical interests served by the concepts.

In addition to the diverse classificatory interests in biology, LaPorte highlights the *semantic vagueness* of the term 'species'. According to him, the initial use of this term is 'not precise enough to favour any one competing present-day camp' (73). This implies that our terms for individual species, such as *'Panthera tigris'*, are vague too, since their meaning is tied to the meaning of the term 'species'. Such imprecision allows biologists today to propose different Species Concepts when they

discover more than one kind in their samples of species. Thus, for LaPorte, semantic vagueness lies at the root of the biologists' disagreement. As he clarifies later, this vagueness is not like the vagueness of term 'tall', which depends on gradual variation along one dimension. Instead it is akin to the multi-dimensional vagueness of the term 'religion', where it is 'not clear just what combination of features is necessary or sufficient for qualification' (156).[25]

What tends to happen to vague biological-kind terms? LaPorte reports that they undergo 'meaning refinement', which is 'a change in terms' use' (70). Later he suggests that this semantic refinement involves stipulation:

> My own account of how terms are revised as science advances is that vagueness and confused suppositions underlying the use of the terms are removed after empirical light is shed on items to which the terms have been applied. Vagueness can be dispelled in any of several ways of modifying a term's meanings, no one of which can claim to preserve the exact original meaning of the term. (90)

This passage spells out LaPorte's view from the semantic perspective. First, when terms such as 'species' and '*Panthera tigris'* are introduced into our language, they are vague. This means that biologists who use these terms eventually face the multiple kinds problem. So they propose different Species Concepts, each defining a different set of essences for individual species. Second, in order to eliminate their vagueness, the meaning of these terms needs to be refined. Third, during this refinement, their meaning can be modified 'in any of several ways'.

[25] He cites Alston (1964), who distinguishes between *degree* vagueness and *combinatory* vagueness. In the latter, borderline cases lack some but not all of the properties or parts attributed to normal cases of the phenomenon. Alston's example is Hinayana Buddhism, which shares many distinctive properties of traditional religions, but lacks their belief in a supernatural being.

Next, let me evaluate LaPorte's arguments. I raise two objections to the argument from diverse classificatory interests. First, LaPorte does not take the role of empirical interests seriously enough. In particular, he does not compare the empirical interests served by different Species Concepts. As I explain in the next section, these include interests in different types of organisms and processes. Neither does he relate these empirical interests to the non-empirical ones. He assumes that, when biologists assess Species Concepts, their empirical interests are dominated by non-empirical ones. That is why he focuses on non-empirical interests such as the biologists' view of their field, the concept's ease of use, and its tidiness. Yet this assumption is open to challenge. I doubt, for instance, that biologists will prefer a concept that serves their empirical interests poorly even if it serves many of their non-empirical interests – say, because it is easier to use and helps them to communicate with the public. Consider an analogy: some philosophers of science claim that scientific theories are assessed on their empirical and non-empirical virtues (McMullin 2008).[26] The latter include simplicity, consilience, and fertility. It would, however, be controversial to suggest that scientists assess theories 'largely' on the basis of these non-empirical virtues.

LaPorte's own examples suggest a more complicated relationship between the empirical and non-empirical interests associated with species concepts. He claims that the Biological Species Concept is easier to use than the Phylogenetic Species Concept because it 'conveniently' groups together interbreeding populations into one polytypic species. But this is convenient only for biologists who are interested in

[26] For critical discussion on the significance of non-empirical virtues, see Sober (1988), Kukla (1994), and Turner (2007), ch. 2 and 8.

sexual reproduction. For those who do not share this empirical interest, the BSC can be less convenient than the PSC because it neglects significant differences between interbreeding populations.[27] Similarly, LaPorte notes that many taxonomists find the BSC untidy because it fails to classify asexual organisms into species. However, this disadvantage may not be salient to those biologists who focus on sexual reproduction.

Such connections between empirical and non-empirical interests are obscured by LaPorte's impression that this controversy 'seems largely to come down to a matter of personal preference that could be decided in any of a number of sensible ways' (73). If empirical interests take priority over non-empirical ones in the species controversy, and if different concepts serve different empirical interests, then it is far from obvious that biologists should allow one concept to prevail over others. In my judgement, LaPorte's failure to examine the empirical interests associated with biological classification undermines his case substantially.

Second objection: LaPorte assumes that the epistemological stalemate over different Species Concepts can only be broken by stipulation. Suppose that he is right to focus on non-empirical interests, and that current concepts offer 'different advantages' in relation to these interests. There remain at least two other strategies for breaking this stalemate. One strategy is to improve a concept so that it combines the advantages of the other concepts. Another strategy is to develop a method for

[27] In a footnote, LaPorte paraphrases a remark by the biologist G. G. Simpson, which makes a similar point: 'many taxonomists like to differentiate with a fine brush organisms in which they specialize, even though they hope that taxonomists working with organisms with which they are not so familiar will use a broad brush, which makes taxa handier for the nonspecialist' (186). But he does not explore its significance.

balancing non-empirical interests such that it selects just one of the current concepts. If either strategy is successful, one concept will still emerge as the correct concept, which determines the only acceptable use for the term 'species'.

I think that LaPorte has plausible grounds to reject the first strategy. It is difficult to see how the Biological Species Concept or the Phylogenetic Species Concept can be improved so as to combine the advantages of both concepts. LaPorte's analysis shows us why, though he does not make this point explicitly: some pairs of non-empirical interests are rivalrous. Serving one of these interests inhibits a concept from serving the other. For instance, biologists who are more interested in fieldwork and communication with non-biologists favour the BSC because it requires less laboratory work than the PSC. Yet this feature stops the BSC from conveying the specialist information valued by biologists who favour the PSC. Similarly, the BSC produces a simpler taxonomy because it groups interbreeding populations together. Yet this feature also prevents the BSC from classifying asexual organisms, and being as comprehensive as the PSC. I infer that biologists face inevitable trade-offs between some of the non-empirical interests associated with the BSC and the PSC.

LaPorte offers an argument that might seem to diminish the appeal of the second strategy. It highlights the difficulty that biologists face in managing their non-empirical interests.

> Suppose I am wrong. Suppose there *is* only one acceptable use of 'species' and that this use accords with the use of speakers from centuries past. Even then, there would be no particular reason to think that this one true use of 'species' will ever be *discovered* to be the one true use. (75)

He goes on to offer two reasons to believe that this hypothetical Species Concept will not be discovered. First, identifying this concept is a complex task. It requires biologists to assess interests that are largely non-empirical and balance them against each other: 'one would have to weigh any number of highly delicate considerations that could easily fail to be properly appreciated or balanced against one another' (75). Because the interests are recalcitrant in this sense, the correct concept 'might easily fail to be adopted'. Second, biologists are not known to be competent at this complex task. They cannot rely on the 'usual empirical methods' to balance non-empirical interests (76).

Thus LaPorte promotes a qualified form of epistemological pessimism: 'even if there *is* a single correct use of "species", there is little reason to think that scientists will discover it rather than settle on some other species concept' (76). But his prediction is based on the scientists' current abilities in assessing and balancing non-empirical interests. He has yet to offer any argument against the development of a new method to assess and balance non-empirical interests 'properly'. Such a method may require the growth of new institutions to establish a consensus on one concept, and enforce it among biologists. I am not claiming that this is either likely or desirable. My point is that LaPorte's epistemological pessimism is not well supported.

What about LaPorte's appeal to semantic vagueness? We might interpret it as another argument for stipulation. According to LaPorte's diagnosis, the biologists' epistemological disagreement is rooted in the semantic vagueness of terms such as 'species' and '*Panthera tigris*'. The cure for this vagueness is to refine these terms by modifying their meaning. LaPorte claims that this can legitimately be done 'in any of

several ways' (90): after refining, the term 'species' can refer to groups defined by either the BSC or the PSC, and the extension of '*Panthera tigris'* can include Sumatran tigers or not.

Unfortunately, I do not think that this argument has independent standing. If we are not already convinced by LaPorte's argument from diverse classificatory interests, then the argument from semantic vagueness will lack force. The latter is limited by the fact that LaPorte's diagnosis does not imply his cure. We can agree with him that the biologists' epistemological disagreement is rooted in the semantic vagueness of terms such as 'species' and '*Panthera tigris'*. But this does not license biologists to stipulate in response – unless we are already committed to a philosophical theory that attributes this vagueness to semantic indecision or context sensitivity. As I pointed out in Chapter 2, there are at least two other theories of vagueness, based on irremediable ignorance and metaphysical indeterminacy. Neither of them warrants this semantic refinement by stipulation.

I note, *ad hominem*, that LaPorte is explicitly not committed to 'any particular account of the truth value of vague statements' (193).[28] He maintains that his position on vague natural-kind terms is compatible with different theories of vagueness, including epistemicism. So he cannot be depending on semantic indecision or context sensitivity to create room for stipulation in biological classification. In fact, LaPorte does not present his views on semantic vagueness and

[28] This point is reiterated in LaPorte (2010), 122, note 13. It puts into question some interpretations of LaPorte's views. See, for instance, Bird (2010), who attributes this more robust conception of vagueness to LaPorte: 'For a kind term "*K*", some things will be determinately *K* and other things will be determinately not *K*. But there will be a boundary of things for which there is no determinate fact of the matter whether they are *K* or not. This means that there will be no determinate fact of the matter that $K_1 = K_2$ for distinct kind terms "K_1" and "K_2"' (125).

semantic refinement as an independent line of argument. Rather they are used to help him make semantic sense of his hypothesis that scientists' conclusions about species essences depend on stipulations. And they are used to contest Kripke's position that the meanings of natural-kind terms do not change with scientific progress. So, at most, the semantic claims are meant to support his main argument about classificatory interests.

I conclude that, on their own, LaPorte's arguments do not show that biologists can solve their multiple kinds problem by stipulation. I draw two other lessons from his argument from diverse classificatory interests. First, the argument helps us to clarify the character of biologists' epistemological disagreement on Species Concepts. Their disagreement reflects classificatory interests that are both rivalrous and recalcitrant. Second, it alerts us to new dimensions of epistemological complacency. When we investigate the conditions under which stipulation can be used, we should not neglect the role of empirical interests. And we should not assume, without evidence, that scientists cannot develop a method to assess and balance their non-empirical interests properly.

5.4 Beyond the epistemological stalemate

I want to make a new argument for stipulation, which avoids the above dimensions of epistemological complacency. This argument requires us to look a little beyond the epistemological stalemate that LaPorte describes. I shall highlight two features of biological practice that are obscured by his account. First, biologists tend to choose

Species Concepts according to their empirical interests. Second, they use some of these concepts successfully in different research programmes. These features suggest that, under some conditions, different groups of biologists can legitimately choose different Species Concepts on the basis of their empirical interests. I will elaborate my argument, then evaluate it for strengths and weaknesses.

Here I draw on some advice from Cracraft, the biologist who conceptualised the Phylogenetic Species Concept. Surveying the controversy over Species Concepts, he picks out a 'consistent refrain' that 'one's own concept is the best for addressing the taxonomic diversity of nature' (2000: 9). He lists examples from various proponents of the Biological Species Concept, the Phylogenetic Species Concept, and other concepts. His description of the controversy resembles LaPorte's, which has various proponents claiming that their concept is the correct one for biology. However, Cracraft warns that 'the notion of "best" is always relative' (10). He urges us to 'look hard at the context of what *best* might mean'.[29] This requires us to ask how general in application a concept is meant to be, and whether a more general concept is always more useful. The official definitions may be misleading. So we should study 'the discussion surrounding, supporting, and justifying' the Species Concepts, and see how they are 'applied in the real world' (13).

What happens when we apply Cracraft's advice? I find two features of biological practice, which are neglected when we focus on LaPorte's epistemological

[29] See also his earlier recommendation in Cracraft (2000), which might resonate with both scientists and philosophers: 'All discussions about species should be approached with skepticism, with a critical mind for the nuances of language, of debating ploys, and an appreciation that arguments and conclusions, while using the same words, might not mean the same thing because those words imply different things to different people and because people argue from different premises. Keeping all of this in mind will be the only way one can begin to make sense of these debates' (5).

stalemate. The first feature concerns how biologists decide on a Species Concept. In response to the controversy, biologists usually choose a concept according to their empirical interests. The biologists Coyne and Orr (2004) vividly illustrate this practice. Although they favour the BSC, they recognise that others with different empirical interests use different concepts. Coyne and Orr distinguish two main empirical interests: some biologists wish to reconstruct the 'history of life', while others want to discover the 'origin of discrete groups in nature' (281). Those who choose the PSC tend to reject the BSC because they believe that reproductive isolation is 'largely irrelevant to reconstructing history'. On the other hand, those who favour the BSC over the PSC see phylogeny as 'largely irrelevant to understanding the discreteness of nature.'

According to Coyne and Orr, biologists who want to explain biodiversity prefer concepts other than the PSC, such as the BSC, the Cohesion Species Concept (CSC), and the Ecological Species Concept (CSC).[30] Sometimes they decide based on the type of *organisms* which interests them. For instance, the BSC is only applicable to sexually reproducing populations. Biologists who are interested in classifying populations that reproduce asexually must choose other concepts such as the CSC and the ESC.[31] Coyne and Orr doubt that it is better to use a single Species Concept for both types of organisms; doing so may impede scientific progress if 'the processes of cluster formation differ between these two groups' (278). At other

[30] The Cohesion Species Concept defines a species as 'the most inclusive population of individuals having the potential for phenotypic cohesion through intrinsic cohesion mechanisms' (Coyne and Orr, 277), while the Ecological Species Concept defines it as 'a lineage (or a closely related set of lineages) which occupies an adaptive zone minimally different from that of any other lineage in its range and which evolves separately from all lineages outside its range' (280).

[31] In choosing these Species Concepts, biologists often stress the significance of asexual reproduction. Most organisms do not reproduce sexually; those who reproduce sexually include some who can also reproduce asexually.

times, biologists decide based on the type of *processes*. Those who are interested in sexual reproduction use the BSC, while those who investigate how hybridization works across different environments use the ESC (280).

In Coyne and Orr's interpretation, each concept is a solution to a different 'species problem' (272). Each species problem reflects a different set of empirical interests – whether they be interests in ancestral and parental relations, sexual reproduction, environmental influences, or genetic mechanisms. To solve a species problem is, therefore, to find a concept that fits a set of empirical interests. For Coyne and Orr, a Species Concept is 'a tool for research, not a hypothesis subject to refutation' (275). Biologists choose different concepts because they want tools for different research tasks; there is no single Species Concept correct for all of biology. This talk of tools is suggestive, but I will not depend on it.[32] Rather my interest lies in their description of how biologists choose a species concept.

Significantly, even biologists who do not share Coyne and Orr's interpretation agree with them about this feature of biological practice. De Queiroz (1999, 2007) proposes to unify the different Species Concepts into a single concept. His General Lineage Concept (GLC) defines species as '(segments of) separately evolving metapopulation lineages' (2007: 881). This is meant to capture the 'common element' in all Species Concepts. However, the GLC allows biologists to specify 'secondary defining properties' for the relevant lineages, such as reproductive isolation, occupation of a minimally distinct adaptive zone, or monophyly and

[32] See also Kitcher (1984), who highlights the tension between Ernst Mayr's claim that the BSC defines the fundamental basis of biological diversity and his defence of it as a 'valuable *instrument*' (119, original italics).

characters sufficient to form a diagnosable cluster. These are the same properties highlighted, separately, by other Species Concepts such as the BSC, the ESC, or Cracraft's PSC.

According to de Queiroz, it is no surprise that biologists focus on different properties.

> Although all modern biologists equate species with segments of population lineages, their interests are diverse. Consequently, they differ with regard to the properties of lineage segments that they consider most important, which is reflected in their preferences concerning species criteria. Not surprisingly, the properties that different biologists consider most important are related to their areas of study. (1999: 65)

So biologists with different 'areas of study' pursue different empirical interests, which motivate them to focus on different properties. For instance, those who are interested in how sexual reproduction works focus on reproductive isolation. Ecologists emphasise adaptive zones, while systematists emphasise monophyly and diagnosability. In de Queiroz's analysis, the BSC, the ESC, and the PSC do not count as distinct concepts. Rather their definitions help biologists to specify different 'operational criteria' or 'lines of evidence' for applying the General Lineage Concept.[33] But I emphasise that, in one crucial respect, he clearly agrees with Coyne and Orr: biologists choose between the BSC, the ESC, and the PSC on the basis of their empirical interests.

The other feature of practice I want to highlight is how biologists use their Species

[33] This point is defended in de Queiroz (2007), 882, and Richards (2010), ch. 5. De Queiroz (2007) adds that the properties mentioned in these operational criteria can also be used to 'define subcategories of the species category' (882). His examples of subcategories are 'reproductively isolated species, ecologically differentiated species, and monophyletic species'.

Concepts. After choosing different concepts on the basis of their empirical interests, biologists use some of them successfully in different research programmes. So the stalemate over which Species Concept is best for biology has hardly inhibited its progress. Instead biologists have created at least three major research programmes: these rely on the Biological Species Concept, the Ecological Species Concept, and the Phylogenetic Species Concept.

According to Ereshefsky (1992, 2010a), biologists use each concept to explore a different causal process – respectively, the process of sexual reproduction, environmental selection pressures, and the process of descent from common ancestry. Stanford (1995) cites the specific explanatory projects made possible by the BSC and the ESC. On the one hand, the BSC helps to investigate 'the phenomenon of gene flow, speciation events in peripheral isolates, founder effects, and the nature and adaptive value of reproductive isolating mechanisms themselves' (73). On the other hand, the ESC helps in an 'ecological explanatory enterprise'. For instance, some biologists may be interested in explaining the different ecological niches occupied by populations that form hybrids. Coyne and Orr (2004) describe how biologists use different versions of the PSC to make inferences about the genealogy of populations, although they also highlight methodological problems associated with each version.

I note that LaPorte also acknowledges the successful use of different Species Concepts, though it plays no obvious part in his arguments for stipulation. According to LaPorte, the BSC and the PSC classify populations into 'natural and scientifically interesting' groups (73). This implies that they already produce biological

classifications with significant explanatory value.

My argument for stipulation draws on both features of biological practice. Faced with their disagreement on what a species is, biologists choose different Species Concepts and thereby stipulate the reference of individual species concepts. They do so according to their empirical interests, rather than any further evidence about the nature of species. Then they use some of these concepts successfully in different research programmes. Together these practices entail that biologists make *multiple successful stipulations* on the basis of their empirical interests. I claim that, from the epistemological perspective, the best explanation for these multiple successful stipulations is that there is a role for stipulation in biological classification – that is, biologists can legitimately choose between the BSC, the ESC, and the PSC, and thereby stipulate the reference of individual species concepts.

To complete this argument, I need to exclude two alternative explanations. They are deflationary in this sense: they try to explain the same biological practices as I do without leaving room for stipulation. In the first explanation, the deflationist appeals to *error*. He suggests that some or all biologists who choose the Biological Species Concept, the Ecological Species Concept, or the Phylogenetic Species Concept have erred. By heeding their empirical interests, these biologists have picked a wrong concept. According to the deflationist, nothing in biological practice favours my account over his. Indeed, my account seems to depend on something like the naturalistic fallacy. As the deflationist points out, the fact that some biologists choose the BSC, others the ESC, and yet others the PSC, does not make their choices legitimate. We cannot infer a right to stipulate simply from the biologists' readiness

to stipulate.

I think this deflationary explanation is flawed. From the epistemological perspective, it might be consistent with the biologists' choices. But what we seek to explain is not just the fact that biologists choose the BSC, the ESC, and the PSC according to their empirical interests, but also the fact that their choices are so successful. If the deflationist's explanation depends only on the biologists' errors, then he misjudges the explananda. He has yet to explain why, if either the BSC, the ESC, or the PSC is the wrong concept, biologists are able to use it successfully in a research programme. Unless he can do so, his account is not as good as one which concedes some room for stipulation. Allowing that the BSC, the ESC, and the PSC are legitimate choices addresses the epistemological mystery surrounding the success of the biologists' stipulations.

The deflationist also misunderstands the basis of my account. It does not depend on the naturalistic fallacy: I do not infer that biologists have a right to stipulate simply from the fact that they do stipulate. Rather I aim to interpret the biologists' practice of choosing the BSC, the ESC, and the PSC according to their empirical interests – in the light of the surprising successes of their choices. These successes are surprising precisely because the choices are based on biologists' empirical interests, rather than further evidence about the nature of species.

At this point, the deflationist might object: on what grounds do we attribute success to the biologists' choices? Earlier I noted that research programmes which rely on the BSC, the ESC, and the PSC have produced biological explanations about the

process of sexual reproduction, environmental selection pressures, and the process of descent from common ancestry. I believe this is success enough for my argument to work. The biologists themselves count these explanations of different species as solid scientific achievements. Moreover, the biological explanations fit what we know about natural kinds in the Kripke-Putnam model. Recall that, in this model, natural kinds help to explain the world because they reflect the world's basic structures. In the research programmes cited above, the species defined by the BSC, the ESC, and the PSC ground their biological explanations.

The second explanation by the deflationist appeals to *luck*. In this explanation, the deflationist acknowledges the biologists' success, but explains it away as their good fortune. He claims that, by heeding their empirical interests, some biologists have lucked into the correct concept. Consider those who choose the Biological Species Concept rather than the Ecological Species Concept and the Phylogenetic Species Concept, because they are interested in sexual reproduction. We see that their choice leads to explanatory success. This may tempt us into supposing that there is room for biologists to choose a Species Concept on the basis of their interests. But the deflationist argues that this is illusory: the biologists who choose the BSC have simply stumbled on the correct concept.

I think that, by focusing on the biologists who choose the BSC, the deflationist again misjudges the explananda. We do not only seek to explain the success of those who choose the BSC. What we have to explain is the distinct successes of the biologists who choose the BSC, the ESC, and the PSC. In three wide-ranging research programmes, biologists have found explanatory success by choosing a Species

Concept according to their empirical interests. The deflationist's appeal to luck is less plausible once we remind ourselves of the biologists' repeated successes. To stumble on one correct concept may be regarded as good fortune; to stumble in the same way on three looks like something more mysterious from the epistemological perspective. By allowing some room for stipulation, my account goes some way towards solving this mystery. Of course, it remains possible that the biologists who choose the BSC, the ESC, and the PSC have only stumbled on concepts that they can legitimately choose on a different basis. I cannot rule this out with the epistemological resources available, but the burden of proof is now on the deflationist to explain what this basis is.

The deflationist might object: on what grounds do we neglect the failure of other biologists' choices? Unlike the BSC, the ESC, and the PSC, some Species Concepts chosen by biologists are not used in successful explanatory programmes. One example is the Genotypic Cluster Species Concept (GCSC).[34] By comparing successful stipulations with failed ones, the deflationist wants to convince us that the biologists who choose the BSC, the ESC, and the PSC have been lucky compared to those who deploy other concepts. Here the deflationist misunderstands the conclusion of my argument. For I conclude only that biologists can legitimately choose the BSC, the ESC, and the PSC on the basis of their empirical interests. I do not thereby infer that biologists have room to choose *any* Species Concept. So the biologists who choose the BSC, the ESC, and the PSC may have been lucky compared to those who choose otherwise. But their luck in this sense is not relevant to my conclusion.

I can now distinguish my argument from what Norton (2008) calls 'the gap argument'. Those who make this argument typically start by highlighting a gap

[34] The GCSC defines a species as 'a (morphologically or genetically) distinguishable group of individuals that has few or no intermediates when in contact with other such clusters' (Coyne & Orr 2004: 273). Coyne and Orr note that the GCSC 'does not yield a particularly fruitful program of research' because it focuses on 'the identification rather than the origin of species' (276). Its lack of 'widespread support' is noted by Mallet (2007), one of its developers.

between the content of scientific theories and the evidence for them.[35] So the gap reflects an underdetermination of theory by evidence (Stanford 2013). In Norton's analysis, the next steps in the argument are as follows:

> My favorite social, cultural, political, ideological, or other factor is able to account for what fills the gap. Therefore, my favorite factor accounts for a portion of the content of our mature scientific theories. (18)

Often this conclusion is construed in terms of a 'contingency' in scientific knowledge. For instance, it might be claimed that the concepts used in scientific theories could be otherwise if the social or political factor were otherwise. More recently, some philosophers have used the gap argument to ascribe a 'social dimension' or 'social character' to scientific knowledge (Longino 2013).

From the epistemological perspective, I find two serious limitations in the gap argument. First, the argument does not show that – if the social or political factor were otherwise – the different concepts in scientific theories would be free from error. So, on the basis of this argument alone, we cannot claim that scientists can legitimately use another set of concepts. Second, the argument fails to show that luck is not involved. It may be that, motivated by the social or political factor, scientists have stumbled onto the correct concepts. So we cannot conclude that scientists can legitimately change their concepts to fit the social or political context.

As I have indicated, my argument from multiple successful stipulations overcomes

[35] I owe this comparison to Philip Kitcher, who mentioned it in a conference discussion. For examples in the philosophical literature on social epistemology, see Longino (2002), ch. 6, Longino (2004), and Machamer and Osbeck (2004). The gap argument also appears in social studies of science; see Woolgar (1988), ch. 4. Kitcher noted the same argument in Shapin and Schaffer (1986), a history of early modern science.

both limitations. It does bear a superficial resemblance to the gap argument: my argument rests partly on the fact that empirical interests influence how the biologists choose their Species Concepts. But, by pointing to the explanatory success that follows from biologists choosing the BSC, the ESC, or the PSC, I show that each choice is not erroneous. Then, by emphasising the multiplicity and uniformity of the biologists' successes, I suggest that these successes from choosing on the basis of empirical interests are not fortuitous.

Let me conclude this section by analysing the strengths and weaknesses of my argument for stipulation. It has three strengths, which I will highlight by comparison with LaPorte's arguments. First, my argument avoids epistemological complacency. Like LaPorte, I do not assume that biologists can solve the multiple kinds problem by stipulation. Instead, I assemble evidence from the epistemological perspective that supports the use of stipulation. This evidence is drawn from biological practice: in particular, I note that biologists choose Species Concepts according to their empirical interests, and that they use the Biological Species Concept, the Ecological Species Concept, and the Phylogenetic Species Concept successfully in different research programmes. Then, on the basis of this evidence, I make an inference to the best explanation.

Earlier, I denied that my argument depends on the naturalistic fallacy. But my argument is naturalistic in a different, and non-fallacious, sense. Like LaPorte and other philosophers of science, I attend to biologists' practices – including their claims on the nature of species, their methods for deriving those claims, and their uses of those claims. Although I do not assume that these claims, methods, and uses are

correct, I aim for a naturalised epistemology of species that defers to biologists where possible and makes the best overall sense of their practices. In particular, my argument tries to make sense of the fact that, despite the controversy over Species Concepts, biologists make multiple successful stipulations on the basis of their empirical interests. In contrast, LaPorte's main argument for stipulation is based more narrowly on the fact that the biologists' controversy is motivated by diverse classificatory interests.

Second, by avoiding epistemological complacency, my argument clarifies the conditions for stipulation and the basis of stipulation under those conditions. I demonstrate a role for stipulation when biologists face a choice between more than one Species Concept. However, each of these Species Concepts must belong to a successful research programme. This condition excludes Species Concepts that are not useful in biological explanations.[36] Correspondingly, biologists have room to stipulate the reference of an individual species concept when they face a choice between multiple kinds. But each kind must reflect a distinctive set of biological structures and ground a set of biological explanations. Under such conditions, biologists can legitimately stipulate on the basis of their empirical interests. That is, they can legitimately choose a Species Concept that fits their empirical interests, and thereby stipulate the reference of individual species concepts.

LaPorte's arguments focus on how biologists choose between one 'natural,

[36] In contrast, other promising concepts include those used to classify the microbial world; see O'Malley and Dupré (2007), and Ereshefsky (2010e). Of course, biologists who develop a new Species Concept cannot know in advance if it will be used in a successful research programme. Likewise, we cannot know in advance if there is room for biologists to choose a particular Species Concept. My aim here is to make the best overall sense of biological practice, not to predict its successes. I thank Lim Chong Ming for discussion on this point.

historical' group of organisms and another (70). Like me, he is proposing that biologists stipulate between Species Concepts which classify populations into different groups that are 'natural and scientifically interesting' (73). However, he does not specify the appropriate basis for their stipulation. He cites the biologists' empirical and non-empirical interests only as the source of their epistemological disagreement on what a species is.

Third, my argument is consistent with the emerging philosophical consensus in favour of species pluralism.[37] According to Ereshefsky (2010b), pluralists deny that there is a single correct Species Concept: 'Biology…contains a number of legitimate Species Concepts' (§3). If I am right that biologists can legitimately choose the BSC, the ESC, and the PSC on the basis of their empirical interests, then biology has at least three legitimate Species Concepts now. Indeed, we can construe my argument as a modest plea for species pluralism, which is made from the epistemological perspective.

It is illuminating to contrast my argument for pluralism with two others. LaPorte's argument also supports the claim that there is more than one legitimate Species Concept.[38] But, according to his argument, all biologists can legitimately choose the same Species Concept – whether it be the BSC or the PSC (74). Other philosophers such as Kitcher (1984) and Stanford (1995) seem to disagree with LaPorte. Their

[37] For a philosophical survey on species pluralism, see Ereshefsky (2010b), §3. Pluralists disagree on the range of legitimate Species Concepts (Dupré 1981, 1993; Kitcher 1984; Ereshefsky 1992), the epistemological and ontological reasons for pluralism (Kitcher 1984; Stanford 1995; Ereshefsky 2010b, §3.1), and the implications of pluralism for realism about species (Kitcher 1984 and Stanford 1995; Ereshefsky 1992 and Brigandt 2003; Devitt 2011).

[38] In his analysis, LaPorte does not adopt the traditional definition of species pluralism. He construes it instead as a distinct Species Concept: 'Pluralism is a "species concept" that offers still another possibility' (74). See Dupré (2004), who alleges that he is 'a pluralist in occasional denial'.

argument for pluralism is based on the claim that different explanatory demands in biology require different Species Concepts: as Stanford concludes, 'we have independent and legitimate explanatory interests in biology which require distinct concepts of species' (76). This suggests that biologists with different explanatory interests cannot use the same Species Concept. At least they cannot do so without much difficulty.[39] I am agnostic on Kitcher and Stanford's apparent disagreement with LaPorte. I doubt that the practices cited in my argument can be used to settle it. At any rate, I have shown that pluralism can be defended without settling it.

What about the weaknesses of my argument for stipulation? I find three main weaknesses. First, I cannot say more to convince anyone who appeals to luck to explain away the biologists' multiple successful stipulations. At most I have shown that this appeal is implausible, given the multiplicity and uniformity of the biologists' successes. Second, I have not addressed the risk of metaphysical caprice. When biologists choose different Species Concepts in accord with their empirical interests, they stipulate different referents for each individual species concept, such as the concept *Panthera tigris*. How does this stipulated species reflect the world's

[39] A caveat: Kitcher and Stanford's claims are inflected by a pragmatism that makes it unclear how far they genuinely disagree with LaPorte. (a) For instance, Stanford (1995) adds that biologists '*could*, of course, simply "stick to our guns"' (76): they could insist on using one species concept, while creating subcategories of species. He complains that this strategy would 'hobble significant investigations in biology' (77). I note that he does not say it would inhibit these investigations.

(b) Stanford (1995) interprets Kitcher (1984) to be claiming that 'certain explanatory demands are *inextricably bound* to certain species concepts': each concept 'represents a legitimate biological explanatory demand and that each demand *necessitates* the use of its corresponding species concept' (72, italics in original). I think these italicised relations need to be qualified. Kitcher (1984) does claim that 'biology needs a number of different approaches to the division of organisms, a number of different sets of "species"' (120). But his construal of this claim appears to be weaker than Stanford's: 'each main type of biological investigation subdivides further into inquiries that are *best conducted* by taking alternative views of the species category' (121, my italics).

(c) In Stanford and Kitcher (2000), they agree on an even weaker claim: 'We suggest that, given different, legitimate, biological interests, there are alternative ways to group organisms…If we are right, then there are several distinct ways to divide up the living world, corresponding to different choices about how to extend the application of a species concept from type specimens' (122).

structures rather than the biologists' interests? Third, I have not addressed the risk of semantic confusion. What happens to the species term *'Panthera tigris'* when biologists stipulate different referents for the corresponding concept? How do they minimise terminological confusion and avoid abruptly changing their subject?

These weaknesses arise because my argument for stipulation is made largely from the epistemological perspective. Earlier I emphasised that – from the epistemological perspective – the most plausible explanation for the biologists' multiple successful stipulations is that there is a role for stipulation in biological classification. But I have yet to interpret stipulation's result from the metaphysical and semantic perspectives. Until I do so, my argument makes only a *prima facie* case for stipulation in biological classification. If I can do so, then luck may also lose its remaining appeal.

5.5 Making sense of stipulation

I have offered a new argument for practical indeterminacy about species essences. My argument draws on biological practices to show that biologists can legitimately choose different Species Concepts, and thereby stipulate the reference of individual species concepts. It also clarifies the conditions for stipulation and the basis for stipulation under those conditions. So I have addressed the risk of epistemological complacency. However, as I noted above, I have yet to respond to the risks of metaphysical caprice and semantic confusion.

To do so, I need to make sense of empirically constrained stipulation from the metaphysical and semantic perspectives. First, I explain how stipulation enables different groups of biologists to focus on the evolutionary structures which interest them. Here I draw on the same empirical theory of evolution that motivates pluralism about Species Concepts. I also distinguish two metaphysical senses in which species count as real kinds, and demonstrate that stipulation does not affect the reality of species in either sense. Second, I explain how stipulation produces a context-sensitive species term. Following Stanford and Kitcher (2000), I distinguish two semantic models for species terms: one model allows for partial reference, the other for contextual reference. In both models, stipulation minimises terminological confusion between the biologists who focus on different evolutionary structures, and avoids abruptly changing the subject of their classification.

Let me start with the metaphysics of stipulation. We know enough about the world's biological structures to explain how stipulated species reflect them. First, as I emphasised in the previous section, the biologists' room for stipulation is limited – thus far – to the Biological Species Concept, the Ecological Species Concept, and the Phylogenetic Species Concept. So biologists cannot choose Species Concepts at whim. The BSC, the ESC, and the PSC are all used in successful research programmes to explore the world's biological structures. Each research programme centres on a different causal process. Correspondingly, the species classified by the BSC, the ESC, and the PSC are defined in terms of these causal processes, not the biologists' whim.

Second, the biologists' room for stipulation fits everything we know about the world. Our best available science tells us that evolutionary structures are produced by a set

of causal processes. This set includes the very same processes cited in the BSC, the ESC, and the PSC – respectively, sexual reproduction, environmental selection, and descent from common ancestry. During evolution, these processes interact in a complex way. They can come into conflict, reinforce each other, or operate in isolation. Neither sexual reproduction, environmental selection, nor descent from common ancestry takes priority within the evolutionary structures. But, together, these processes account for the diverse populations that biologists want to classify.

It so happens that, among the populations that superficially resemble each other, each causal process is associated with a different lineage of these populations. One lineage consists of interbreeding populations, another consists of populations responding to the same environmental pressures, another consists of populations with a common ancestry. Yet these lineages overlap: some populations belong to some, but not all, of the lineages. Such populations, including LaPorte's Sumatran tigers, show up during classification as borderline cases of a species.

Third, the basis of stipulation used by the biologists also fits what we know about the world. Their interests in different causal processes determine the evolutionary structures which they want to investigate. Biologists who are interested in sexual reproduction use the BSC to investigate the evolutionary structures associated with sexual reproduction. Those who are interested in environmental selection use the ESC to study the environment's impact on evolutionary structures. And those who are interested in genealogical relations use the PSC to study the role of descent. Sometimes the type of organisms in which biologists are interested influences the evolutionary structures on which they focus. For instance, it is difficult to reconstruct

the patterns of sexual reproduction in some extinct populations. As a result, biologists who are interested in such populations tend to study structures that are not associated with sexual reproduction.

I claim that biologists can legitimately choose a Species Concept on the basis of their empirical interests. This entails that biologists with different empirical interests can legitimately use different species classifications.[40] In this sense, the classifications variously defined by the Biological Species Concept, the Ecological Species Concept, and the Phylogenetic Species Concept do reflect *biological interests*. But the species identified within these classifications do not thereby fail to reflect the world's *biological structures*. In every classification, the essence of each species is determined by a causal process that is responsible for some evolutionary structures. No metaphysical caprice is indulged: no essences have been created out of biological interests. When different groups of biologists stipulate different referents for an individual species concept, they are simply dividing their attention between the different structures that interest them. What matters to them in explaining evolutionary structures is distinguishing the contributions of the causal processes, not identifying one of the processes as *the* basis for defining all species.

Here it may help to borrow an analogy from LaPorte (2004), though he is referring to higher taxa rather than species. According to LaPorte, different systems of biological classification 'just reflect different information' (80). There is 'no fact of the matter' about which system is best for all of biology. He compares the different systems of

[40] Kitcher (1984) is prescient on this point. He draws up a partly speculative taxonomy of species concepts. Then he suggests that this taxonomy 'already helps us to see how different views of species may be produced by different biological priorities' (124).

classification to the 'number of legitimate maps for a given location, each representing certain features of the represented location at the expense of others'. Some maps, for instance, represent transport options in an area, while others highlight its weather patterns. Following LaPorte's suggestion, we might think of the classifications based on different Species Concepts as maps that represent different evolutionary structures. Each group of biologists, interested in a different set of structures, comes to rely on a different map.

I have been appealing to the same view of evolutionary structures that motivates species pluralists such as Ereshefsky, Kitcher, and Stanford. Ereshefsky (2010b) sums up this view well: 'Evolutionary theory, a well substantiated theory, tells us that the organic world is multifaceted.' Each facet of this world corresponds to what I describe as the evolutionary structures associated with a specific causal process. In Ereshefsky (1992), he spells out this view more fully:

> All of the organisms on this planet belong to a single genealogical tree. The forces of evolution segment that tree into a number of different types of lineages, often causing the same organism to belong to more than one type of lineage. The evolutionary forces at work here include interbreeding, selection, genetic homeostasis, common descent, and developmental canalisation….So the forces of evolution segment the tree of life into a plurality of incompatible taxonomies: one taxonomy consisting of interbreeding units, another consisting of ecological units, and a third consisting of monophyletic taxa. (676-7)

These three taxonomies are the species classifications defined, respectively, by the BSC, the ESC, and the PSC. Ereshefsky reminds us that this view of evolutionary structures is fallible since it rests on 'empirical matters'. For instance, it may turn out to be wrong in some respects: 'perhaps some of the above-mentioned forces do not

exist, or those forces lack the ability to produce stable taxonomic entities' (677).

Citing what 'our best available science' says about evolutionary structures is in keeping with my naturalistic stance. As I explained earlier, I am trying to make the best overall sense of how biologists classify populations into species – in the light of what biologists actually claim about species. Current biology tells us that the world's biodiversity does not rest on one fundamental basis. Rather it arises from the complex interaction of at least three different processes. Of course, as Ereshefsky emphasises, current biology may turn out to be wrong. I concede that biologists may yet discover a more fundamental essence underlying the processes of sexual reproduction, environmental selection, and descent from common ancestry, though this possibility seems remote. Nothing in my argument for stipulation rules out this discovery.

If such a discovery were made, then biologists might choose to define species in terms of the more fundamental essence.[41] This would offer a reason for biologists to accept De Queiroz's General Lineage Concept, which interprets the Biological Species Concept, the Ecological Species Concept, and the Phylogenetic Species Concept as operational criteria for identifying species.[42] Accordingly, I would have to re-interpret the biologists' room for stipulation: before the discovery of this fundamental basis, biologists could legitimately choose a Species Concept on the basis of their empirical interests. In doing so, they would be stipulating the reference

[41] I thank Vasso Kindi and Gabor Zemplén, who urged me to clarify this possibility during a conference discussion.

[42] See my analysis of the General Lineage Concept in §6.4. Richards (2010) emphasises that the GLC's plausibility depends on 'how universal the theoretical concept based on the idea of segments of a population lineage can be' (142). I take it that this, in turn, depends on what our best theory of evolution tells us about the correct basis for segmenting population lineages.

of individual species concepts. Although these stipulations turned out to be erroneous, they would have helped biologists to discover the true essences of individual species.

I have shown that, by choosing different Species Concepts, different groups of biologists focus on the evolutionary structures which interest them. My account makes it implausible that the biologists who choose the BSC, the ESC, and the PSC on the basis of their empirical interests have only stumbled on concepts that they can legitimately choose on a hitherto unknown basis. It also implies that the species they stipulate are *natural* kinds. At least they count as natural in the Kripke-Putnam model, which requires that natural kinds reflect the world's basic structures. Now I want to consider an ostensibly different metaphysical challenge: are these species *real* kinds? This challenge is complicated by the fact that philosophers disagree on what makes a natural kind real.

Bird and Tobin (2015) help us by distinguishing two versions of realism about natural kinds.[43] Weak realism, or naturalism, requires only that these kinds demarcate natural divisions in the world. Such divisions must reflect 'the structure of the natural world rather than the interests and actions of human beings'.[44] On the other hand, strong realism requires that the kinds exist as 'a special sort of entity in our ontology' – distinct from the individuals in these kinds, and independent of human

[43] This distinction is elaborated in Bird and Tobin (2015), §1.1.1 on 'Naturalism (weak realism)' and §1.2 on 'Natural Kind Realism'. Citing Kitcher, Dupré, and Ereshefsky, they claim that 'pluralism is the weakest form of realism about kinds in the philosophy of biology' (§2.1.1). See also Bird (2009), 503-4.
[44] Putnam (1975) brings up this version of realism in a section entitled 'Let's be realistic' (235-8). Here is how he characterises anti-realism: 'To a strongly anti-realist intuition it makes little sense to say that what is in the extension of Archimedes' term *χρυσός* is to be determined using *our* theory. For the antirealist does not see our theory and Archimedes' theory as two approximately correct descriptions of some fixed-realm of theory-independent entities' (236). I owe this point to Bill Child.

minds. Weak realism does not entail strong realism. To clarify this point, Bird and Tobin compare strong realism about kinds to realism about universals. One can reject realism about universals, and yet be committed to the genuine similarities and differences between things. That is, indeed, the position held traditionally by the nominalist. Similarly, one can deny that kinds exist as real entities, and yet be committed to the natural divisions among things.

For the weak realist, the biologists' stipulated species count as real. He requires only that these species be natural in the sense required in the Kripke-Putnam model.[45] As I explained earlier, the classifications chosen by different groups of biologists reflect their interests in different causal processes and evolutionary structures. But this does not entail that the species identified within these classifications are constituted by 'the interests and actions' of the biologists. Rather, in each classification, the essences of species are determined by the various causal processes in which biologists are interested. Thus these species still reflect 'the structure of the natural world'. Kitcher (1984) defends a similar stance on realism under the label 'pluralistic realism'. According to Kitcher, the reality of species is about 'whether the division of organisms into species corresponds to something in the *objective structure* of nature' (128, original italics). It is therefore possible for biologists to focus on different, equally objective, structures by classifying the same populations into different species: more generally, 'the patterning of nature generated in different

[45] I add that this is how some biologists understand what it means to claim that a species is real. See, for instance, Cracraft (2000), 11: 'definitions do not necessarily make things real, and some definitions might actually lead us to identify spurious ("unreal") entities. In what sense might this be true? By saying that something is a real discrete entity, we imply that it *participates in one or more natural processes*' (my italics). On the other hand, Mallett (1995) warns that 'real' and 'reality' are 'tricky words' used in more than one way when biologists debate species concepts (2). He uses it to mean that species have something like an 'essence' – 'a hypothetical pure, albeit obscure, truth that underlies the messy actuality'.

areas of biology may cross-classify the constituents of nature'.

What of strong realism about these species? Some of Kitcher's illuminating remarks relate to this question:

> …if realism about species is construed as the bare claim that species exist independently of human cognizance of them, then anyone who accepts a modest realism about sets can endorse realism about species. Organisms exist and so do sets of those organisms. The particular sets of organisms that are species exist independently of human cognition. So realism about species is trivially true. (1984: 128)

Whereas Kitcher interprets weak realist claims about species to be meaningful theses that 'cry out for analysis', he treats strong realist claims as 'trivially true'. According to him, species are nothing but sets of organisms. Like all sets, or at least the theoretically significant ones, species are real entities that 'exist independently of human cognizance of them'. But this strong realism about species is too 'cheap': it fails to take heed of 'what provokes biologists and philosophers to wave banners for the objectivity of systematics' (128). In Kitcher's judgement, weak realism captures better what these biologists and philosophers mean by the reality of species.

For some philosophers, the resemblance between strong realism about kinds and realism about universals makes the former as metaphysically suspect as the latter. Devitt (2011) says: 'I follow Quine in thinking it arises from a pseudo-problem…Indeed, it seems to me best to discuss the issues of realism in biology without any commitment to this vexed, millennia-old, metaphysical issue' (158).[46] In contrast, Hawley and Bird (2011) are especially interested in the 'metaphysical status' of natural kinds in our ontology: 'we take the kindhood question seriously, and we examine how best to

[46] More generally, Boyd (1999a) argues that questions about the 'reality' or 'realism' of kinds should not be construed as asking about the 'metaphysical status' of kinds 'considered by themselves' (159). For him, the questions at stake are about 'the accommodation of representational and inferential practices to real causal structures in the world' (15). Slater (2014), §6, defends and extends Boyd's proposal.

develop a realist view of natural kinds' (206). Hawley and Bird do not contest the significance of weak realism, but they consider strong realism to be neither trivial nor dubious. They explore a strong realist view which construes natural kinds as complex universals.

I will remain agnostic on whether, in our best ontology, the biologists' stipulated species count as real entities in the strong sense. Should species be theorised as entities that depend for their existence on human minds? Or should they be treated as sets or universals? These are debates in metaphysics far removed from my philosophical purposes. However, I want to clarify that these debates are not at all affected by my conclusion that biologists can legitimately stipulate the reference of species concepts. As I emphasised above, the biologists' room for stipulation does not entail that the species are constituted by either their stipulations or their interests. Therefore, it does not entail that the stipulated species are mind-dependent entities. It does not exclude them from being mind-independent entities such as sets or universals.

Next I consider the semantics of stipulation. Recall that both Kripke and Putnam agree that their causal-historical model of reference needs to be modified for non-ideal situations. How can we make sense of what happens to a species term upon stipulation? To do so, we need a grasp of the semantic structures that connect our species terms to evolutionary structures. While biology tells us a lot about the evolutionary structures that involve species, linguistics has less to say about the semantic structures that involve species terms. So my interpretation of stipulation from the semantic perspective has to be more speculative than that from the metaphysical perspective.

Here I appeal to two models of scientific terms that were originally developed to make semantic sense of theoretical change.[47] The first model, proposed in Field (1973), allows for a term's *partial reference*. In this model, a token of a scientific term can partially denote more than one thing. For Field, partially denoting terms are referentially indeterminate: 'there is no fact of the matter as to what they denote (if they are singular terms) or as to what their extension is (if they are general terms)' (462). Their indeterminacy is often exposed when scientists make advances during a scientific revolution. Once scientists discover referential indeterminacy in a term, they can refine its denotation. After refinement, each token of the term will fully denote just one of the things which it partially denoted before.

Field's example comes from Isaac Newton's use of the term 'mass'. He argues that each token of this term in Newton's language partially denoted proper mass and partially denoted relativistic mass. Our lack of a sound basis for attributing only proper mass or only relativistic mass to Newton's term suggests that '*there is no fact of the matter* as to which of these quantities he was referring to' (467, original italics).[48] According to Field, the term underwent denotational refinement after physicists accepted Einstein's special theory of relativity. Indeed, it was refined by two groups of physicists in different ways: 'some physicists have refined "mass" into a word for relativistic mass, while others have refined it into a word for proper mass'

[47] These models belong to the philosophical literature on the semantics of theoretical terms in science. Andreas (2013) and Koslicki (2008), section V, survey this literature. I learnt especially from the philosophers who modify Kripke and Putnam's model: these include Kitcher (1978), Boyd (1993), Stanford and Kitcher (2000), and LaPorte (2004), chapter 5. Earlier, I discussed the model in LaPorte (2004), which allows for semantic vagueness and refinement.

[48] His argument for referential indeterminacy in pre-relativity uses of 'mass' is contested in Earman and Fine (1977), and Kitcher (1978), 546. See also my reference in §4.4 to Papineau's position on this issue. I will not try to settle their debate here: my interest is in how Field's model can be used to modify Kripke and Putnam's, not whether he is right to apply it to the term 'mass'.

(479). Field believes that 'denotational refinement is a fairly common feature of scientific revolutions' (480). Moreover, he predicts that many current scientific terms are referentially indeterminate.

The second model, developed by Kitcher (1978), makes room for a term's *contextual reference*. Here different tokens of a scientific term can refer to different things, depending on the contexts in which the tokens are produced. Kitcher's model is largely similar to the Kripke-Putnam model: the reference of a term is usually fixed by an 'initiating event' in which the referent is causally involved, or one in which the referent is picked out by description (537). But he explicitly allows for the reference of different tokens of the same term to be fixed by different events that are historically connected.[49] So these different tokens may have distinct but related referents. Unlike Field's model, this model does not attribute referential indeterminacy to the term. So it does not call for any denotational refinement. Rather it claims that 'the connections of terms to the world are often extended in subsequent uses' through 'continued reapplication and redefinition' (539).

Kitcher uses his model to explain Joseph Priestley's use of the term 'dephlogisticated air'. In Kitcher's account, the reference of this term is contextually sensitive. When Priestley began to use the term 'phlogiston', the reference of its tokens was fixed by an initiating event which preceded Priestley. At this event, the founder of the phlogiston theory attempted to fix the reference of 'phlogiston' by definition, as the

[49] In a footnote, Kripke (1981) acknowledges the following artificiality in his account of initial baptisms: 'it may be hard to say which items constitute the original sample. Gold may have been discovered independently by various people at various times' (139). He adds: 'I do not feel that any such complications will radically alter the picture.' To my knowledge, Kripke never discusses Kitcher's possibility that term-tokens may end up with distinct but related referents.

substance which is emitted in combustion. Because nothing is emitted in all cases of combustion, Priestley's early tokens of 'phlogiston' failed to refer. Correspondingly, his early tokens of 'dephlogisticated air' also failed to refer. However, Kitcher argues that at least some of Priestley's later tokens of 'dephlogisticated air' succeeded in referring to oxygen. Their reference was fixed anew by pointing to the breathable gas produced in Priestley's experiments.

How are these two models applicable to species terms? Stanford and Kitcher (2000) help us part of the way.[50] Surveying the history of biology, they distinguish historical tokens of species terms whose reference was 'clearly fixed by description', and other historical tokens of the same terms whose reference was 'surely fixed through type specimens' (122). They conclude as follows:

> In our judgment, the history of use of species-names is full of differential reference of different tokens of the same type, overlain with partial reference to many of the species divisions we recognize and some more besides. (123)

Stanford and Kitcher also detect contextual reference and partial reference in how contemporary biologists use species terms. When biologists in different disciplinary contexts rely on different species concepts, they end up with different species classifications. In these contexts, 'different tokens of species-names refer differently'

[50] I also learnt from their critical discussion of the Kripke-Putnam model. A minor point of disagreement: I reject their analysis of Field's model, in which they define partial reference as linking 'a term(-type) to different potential referents' (Stanford & Kitcher 2000: 118-9). This seems to conflict with Field's explicit warning in a footnote: 'This definition and that which follows are intended to apply to term *tokens*, not to term *types*. This is necessary in order to distinguish indeterminacy from ambiguity' (Field 1973, 475).

Boyd (1993, 2013) also applies Field's semantic apparatus of partial denotation and denotational refinement in his Homeostatic Property Cluster model of natural kinds. He argues that Field's apparatus was 'one of the first important developments' of naturalistic conceptions of reference and natural kinds (2013, 80).

(122). However, 'exactness' in species classification is not always needed. For instance, biologists need not choose a particular species concept during some 'inter-field communication' (124). In such contexts, the tokens of species terms partially refer to more than one lineage of populations.

I want to extend their analysis, in order to evaluate the impact of stipulation on species terms. In Field's model: before stipulation, a species term is referentially indeterminate. Each token of the term partially denotes one lineage with interbreeding populations, partially denotes one lineage responding to the same environmental pressures, and partially denotes one lineage with a common ancestry. When biologists with shared empirical interests stipulate the reference of a species concept, the corresponding species term undergoes denotational refinement. After stipulation, each token of the term fully denotes just one of the three related lineages. In Kitcher's model: stipulation does not lead to denotational refinement of a species term. Rather the interests that lie behind stipulation help to fix anew the reference of tokens of the term in a disciplinary context. So biologists with different empirical interests refer to different lineages even though they use the same species term. In each context, tokens of the term refer to just one of the three lineages.

Either one or two of the above models may be needed to account for the semantics of a species term during stipulation. I cannot say more without examining the detailed history of a species term: sometimes the term requires denotational refinement, sometimes it does not. It depends on when and how the term is introduced by some biologists, and then spread to others. Sometimes all the tokens of the term refer to one specific lineage of populations, sometimes they refer to

different lineages. As Stanford and Kitcher suggest, the semantic structure of a species term can involve both partial reference and contextual reference. What matters to me is that both Field and Kitcher's models result in a context-sensitive species term: different tokens can refer to different lineages, according to the disciplinary contexts in which the tokens are produced.

This context sensitivity explains how biologists minimise terminological confusion even though they associate different kinds with a specific species term. First, biologists who share the same empirical interests work in a disciplinary context in which their tokens of the species term refer to the same lineage of populations. Among biologists who are interested in sexual reproduction, the context of their discussions determines a common reference for their tokens of the species term. These tokens refer to the lineage with interbreeding populations, rather than the lineage of populations responding to the same environmental pressures or the lineage of populations with a common ancestry. Second, outsiders who are trying to make sense of the biologists' discussions can turn to their disciplinary context to clarify the reference of their tokens. Third, if the biologists need to discuss the other related lineages, they can specify a non-standard reference for some tokens. Moreover, where misunderstandings are likely to arise, biologists can always adopt the more explicit terminology suggested by Ereshefsky (1992), which includes the terms 'biospecies', 'ecospecies', or 'phylospecies'.

Field and Kitcher's models also explain how biologists avoid abruptly changing the subject of their classification. In Field's model: stipulation leads to a denotational refinement of the species term, rather than an abrupt or arbitrary change in its

reference. Before stipulation, biologists use tokens of a species term to partially denote more than one lineage of populations. After stipulation, biologists use each token to fully denote just one of these related lineages. There is evidently no abrupt change in the subject of the biologists' classification. Collectively, before and after stipulation, biologists are studying the same three lineages and the evolutionary structures in which they participate. Stipulation enables them to refer, more precisely, to one of the lineages in order to investigate the subset of evolutionary structures that interests them.[51]

In Kitcher's model: the interests behind stipulation help to fix the reference of tokens of a species term. So different stipulations are associated with different referents for these tokens. But this does not mean that biologists from different disciplinary contexts have abruptly changed their subject. After all, all these biologists are still interested in lineages of one sort or another. Moreover, we can show through historical investigation that all the initiating events that fix the reference of different tokens are historically connected. They belong to the same scientific enterprise in which biologists from different disciplinary contexts investigate related evolutionary structures.

5.6 Conclusion

In this long chapter, I explained why biologists face the multiple kinds problem when

[51] Field notes that his model also addresses the threat of incommensurability: 'it shows that we can accept the claim that we can't always *equate* a term from one theory with a term from a later theory, and still deny the incommensurability thesis, i.e., the thesis that earlier and later terms cannot objectively be compared with respect to referential properties' (479).

they classify populations into species. And I demonstrated how they solve this problem by empirically constrained stipulation. I argued that, in the Kripke-Putnam model of natural kinds, this use of stipulation brings the risks of metaphysical caprice, epistemological complacency, and semantic confusion. To address the risks, an account of stipulation ought to clarify how a stipulated kind reflects the world's structures, how stipulation is constrained by the scientists' conditions, and how stipulation affects the kind term's reference.

Next I explained why biologists' longstanding controversy over what a species is leads them to the multiple kinds problem too. As a result of their different Species Concepts, they find more than one kind in their samples of a putative species. Then I evaluated LaPorte's arguments for stipulating one of these overlapping kinds to be the referent of the species concept. While his arguments clarify the diverse classificatory interests behind the biologists' disagreement, they do not make a convincing case for stipulation. I criticised LaPorte for not taking the differences in biologists' empirical interests seriously, and for being unduly pessimistic about balancing their non-empirical interests. I also noted that the semantic vagueness of a species term does not, by itself, licence biologists to stipulate its reference.

Finally, I offered a new, and more naturalistic, argument for stipulation. My argument from the epistemological perspective is based on two features of biological practice that are obscured in LaPorte's account. First, biologists tend to choose different Species Concepts according to their empirical interests, and thereby stipulate the reference of individual species concepts. Second, biologists use the Biological Species Concept, the Ecological Species Concept, and the Phylogenetic

Species Concept successfully in different research programmes. So I made an inference to the best explanation: the best explanation for these multiple successful stipulations is a role for stipulation in biological classification. As part of my argument, I rejected two alternative explanations that appeal to the biologists' error and luck.

I also addressed the risks of metaphysical caprice and semantic confusion. First, by appealing to what biology tells us about evolutionary structures, I explained that stipulation enables biologists to investigate the different causal processes and evolutionary structures that interest them. My interpretation maintains that the stipulated species are natural kinds. Although the classifications defined by the BSC, the ESC, and the PSC reflect different biological interests, the species within these classifications continue to reflect biological structures. Second, following Stanford and Kitcher, I proposed two semantic models in which stipulation produces a context-sensitive species term. This context sensitivity explains how different groups of biologists minimise terminological confusion although they associate different kinds with the same species term. Both models contain semantic resources to explain why stipulation does not abruptly change the subject of the biologists' classifications. Here I acknowledge that my arguments from the semantic perspective are more speculative than those from the metaphysical perspective; I can, at most, claim that stipulation need not lead to semantic confusion.

Chapter 6

Stipulation and subjectivity: a defence

6.1 Introduction

In this thesis, I have argued that the study of borderline consciousness is central to the science of phenomenal consciousness. Borderline consciousness poses a significant epistemological challenge to scientists who investigate phenomenal consciousness as a natural kind. When these scientists discover more than one overlapping kind in their samples of conscious creatures, how can they identify *the* kind to which all and only conscious creatures belong? According to my assessment, three recent philosophical responses to this multiple kinds problem are unduly pessimistic. By examining how biologists classify diverse populations into species, I have demonstrated that the same problem in biology can be solved by empirically constrained stipulation.

This supports my proposal on borderline consciousness. In Chapter 1, I proposed that, through stipulation, different groups of scientists can legitimately use the concept of phenomenal consciousness to refer to different kinds discovered in their samples of conscious creatures. Each group of scientists will refer to one of the overlapping kinds, in accord with their empirical interests. After stipulation, each group can classify at least some borderline conscious creatures more precisely, in accord with their conceptual decisions. But I noted a problem with this proposal. Using stipulation to re-classify borderline conscious creatures may strike some philosophers as *obviously* wrong, so wrong that it hardly warrants debate. Now I

need to address their objections.

Before I do so, let me retrace the trajectory of my arguments in Chapters 2 to 5. In Chapter 2, I analysed the concept of borderline consciousness in terms of phenomenal consciousness and borderline cases. My analysis explained how our rough-hewn behavioural and neural criteria for attributing phenomenal consciousness to others produce borderline cases. These criteria are, however, provisional: we expect science to improve them, and so classify at least some borderline conscious creatures more precisely. I defended this concept of borderline consciousness against two challenges from McGinn (1996) and Antony (2006b, 2008), who argue that borderline consciousness is inconceivable.

In Chapter 3, I distinguished two epistemological challenges that borderline consciousness poses to consciousness science: the problem of classification and the problem of multiple kinds. I concluded that the more promising strategy for re-classifying borderline conscious creatures is to build neural theories of phenomenal consciousness. Then I clarified some epistemological norms that govern this strategy. Following Block (2007a, b) and Shea and Bayne (2010), I also interpreted this strategy, in metaphysical terms, to be the investigation of phenomenal consciousness as a natural kind. Next, I explained why borderline consciousness leads scientists to discover more than one overlapping kind in their samples of conscious creatures. I clarified this multiple kinds problem by contrasting it with the phenomenon of multiple realisability in mental kinds. Then I demonstrated how the problem produces three theoretical impasses on the neural structure of phenomenal consciousness (Irvine 2013), the development of foetal consciousness (Derbyshire &

Raja 2011; Chin 2011), and the possibility of artificial consciousness (Prinz 2003, 2005). I emphasised the intractability of these impasses by showing that the epistemological norms cited earlier do not help with them.

In Chapter 4, I explained and evaluated three recent responses to the multiple kinds problem in consciousness science. These responses from Irvine (2013), Prinz (2003, 2005), and Papineau (2002, 2003) are pessimistic, in different degrees, about the science of phenomenal consciousness. Irvine recommends, on methodological grounds, that scientists eliminate the concept of consciousness from their research. Prinz argues that scientists are irremediably ignorant of the nature of phenomenal consciousness, although they can discover a lot about human consciousness. Papineau argues that consciousness science has limited prospects because the concept of consciousness-as-such suffers from a special form of referential indeterminacy. According to my evaluation, their arguments do not warrant this pessimism. Moreover, they raise the possibility that scientists can solve the multiple kinds problem by empirically constrained stipulation.

In Chapter 5, I demonstrated how biologists solve the multiple kinds problem by stipulation. Due to their longstanding controversy over what counts as a species, biologists face the multiple kinds problem when they classify populations into species. LaPorte (2004, 2010) argues that biologists can legitimately choose between different Species Concepts, and thereby stipulate the reference of individual species concepts. I explained how his arguments are based on the biologists' diverse classificatory interests and the semantic vagueness of their species terms. In my assessment, the arguments do not make a compelling case for stipulation, although

they clarify the biologists' disagreement. I proposed a new, more naturalistic, argument for stipulation based on the fact that biologists tend to choose different Species Concepts according to their empirical interests, and the fact that they use at least three of these concepts successfully in different research programmes. In my analysis, this use of stipulation is neither epistemologically complacent nor metaphysically capricious. According to two semantic models, it also need not lead to semantic confusion.

In this final chapter, I shall defend my proposal on borderline consciousness against the four key objections which I raised in Chapter 1. These objections claim that using stipulation to re-classify borderline conscious creatures is naturalistically unsound, metaphysically presumptuous about vagueness, semantically shifty, and subjectively absurd. First, I argue that empirically constrained stipulation is naturalistically sound on both methodological and metaphysical grounds. Second, I argue that this use of stipulation in consciousness science is consistent with all four theories of vagueness in the philosophical literature. Third, I examine its semantic consequences. I deny that this use of stipulation abruptly changes the subject of consciousness science. Fourth, I argue that its use is consistent with our introspective and intuitive views on the determinacy of conscious experience. I conclude with a set of conceptual considerations that point to the multiplicity of phenomenal consciousness.

I will not be offering a polemical defence of stipulation in consciousness science. Rather my aim is to show that none of the four objections against stipulating consciousness is decisive. Where possible, I start by clarifying or elaborating these objections. My responses to them draw on philosophical resources about natural

kinds, borderline cases, and phenomenal consciousness. Many of these resources were developed in Chapters 2 to 5; here I show how they can be connected. Through the dialectic of objections and responses, I hope to highlight the strengths and weaknesses in my account of stipulation's role in building neural theories of phenomenal consciousness. I also identify areas for further philosophical research.

6.2 Naturalistically unsound?

The first objection claims that using stipulation to re-classify borderline conscious creatures is naturalistically unsound. This objection is based on both the methodological and ontological stances in philosophical naturalism.[1] *Methodological naturalists* see science and philosophy as 'engaged in essentially the same enterprise, pursuing similar ends and using similar methods' (Papineau 2007). Often they assert 'some kind of general authority for the scientific method' in both science and philosophy. This methodological stance implies that, in debates on the neural nature of phenomenal consciousness, we should respect the established methods of science, especially empirical testing and inference to the best explanation. Stipulation will only hinder scientific discovery. De Brigard and Prinz (2010) warn of this danger: those who stipulate 'precisified definitions' of consciousness and other

[1] I learnt from the surveys of philosophical naturalism in Papineau (2007) and Jacobs (2009). As Papineau (2007) points out, there remains disagreement on which philosophical commitments are essential to these stances. What matters in each debate is to specify the relevant philosophical commitments, and then to assess their cogency. Here I follow Papineau's lead: 'The important thing is to articulate and assess the reasoning that has led philosophers in a generally naturalist direction, not to stipulate how far you need to travel along this path'. This advice is also rehearsed in the introduction of Papineau (1993).

For critical analysis of philosophical naturalism, see Kitcher (1992) and Friedman (1997). The naturalistic trend has flourished in contemporary philosophy of science; see the review of this trend, and its problems, in French and Massimi (2013).

related concepts 'often replace the folk concept with idiosyncratic definitions that settle crucial questions by fiat rather than facilitating the process of scientific investigation and discovery' (52).

Ontological naturalists, on the other hand, urge philosophers to avoid metaphysical speculation about the contents of reality. Reality has 'no place for "supernatural" or other "spooky" kinds of entities' (Papineau 2007). Instead, it consists fundamentally of nature's objects, properties, and kinds. Here nature is taken to be 'the order of things accessible to us through observation and the methods of the empirical sciences' (Jacobs 2009). If phenomenal consciousness is a natural kind, then its nature should be discovered by science. Stipulation will only produce an artificial kind with limited use in making inductive generalisations and causal explanations. This ontological stance is presented in Block (2002b). According to him, we who believe that phenomenal consciousness is real and has a 'scientific nature' should not define it through a 'decision' on 'extrapolating a concept of consciousness grounded in our physical constitution to other physical constitutions' (421).

I deny that this objection from naturalistic unsoundness applies to my proposal. Let me start from the methodological stance. My proposal to stipulate what counts as phenomenal consciousness should not trouble methodological naturalists. It does not try to pre-empt scientific methods. In Chapter 3, I observed that the term 'jade' is now used to refer to more than one chemical kind. This practice is motivated by commercial and cultural considerations (Hacking 2007b). In Chapter 4, I noted that we might re-classify some borderline conscious creatures as conscious, in order to express our moral concern (Papineau 1993; Putnam 1964). But I did not lean on

these commercial, cultural, and moral considerations to support my proposal.

Rather it is supported by my examination of the methods in biology. In Chapter 5, I argued that biologists can solve the multiple kinds problem by empirically constrained stipulation. As I emphasised, one strength of my argument is its naturalism. By attending to two neglected features of biological practice, I argued that biologists can legitimately choose between different Species Concepts according to their empirical interests, and thereby stipulate the reference of individual species concepts. Moreover, I argued that this use of stipulation does not hinder discovery in biology. Instead it enables different groups of biologists to investigate the evolutionary structures that interest them.

Here is another strength of my argument: it clarifies the conditions for stipulation in biology and the basis of stipulation under those conditions. I showed that biologists have room to stipulate the reference of a natural-kind concept when they discover more than one overlapping kind in their samples of the putative kind. But each kind has to reflect a distinct set of the world's structures and ground a distinct set of explanations about the world. Under such conditions, biologists can legitimately stipulate on the basis of their empirical interests. In Chapter 3, I identified the *norms of simplicity* which are usually cited when inference to the best explanation is used to justify a theoretical identity. Then I emphasised that these norms do not suffice to solve the multiple kinds problem. They can now be supplemented by the above norms about the conditions for stipulation and the basis of stipulation. I take these new norms to be among the *norms of stipulation* which govern how inference to the best explanation is used to break a theoretical impasse produced by the multiple

kinds problem.

I propose that scientists who face the multiple kinds problem in consciousness science should stipulate under the same conditions as biologists. This produces a unified approach to the current impasses on phenomenal consciousness, borderline consciousness, and artificial consciousness. Consider the impasse on the neural structure of phenomenal consciousness. The conditions for stipulation appear to be met. As Irvine (2013) shows, scientists have found more than one overlapping kind in their sample of conscious humans: 'the differing operationalisations split "consciousness" into one of many varied and fine-grained scientific kinds' (160). According to the norms of stipulation, different groups of scientists can legitimately use the concept of phenomenal consciousness to refer to different kinds, so long as each kind reflects a distinct set of neural structures and grounds a distinct set of neural explanations.

This means that scientists who are interested in transient neural synchrony and local recurrent processing can stipulate that the concept refers to the kind defined by these neural structures. Similarly, global workspace theorists who are interested in more sustained neural synchrony and global recurrent processing can stipulate that the concept refers to the kind defined by those neural structures. Both kinds are already cited in neuroscientific explanations. Through stipulation, different groups of scientists can focus on different kinds, in accord with their empirical interests. After stipulation, each group of scientists can classify some borderline conscious creatures more precisely, in accord with their conceptual decisions. For instance, scientists who focus on transient neural synchrony and local recurrent processing can re-

classify some late-stage foetuses as conscious. Global workspace theorists can re-classify creatures without more sustained neural synchrony and global recurrent processing as non-conscious.

In contrast, the conditions for stipulation are not yet fulfilled in the impasse on the possibility of artificial consciousness. The thought experiment in Prinz (2003) suggests that scientists will find at least two overlapping kinds in their samples of conscious humans. The first kind is defined by a functional structure constitutively related to the behaviour with which we identify conscious humans; this structure is common to conscious humans and our artificial duplicates. The second kind is defined by a biological structure in humans which realises this functional structure. From a naturalist perspective, Prinz's thought experiment cannot prove that both hypothetical kinds are natural kinds. We do not know, on the basis of his thought experiment, if the functional kind will be cited in any scientific explanations.

I want to highlight two weaknesses in how I defend stipulation's role in consciousness science. Both point to areas that need further investigation, though they are beyond the scope of this thesis. First, I have only looked at some biological practices for evidence that the multiple kinds problem can be solved by empirically constrained stipulation. LaPorte (2004, 2010) claims that stipulation has a wider role in biological and chemical classifications.[2] I am keen to discover if the argument from multiple successful stipulations applies to his other cases. Boyd (1993, 1999, 2013) also cites some plausible cases of stipulation in biological and chemical

[2] See LaPorte (2004) on the stipulation of higher biological taxa (76-85) and chemical substances (103-110). His arguments on chemical substances are contested in Slater (2005), LaPorte (2010), Bird (2007, 2010), Hendry (2010, 2012), and Massimi (2012).

classifications. These claims need to be evaluated – against the risks of metaphysical caprice, epistemological complacency, and semantic confusion that I have distinguished. We might uncover other conditions for stipulation, including social conditions, that are not met by consciousness science.

Second, I have not shown that empirically constrained stipulation is the best response to the multiple kinds problem in consciousness science. Rather I have shown that, if the right conditions are fulfilled, then stipulation is a legitimate response. How does stipulation compare to other responses in the philosophical literature? It is better motivated than the pessimism of Irvine's methodological eliminativism, Prinz's moderate mysterianism, and Papineau's referential indeterminacy, which I criticised in Chapter 5. In the debate on the neural structure of phenomenal consciousness in humans, Block (2007 a, b; 2008) and Prinz (2003, 2005) are more hopeful than I am that inference to the best explanation can break the impasse without using stipulation.[3] These claims need to be evaluated more thoroughly – given my arguments on the intractability of the multiple kinds problem. We might find other responses that are as good as, if not better than, stipulation.

Finally, I turn to the metaphysical stance in naturalism. My arguments on the metaphysical status of stipulated species should reassure ontological naturalists. In Chapter 5, I showed that different groups of biologists can legitimately stipulate that a species concept refers to different natural kinds. None of the stipulated species are constituted by the biologists' interests and actions. It is true that biologists with

[3] See also Shea (2012), who responds to Block (2007a). He explicitly raises the possibility of multiple kinds between which we cannot decide: 'We leave for another day the question of what to say about phenomenality if it turns out that there are many different natural kinds (or none)' (333).

different empirical interests can legitimately use different species classifications. So the classifications defined by competing Species Concepts do reflect biological interests. However, the species identified within these classifications do not thereby fail to reflect biological structures. Each stipulated species has an essence that is determined by a causal process responsible for some evolutionary structures. By stipulating a species concept differently, biologists are dividing their attention between the evolutionary structures that interest them.

Similarly, in consciousness science, the stipulated kinds in my proposal will be *natural* kinds. If the conditions for stipulation are fulfilled, then different groups of scientists can legitimately use the concept of phenomenal consciousness to refer to different natural kinds, in accord with their empirical interests. So their classifications will reflect these empirical interests. But the kinds will not thereby fail to reflect neural structures. Each stipulated kind will have an essence that is determined by a neural structure associated with phenomenal consciousness. By stipulating the concept of phenomenal consciousness differently, scientists will be dividing their attention between the neural structures that interest them.

Are these stipulated kinds *real* kinds? In Chapter 5, I drew on a distinction that Bird and Tobin (2015) make between weak and strong realism about natural kinds. Weak realists require only that these kinds demarcate natural divisions in the world. Such divisions must reflect 'the structure of the natural world rather than the interests and actions of human beings.' On the other hand, strong realists require that the kinds exist as 'a special sort of entity in our ontology' – distinct from the individuals in these kinds, and independent of human minds.

In Chapter 5, I argued that stipulated species are real enough for the weak realist and need not be unreal for the strong realist. I think the same arguments apply to the kinds stipulated in consciousness science according to my proposal. By the weak realist's standards, these stipulated kinds count as real. Their essences are determined by the neural structures associated with phenomenal consciousness, rather than the scientists' interests or actions. By the strong realist's standards, the stipulated kinds need not be unreal. The scientists' room for stipulation does not entail that phenomenal consciousness is constituted by the scientists' interests or actions. Therefore, it does not entail that the stipulated kinds are mind-dependent entities. It does not exclude them from being mind-independent entities such as sets or universals.

I want to mention another weakness in my defence. I cannot cite much neuroscience to make sense of stipulated consciousness from the metaphysical perspective. In Chapter 5, I cited what our best available biology says about evolutionary structures to make sense of stipulated species from the metaphysical perspective. I noted that sexual reproduction, environmental selection, and descent from common ancestry interact in complex ways to produce the evolutionary structures which biologists investigate and the diverse populations which they observe. Unfortunately, scientists know much less about neural structures than evolutionary ones. I foresee that they will find out much more about the contributions made by different types of neural synchrony and recurrent processing. Only then can we make more sense of the multiple kinds discovered by consciousness science.

6.3 Metaphysically presumptuous?

The second objection claims that stipulation is metaphysically presumptuous about vagueness. Stipulation is licensed by only two controversial theories in the philosophy of vagueness. In the first theory, our semantic indecision about a term explains its borderline cases (Lewis 1993; Barnes 2009). Once we have decided on more exact boundaries for the term, we can classify its borderline cases more precisely. The second theory holds that a special kind of contextual sensitivity explains borderline cases (Soames 1999; Shapiro 2006; Åkerman 2012). For instance, the exact boundaries of a term may depend on the judgements made by speakers in each context. If either theory of vagueness is correct about borderline consciousness, then scientists can stipulate more exact boundaries for the term 'phenomenal consciousness', in order to classify borderline conscious creatures more precisely.

However, there are at least two other theories of vagueness, which appeal respectively to irremediable ignorance and metaphysical indeterminacy. The first theory claims that, although a vague term has borderline cases, there is a fact of the matter about where its exact boundaries lie. Unfortunately, we are irremediably ignorant of these exact boundaries (Williamson 1994; Sorensen 2006). In the second theory, there is no fact of the matter about where these exact boundaries lie. Borderline cases arise because of metaphysical indeterminacy – indeterminacy in the world itself, rather than in our representation of the world (Williams 2008; Barnes 2009). Neither theory of vagueness supports the stipulating of consciousness.

This objection from metaphysical presumption highlights the lack of consensus among philosophers who theorise about vagueness. Fortunately, my proposal does not draw support from any controversial theory of vagueness. It is instead justified by my investigation into how biologists classify diverse populations into species. I argued that, given the same conditions for stipulation, scientists can solve the multiple kinds problem in consciousness science in the same way that they did in biology. I avoided claiming that stipulation can resolve all forms of vagueness in scientific classification. Indeed, in Chapter 5, I warned that vagueness by itself does not warrant stipulation in science. There I argued that the vagueness of the terms 'species' and '*Panthera tigris*' does not license biologists to stipulate referents for these terms – unless we are already committed to a philosophical theory that attributes their vagueness to semantic indecision or contextual sensitivity. The same warning applies to the vagueness of 'phenomenal consciousness'.

I think a more significant challenge lies behind the objection from metaphysical presumption. Even if my proposal does not draw support from any theory of vagueness, does it commit us to the presence of semantic indecision or contextual sensitivity in the term 'phenomenal consciousness'? If so, there seems to be a risk that my naturalistic arguments for stipulation might be trumped by future philosophical research into vagueness. For instance, philosophers might come to a consensus that either irremediable ignorance or metaphysical indeterminacy is our best explanation of vagueness.

Let me offer three responses to this challenge. The first response is *irenic*: it avoids any disagreement with philosophers who build theories of vagueness. I already

gestured at this response at the end of my conceptual analysis in Chapter 2. We start by distinguishing borderline cases which are re-classifiable by science from those which are not. Then we construe the former as apparent examples of vagueness, not genuine ones.[4] This means that our explanation of borderline consciousness is no longer bound by philosophers' best explanation of borderline baldness. Without worry, we can now explain borderline consciousness using semantic indecision or contextual sensitivity. Indeed, we can build a new theory of apparent vagueness in scientific classifications – independent of current, or future, theories of vagueness.

My second response is more *resolute*. It asserts our right to count borderline conscious creatures as genuine examples of vagueness, and to account for them using semantic indecision or contextual sensitivity.[5] We have this right even if philosophers decide that the other, more familiar, borderline cases are best explained by irremediable ignorance or metaphysical indeterminacy. Again we distinguish borderline cases which are re-classifiable by science from those which are not. But we construe the former as examples of kind vagueness, not Sorites-susceptible vagueness. We can insist that philosophical research into vagueness accommodate more than one explanation of vagueness, in order to account for different forms of vagueness. Already, in Chapter 5, we saw that Alston (1964) and

[4] I owe this suggestion to Williamson (1994), §7.7. Here is how he interprets the hypothetical discovery of an essence for heaps: '…we should presumably say that "heap" was not as vague as had been supposed, because the cases that appeared on the borderline were not really so. That would show that "heap" was not a good example; it would not show that anything was wrong with our account of vagueness' (213).

[5] Here we might draw moral support from Lewis (1994). He allows for semantic indecision when we analyse our 'traditional mental vocabulary' and search for its referents: 'any interesting analysandum is likely to turn out vague and ambiguous. Often the best that any one analysis can do is to fall safely within the range of indecision' (416). He also leaves room for semantic satisficing: 'analysis may reveal what it would take to deserve a name perfectly, but imperfect deservers of the name may yet deserve it well enough.'

LaPorte (2004) distinguish combinatory vagueness from degree vagueness.

I do not believe that philosophers who focus on Sorites-susceptible vagueness have sufficient standing to dismiss this resolute response. As I noted in Chapter 2, they tend to ignore what happens to borderline cases in science. Instead they rely on a limited range of cases to support their theories. Epistemicists like Williamson (1994) and Sorensen (2001) discuss being bald, being thin, being old, being noonish, being small, and being a heap. Theorists of metaphysical indeterminacy, such as Williams (2008) and Barnes (2009), concentrate on the composition of ordinary objects, the persistence of selves, and the identity of quantum particles.

Moreover, some theorists of metaphysical indeterminacy already allow for more than one explanation of vagueness. Williams (2008), for example, urges that philosophers be less 'imperialistic' about the nature of vagueness:

> Many familiar semantic and epistemic theorists present their theory as *imperialistic*, i.e. putatively covering every instance of indefiniteness. But theorists might be more pluralistic than this, taking indefiniteness to be realized by different kinds of phenomena in different cases. (766)

According to Williams, friends of metaphysical indeterminacy should be pluralists in this sense. They need not insist that all borderline cases are the result of 'indeterminacy in how the world is'; some may be explained by 'a lack of semantic conventions specifying the conditions under which an object gets picked out' (779). Prinz (1998) and Barnes (2009) share his view. Although Prinz claims the 'best metaphysical theories of objects, kinds, and properties may leave us with some indeterminacy', he concludes that the 'vagueness of language cannot all be

attributed to vagueness in the world' (§34-5). Barnes defines metaphysical indeterminacy modestly as 'the view that at least some cases of indeterminacy have their source in how the world is non-representationally' (372).

My third response is *compatibilist*: it argues that the proposal to stipulate what counts as consciousness is compatible with both irremediable ignorance and metaphysical indeterminacy. We should distinguish the roots of vagueness from our response to it. The four theories of vagueness find its roots in either semantic indecision, irremediable ignorance, metaphysical indeterminacy, or contextual sensitivity. But, to my knowledge, no one has shown that these theories must disagree on our response to vagueness. More specifically, neither irremediable ignorance nor metaphysical indeterminacy excludes stipulation as a reasonable response to borderline consciousness.

We can make sense of stipulated consciousness in the light of irremediable ignorance or metaphysical indeterminacy. Suppose that we begin by being irremediably ignorant of the referent for the term 'phenomenal consciousness'. Before the stipulation that I propose, the term refers to an unknowable kind.[6] Afterwards, different groups of scientists will use the same term to refer to different natural kinds, in accord with their empirical interests. For most, if not all, of these groups, the referent of the term will shift from the unknowable kind to a known kind. Only by luck will a group of scientists stipulate the kind to which the term originally referred. Next, suppose that the term 'phenomenal consciousness' refers

[6] See Chapter 5, note 11. For simplicity's sake, I assume that the natural-kind term 'phenomenal consciousness' refers to a kind. Those who reject this assumption can recast my arguments in terms of the extension of the term.

to a metaphysically indeterminate kind.[7] Again, after stipulation, different groups of scientists will refer to different natural kinds. For all of these groups, the referent of the term will shift from a more metaphysically indeterminate kind to a less metaphysically indeterminate one.

Does a reference shift abruptly change the subject away from conscious experience? I will address this concern in the next section. For reasons that I explain below, I think that semantic indecision or contextual sensitivity makes better sense of what happens to a natural-kind term upon stipulation. So I do not really need the compatibilist response. But it may prove useful to those who want to use irremediable ignorance or metaphysical indeterminacy to explain all forms of vagueness. What about the choice between the irenic and resolute responses? This seems to me a tactical matter. If some philosophers insist that vagueness excludes borderline cases that are re-classifiable by science, or that it excludes borderline cases that do not generate Sorites paradoxes, then I am inclined to let them have the term 'vagueness'. We can wait for their imperialistic mood to pass. Meanwhile, as Weiner (2007) reminds us, the full range of borderline cases in scientific classifications needs to be investigated. I have only examined some cases that can be resolved by stipulation and clarified the conditions for stipulation in them. I have yet to distinguish them systematically from cases that cannot be resolved by it.[8]

[7] For philosophical accounts of metaphysically indeterminate kinds, see Prinz (1998), §3, Boyd (1999), 143-4, and Barnes (2009), 377.
[8] I owe this criticism to Luca Barlassina, who raised it during a conference discussion in Bochum.

6.4 Semantically shifty?

The third objection finds stipulation to be semantically shifty. Let me cite two philosophers who mention it in relation to vagueness. Williamson (1994) raises this objection generally in a non-scientific context. Suppose that we stipulate a precise boundary for the vague term 'thin'. According to him, both the meaning and extension of the term are likely to change.

> On just about any view, such a stipulation changes the meaning of 'thin' in those contexts in which it has authority. One needs to know more than was previously necessary to understand the word. There is also likely to be a slight change in its extension. (214)

For Williamson, these changes in meaning and extension amount to a change of subject. Stipulation does not 'tell us anything about the truth or falsity of utterances already made using "thin" in its old sense' (214). He adds that this change in meaning 'can be beneficial' because 'the exact location of a cut-off point is sometimes less important than the fact that we know its location'. Moreover, stipulating a more precise boundary for the vague term may allow us to ask 'new and sometimes better' questions. But he warns that it does not 'answer old questions'. In Williamson's judgement, stipulation 'may be quickly dismissed' as a strategy for handling borderline cases.[9]

Prinz (1998) highlights the same concerns in a scientific context much like ours.

[9] See also Sorensen (2012), §1. He asks if removing one head from a two-headed man counts as decapitation: 'We could give the appearance of settling the matter by stipulating that "decapitate" means "remove a head"…But that would amount to changing the topic to an issue that merely sounds the same as decapitation.'

Suppose that there are numerous biological kinds that subsume all the existing creatures which we call 'dogs' (§39). These 'highly similar' kinds differ only slightly 'in their phylogeny or in their permissible degrees of genetic variance'. More than one of them may be involved in our initial baptism of the term 'dog'. If scientists discover these kinds, and stipulate that the term 'dog' refers to just one of them, this will be 'mere fiat' (§40). Stipulation will 'alter the meaning' of the original term, whose reference is fixed 'in a context indifferent to the distinction' between these kinds. According to Prinz, some natural-kind concepts 'may suffer from an irresolvable indeterminacy'.

How might these semantic worries apply to stipulated consciousness? The objection goes: When scientists stipulate the kind to which all conscious subjects belong, they will likely shift the reference of the term 'phenomenal consciousness'. By doing so, they risk abruptly changing the subject of their classification. It might no longer be about conscious experience – the mental phenomenon with borderline cases in which we are interested. Block (2002) spells out this danger in consciousness science: 'Stipulations need not stick when it comes to the phenomenal realist conception of consciousness' (422). After scientists use stipulation to re-classify borderline conscious creatures, our original question about borderline consciousness would remain unanswered. We could raise it anew by asking if the stipulated consciousness of any re-classified creatures '*feels the same*' as our consciousness.

I think that this objection can be defused by considering the details of my proposal. My proposal does not allow scientists to stipulate subjects in an arbitrary fashion. So it is not comparable to arbitrary stipulations that obviously do change the subject

under consideration. For instance, Williamson (1994) likens the stipulation of a precise boundary for the term 'thin' to the stipulation of birth-dates for the emperors of Rome (214), while Block points to the stipulation that 'Tuesdayconsciousness' refers to consciousness that occurs on Tuesday (422). None of these would be apt comparisons for the empirically constrained stipulations in my proposal.

My proposal is modelled after the biologists' solution to the multiple kinds problem. When different groups of biologists heed their empirical interests and stipulate that a species concept refers to different species, they do not abruptly change the subject of their classification. As I explained in Chapter 5, these biologists are simply dividing their attention between the evolutionary structures that interest them. Together these evolutionary structures account for the diverse populations that the biologists want to classify. Similarly, if scientists stipulate consciousness under the same conditions, they will be dividing their attention between the neural structures that interest them. These structures account for the behavioural patterns that scientists associate with phenomenal consciousness. Like the biologists who stipulate species, these scientists who stipulate consciousness will not be abruptly changing subject.

In Chapter 5, I used two semantic models developed by Field (1973) and Kitcher (1978) to analyse what happens to a species term upon stipulation. My analysis suggests that, in both models, stipulation results in a context-sensitive term. The same analysis applies to the term 'phenomenal consciousness' in consciousness science. What will happen to this term in the debate on the neural structure of phenomenal consciousness? In Field's model: before stipulation, the term

‘phenomenal consciousness’ has indeterminate reference. Each token of the term partially denotes a kind defined by transient neural synchrony and local recurrent processing, and partially denotes a kind defined by more sustained neural synchrony and global recurrent processing. If scientists with shared empirical interests stipulate the reference of the concept of phenomenal consciousness, the term ‘phenomenal consciousness’ will undergo denotational refinement. After stipulation, each token of the term will fully denote just one of the two kinds.

In Kitcher’s model: stipulation will not lead to denotational refinement of the term ‘phenomenal consciousness’. Rather the interests that lie behind stipulation will help to fix anew the reference of tokens of the term in a disciplinary context. So scientists with different empirical interests will use the same term to refer to different kinds. For instance, among global workspace theorists, the reference of their tokens of ‘phenomenal consciousness’ will be fixed to the kind defined by more sustained neural synchrony and global recurrent processing.

Both models explain how scientists avoid abruptly changing the subject even though their stipulations will shift the reference of the term ‘phenomenal consciousness’. In Field’s model: stipulation will lead to a denotational refinement of the term, rather than an abrupt or arbitrary shift in its reference. Before stipulation, scientists use tokens of the term to partially denote the two kinds in their sample of conscious humans. After stipulation, they will use each token to fully denote just one of these two kinds. There will be no abrupt change in subject. Collectively, before and after stipulation, scientists will be attending to the same kinds and the same neural structures in which the kinds participate. Stipulation will enable them to refer, more

precisely, to one of these kinds in order to investigate the subset of neural structures that interest them.

In Kitcher's model: the interests behind stipulation will fix anew the reference of tokens of the term 'phenomenal consciousness'. But this does not entail that scientists will abruptly change their subject. After all, each token of the term will refer to one of the two kinds that scientists find in their samples of conscious humans. Both kinds are defined by neural structures which account for the behavioural patterns that scientists associate with phenomenal consciousness. Moreover, the initiating events that fix the reference of different tokens will be historically connected. They will belong to the same scientific enterprise in which scientists from different disciplinary contexts investigate neural structures related to phenomenal consciousness.

Here are two weaknesses in my defence against semantic shiftiness. First, I have relied on two relatively abstract models of reference to analyse the semantic consequences of stipulation. I think Field and Kitcher's models suffice to show that scientists need not abruptly or arbitrarily change the subject away from conscious experience even though their stipulations will shift the reference of the term 'phenomenal consciousness'. To make a more complete analysis, I will need to examine the detailed history of the term 'phenomenal consciousness' – including when and how the term is introduced into consciousness science by some philosophers and scientists, and then spread to others. Without this history, it is not yet possible to distinguish and evaluate the roles of partial reference and contextual reference in the semantic structures associated with the term 'phenomenal

consciousness'.

Second, in my analysis, I have focused on the impact of stipulation on the reference of the term 'phenomenal consciousness'. I have yet to consider its impact on the meaning of the term. This is mainly because I have not found models of meaning in philosophy of science to complement Field and Kitcher's models of partial and contextual reference.[10] However, two theories of vagueness may be of some help here. Both theories clarify what happens to the meaning of the term 'phenomenal consciousness' upon stipulation. If the term's vagueness arises from semantic indecision, then the stipulations I propose will only make its meaning more precise. If the term's vagueness arises from contextual sensitivity, then the stipulations appropriate to various contexts will not modify its meaning.

What if the term's vagueness arises from irremediable ignorance or metaphysical indeterminacy? For now, let me offer a general reason to deny that the stipulations in my proposal will change the meaning of the term 'phenomenal consciousness' in a problematic way. In Chapter 5, I showed that stipulation under some conditions solves the multiple kinds problem in biology. If I am right, then empirically constrained stipulations do not constitute a change of meaning; the possibility of such stipulations was 'part of the original enterprise'.[11] It does not matter that the laymen or scientists who introduced the species terms were ignorant of the stipulations to come. It does not even matter that they were ignorant of stipulation's

[10] Kitcher (1978) blames the 'philosophical difficulties of the notion of meaning' (521). He notes that, after Kuhn and Feyerabend, many philosophers have 'urged the benefits of doing as much semantics as possible within the theory of reference.'

[11] Compare Kripke (1980), 138: 'scientific discoveries of species essence do not constitute 'a change of meaning'; the possibility of such discoveries was part of the original enterprise.' He makes similar points, on 119-120, about the meanings of the terms 'gold' and 'tiger'.

role in scientific classifications. For they too would acknowledge that the scientific enterprise always leaves room for the discovery of new methods and norms. Similarly, empirically constrained stipulations will not change the meaning of the term 'phenomenal consciousness. If we accept the legitimacy of such stipulations in consciousness science, then we should acknowledge these stipulations as part of the science. Whether my reasoning here is decisive depends on the future models that philosophers develop for the meaning of scientific terms.

6.5 Subjectively absurd?

The fourth, and final, objection criticises stipulation for being subjectively absurd. From our subjective perspective, phenomenal consciousness does not appear to be something that can be stipulated. So this objection denies that it can be treated like species: even if scientists can solve the multiple kinds problem in biology by stipulating species, they cannot do the same in consciousness science by stipulating consciousness. In §6.2, I noted that my arguments for stipulating species support my proposal for stipulating consciousness. If this objection from subjective absurdity is sound, then it gives us ground to reject the analogy between the conditions for stipulation in biology and those in consciousness science.[12]

Suppose we look at phenomenal consciousness through introspection – examining our own experience from the inside, rather than another's experience from the outside. When we do so, we are likely to be impressed by the metaphysical

[12] It is reminiscent of the rebuke in Nagel (1974): 'Every reductionist has his favorite analogy from modern science. It is most unlikely that any of these unrelated examples of successful reduction will shed light on the relation of mind to brain' (519).

determinacy of phenomenal consciousness. Either we have experience or we do not have it. We may have specific states of experience that are faint or fleeting; but there is a fact of the matter whether we have experience at all. This fact is not one which depends on another's decision. Therefore, it seems absurd to propose that scientists use stipulation to re-classify borderline conscious creatures.

Chalmers (1996) comes close to making this objection against stipulation when he criticises functional analyses of phenomenal consciousness. Instead of an introspective view of our own experience, he appeals to an intuitive view of the mouse's experience.

> Does a mouse have conscious experience? Does a virus? These are not matters for stipulation. Either there is something it is like to be a mouse or there is not, and it is not up to us to define the mouse's experience into or out of existence. To be sure, there is probably a continuum of conscious experience from the very faint to the very rich; but if something has conscious experience, however faint, we cannot stipulate it away. (105)

To rule out the possibility of stipulation, Chalmers cites the determinacy of the mouse's conscious experience: 'Either there is something it is like to be a mouse or there is not.' He is implicitly making an assumption about stipulation. According to this assumption, we can legitimately stipulate whether the mouse has experience only if it is somehow indeterminate whether the mouse has experience. More generally, we can legitimately stipulate more precise boundaries for a kind only if it is indeterminate. Since we intuit that it is determinate whether the mouse has experience, we can infer – by *modus tollens* – that there is no room to stipulate whether the mouse has experience.

I shall offer two responses to this objection. The first response is *irenic*. It accommodates our views on the determinacy of experience: either we have experience or we do not have it, either there is something it is like to be a mouse or there is not, there is a fact of the matter whether a creature is conscious or not, etc. I claim that we can make sense of stipulated consciousness without committing to any metaphysical indeterminacy of phenomenal consciousness. Indeed, in §6.3, I offered three interpretations of my proposal for stipulating consciousness which do not appeal to metaphysical indeterminacy. None of these interpretations compel us to deny that there is a fact of the matter whether the mouse has experience.

Consider the interpretation based on semantic indecision. If borderline consciousness arises from our semantic indecision over the term 'phenomenal consciousness', then the reference of this term is indeterminate. Before stipulation, it refers indeterminately to more than one kind – say, the kind of creatures with transient neural synchrony and local recurrent processing, and the kind with more sustained neural synchrony and global recurrent processing. In this interpretation, there is room for stipulation even if the kinds themselves are determinate. So we have no reason to doubt that there is a fact of the matter whether the mouse belongs to each of these kinds.

What about the interpretations that cite contextual sensitivity and irremediable ignorance? If borderline consciousness arises from the contextual sensitivity of the term 'phenomenal consciousness', then this term refers to different kinds in different contexts of stipulation. Before a context is specified, the term refers indeterminately to more than one kind. If borderline consciousness arises from our

irremediable ignorance, then the term ‘phenomenal consciousness’ has determinate reference. Before stipulation, it refers determinately but unknowably to one of the kinds. Again, stipulation does not require that the kinds themselves be indeterminate. Discovering room for stipulation adds no reason to deny that there is a fact of the matter whether the mouse belongs to each of these kinds.

Does my proposal permit scientists to ‘define the mouse’s experience into or out of existence’? That would be a tendentious characterisation of the stipulations in my proposal. I do not construe borderline consciousness as phenomenal consciousness that can be defined ‘into or out of existence’. I do concede that whether scientists classify the mouse as conscious depends on which kind they stipulate to be the referent of the concept of phenomenal consciousness. Under the right conditions, scientists with different empirical interests can legitimately stipulate different kinds. If the mouse is a borderline conscious creature, then it belongs to at least one but not all of these kinds. After their stipulations, some scientists may classify the mouse as a conscious creature; others may not. But, in the interpretations based on semantic indecision, contextual sensitivity, and irremediable ignorance, this possibility is not produced by there being no fact of the matter whether the mouse has experience.

Moreover, in two crucial senses, whether the mouse has experience is not determined by the scientists’ stipulations. First, whether the mouse belongs to each of the kinds stipulated is not decided by the scientists. Second, none of the kinds stipulated is constituted by the scientists’ interests or actions. Rather each kind has an essence that is determined by neural structures associated with phenomenal

consciousness. Therefore the boundaries of these stipulated kinds are also not decided by scientists.

I see a weakness in this irenic response. By accommodating our introspective and intuitive views on the determinacy of experience, we can no longer explain borderline consciousness in terms of metaphysical indeterminacy. Losing this option undermines part of my compatibilist response in §6.4, where I argued that my proposal for stipulating consciousness is consistent with the metaphysical indeterminacy of phenomenal consciousness. In general terms, this loss does not affect my arguments for stipulating consciousness, since we can still explain borderline consciousness through semantic indecision, irremediable ignorance, or contextual sensitivity. But, ideally, I should like to leave open the possibility that, before stipulation, some borderline consciousness arises from metaphysical indeterminacy. Moreover, I do not want to rule out the possibility that some of the kinds stipulated by scientists will themselves be indeterminate.

So my next response is more *resolute*: it challenges our introspective and intuitive views on the determinacy of experience. On methodological grounds, I deny that naturalists about phenomenal consciousness have to accommodate these introspective and intuitive views. In §6.2, I argued that my proposal for stipulating consciousness is naturalistically sound. From this naturalist perspective, there is no reason to trust that either our introspection or intuition provides access to the nature of phenomenal consciousness.

Take the introspective view that I described earlier: either we have experience or we

do not have it. This view seems authoritative because we are examining our own experience 'from the inside'. But why should phenomenal consciousness be transparent in this way to itself? Naturalists agree, of course, that we can learn about experience from introspection. But we have no reason to believe that introspection offers special insight into the nature of experience – including its determinacy or lack thereof. Instead, as I noted in §6.2, naturalists assume that phenomenal consciousness has a hidden nature which can be uncovered by scientific methods. If our best science says that phenomenal consciousness is metaphysically indeterminate, then an introspective view cannot trump this scientific judgement.

Here I heed a warning from Loar (1997). He urges naturalists to reject the assumption that the conscious mind is transparent to itself: 'on a naturalist view of human nature, one ought to find it puzzling that we have such a first-person insight into the nature of our mental properties' (614). This naturalist view is contrasted with 'a dualist conception of Platonic insight into mental essences'. In addition, Loar uses an account of phenomenal concepts to explain why we expect the nature of phenomenal properties to be transparent. His explanation has been somewhat neglected in the philosophical literature on consciousness. I sketch the explanation below because it complements my methodological challenge. It helps naturalists about phenomenal consciousness understand why we too readily assume transparency.

In Loar's account, phenomenal concepts are type-demonstratives whose reference is

fixed from the subjective perspective (1997: 597).[13] For instance, the concept of pain has the demonstrative structure '*that* type of sensation'. The more general concept of phenomenal property has the structure '*that* type of determinable property associated with determinates such as sensations, sensory experiences, etc.'[14] Each concept does not pick out a phenomenal property via its contingent properties. Rather each concept uses an instance of a phenomenal property, or an image of the property, to narrow down its reference (604). In this sense, a phenomenal concept picks out its referent 'directly' – without invoking the referent's contingent properties. However, as Loar emphasises, this 'direct grasp' is different from that of a theoretical description which picks out the same property via its essential properties (609).

We easily conflate these two types of direct grasps. That is, we forget that one can conceptualise a property *introspectively* by distinguishing it in our experience, and also conceptualise it *theoretically* by describing its essence. According to Loar, this creates an illusion.

> The illusion is of *expected transparency*: a direct grasp of a property ought to reveal how it is internally constituted, and if it is not revealed as physically constituted, then it is not so. The mistake is the thought that a direct grasp of essence ought to be a transparent grasp, and it is a natural enough expectation. (609, original italics)

[13] Loar's 'recognitional' model of phenomenal concepts closely resembles Papineau's 'quotational' model, which I discussed in §4.4. Both models are used to defuse anti-physicalist arguments about the nature of consciousness.

In the philosophical literature on consciousness, this 'phenomenal concept strategy' is widely seen as a promising defence of physicalism. See Van Gulick (1993); Sturgeon (1994); Block (2002), IV; Papineau (2002, 2006); McLaughlin (2003), 191-3; Alter and Walter (2007), Part II. For criticism of this strategy, see Chalmers (2007), Tye (2009), ch. 3, and Shea (2014).

[14] Although Loar's article is titled 'Phenomenal States', he generally puts his arguments in terms to phenomenal properties or qualities. In his analysis, the most general phenomenal concept is the concept of '*phenomenal* (state, quality)', which refers to the 'highest ranking phenomenal determinable' (610).

Under this illusion, we expect an introspective conceptualisation of a phenomenal property to reveal the property's essence – much like a theoretical description of its essence. As Loar suggests, our expectation of transparency helps to explain why some philosophers believe that phenomenal consciousness is not physically constituted. Through introspection, phenomenal consciousness does not appear to be physically constituted. If we expect transparency, then we naturally conclude that this consciousness is not physically constituted. I suggest the same illusion contributes to the unwarranted confidence that phenomenal consciousness is metaphysically determinate. Since our introspection does not reveal this consciousness to be metaphysically indeterminate, we conclude that it is metaphysically determinate.

What about our intuitive views on the determinacy of experience? These views do not rest on our introspection of experience. Rather they depend on our intuitions about another creature's experience. Earlier I cited Chalmers' view: 'Either there is something it is like to be a mouse or there is not' (1996: 105). He must find this view intuitively obvious, since he offers no evidence for it. He just assumes it in an argument against functional analyses of consciousness. In Chapter 2, I also discussed McGinn's claim that 'either a creature definitely is conscious or it is definitely not' (1996, 14). He appeals to our imagination to support this claim: 'we cannot imagine the position of a creature for whom it is indeterminate whether there is such an "inner" subjective aspect' (15).

Here is how Papineau (2002) spells out the intuitive view in relation to octopuses and silicon doppelgängers. He imagines an interlocutor who says:

> …surely, it seems, there must be a fact of the matter whether it is *like anything at all* for such creatures. Maybe there is unclarity about how exactly to classify specific states of consciousness in alien creatures. But it can't be unclear whether they have any such states to start with. (202-3, original italics)

I note that the interlocutor's view, as presented above, is confused. Even if there is a fact of the matter whether borderline conscious creatures are conscious, there can be unclarity among scientists about that fact. The former is a matter of metaphysics, the latter an issue of epistemology. His intuitive view is really about the metaphysical determinacy of consciousness: 'Either there is some spark of consciousness present, or there isn't.' Papineau (2003) captures the view with an apt metaphor: 'Surely a light is on, or it is not' (219).

From the naturalist perspective, I see no reason to trust these intuitive views when they are not supported by empirical evidence. Why would our intuition offer insight into the nature of phenomenal consciousness? Recall that intuition also assures many of us that the mind is not physically constituted – a view that is, in fact, philosophically contested. Therefore, I take our intuitive views about the determinacy of experience to be on par epistemically with the introspective view. If our best science implies that phenomenal consciousness is metaphysically indeterminate, intuition cannot overrule scientific judgement. I have in mind another naturalist's warning. When Block (1980) analyses the 'logic' of appeals to intuition, he warns that intuitive views on the mind, however strong, cannot outweigh a scientific theory with empirical support. 'At best, intuition reveals facts about our *concepts* (at worst, facts about a motley of factors such as our prejudices, ignorance, and still worse, our lack of imagination – as when people accepted the deliverance of

intuition that two straight lines cannot cross twice)' (103-4).

Let me highlight two weaknesses in this resolute response to the objection from subjective absurdity. They turn out to be related in a surprising way. First, my account of the introspective and intuitive views, as they are seen from a naturalist perspective, remains incomplete. I have explained, in naturalist terms, why our views on the determinacy of experience are epistemologically dubious. I have also explained, in naturalist terms, how an illusion of transparency reinforces our mistaken confidence in these views. But I have yet to explain how the views themselves arise. Second, by focusing on a methodological challenge, I have failed to assess the introspective and intuitive views on conceptual grounds. In particular, I have gone along with the assumption that all these views address the metaphysical determinacy of phenomenal consciousness. But, as I argue below, it is likely that this assumption rests on some conceptual confusion.

Consider again the introspective view: either we have experience or we do not have it. But this view does not imply that there is a fact of the matter whether all creatures have experience. Therefore, it does not imply that there is fact of the matter whether borderline conscious creatures such as octopuses, fishes, frogs, late-stage human foetuses, and vegetative-state patients have experience. We might experience our own state of phenomenal consciousness as determinate. But this does not mean that phenomenal consciousness is everywhere determinate. In conceptual terms, we ought not to confuse the determinacy *in* our experience with the determinacy *of* experience. The former may be produced by something other than the latter, such as the determinacy of our sense of self, or the determinacy of

our sense of ownership.

Similarly, we should be more cautious in interpreting intuitive views on the determinacy of experience. Take the view which Chalmers cites: 'Either there is something it is like to be a mouse or there is not' (2006, 105). Strictly speaking, this says that there is a fact of the matter whether the mouse has experience. It does not say that there is a fact of the matter whether all creatures – including all borderline conscious ones – have experience. Indeed, nothing in the rest of his passage entails the general claim that phenomenal consciousness is metaphysically determinate. In conceptual terms, we ought not to confuse the determinacy of experience in a *particular species* with the determinacy of experience *per se*.

If I am right about the conceptual confusion, then it may help to explain how our views on the metaphysical determinacy of phenomenal consciousness arise. I only have room here to venture two psychological hypotheses, which need to be tested. My first hypothesis is that, because we confuse the determinacy in our experience with the determinacy of experience, we come to believe that there is a fact of the matter whether every creature has experience. According to my second hypothesis, we are prone to projecting the determinacy of phenomenal consciousness in familiar species onto other species. When we are asked to assess intuitively the determinacy of phenomenal consciousness, we tend to imagine familiar species which are either conscious or not conscious. Then we extend our sense of determinacy onto other species, including species that are currently classified as borderline conscious.

6.6 The multiplicity of subjectivity

I have been examining what stipulating consciousness means for the determinacy of phenomenal consciousness. Now I want to consider what it means for the multiplicity of phenomenal consciousness. According to my proposal, different groups of scientists can legitimately use the concept of phenomenal consciousness to refer to more than one overlapping kind discovered in their samples of conscious creatures. In this sense, my proposal implies that there are more than one subjectivities – just as there are biospecies, ecospecies, and phylospecies. How can we make sense of this multiplicity of subjectivity?

Earlier, in §6.2, I conceded that we do not know enough about the neural structures associated with phenomenal consciousness to make sense of stipulated consciousness from the metaphysical perspective. Here I shall pursue an interim strategy: a more thorough conceptual analysis of phenomenal consciousness may illuminate some of the overlapping kinds that can be legitimately stipulated to be the referent of the concept of phenomenal consciousness. First, I show that, contrary to appearances, philosophers who debate the nature of phenomenal consciousness generally acknowledge that there may be more than one kind associated with the concept of phenomenal consciousness. Second, in conceptual terms, I distinguish three kinds of consciousness that philosophers and scientists often associate with being a subject of experience: the qualitative, perspectival, and first-personal.

This conceptual complexity is obscured when philosophers and scientists borrow

Nagel's phrase to define phenomenal consciousness. In their standard definition, a creature's phenomenal consciousness is 'what it is like' to be that creature. As I showed with examples in §2.2, philosophers also use this phrase to connect phenomenal consciousness conceptually with subjectivity and qualia. Nagel (1974) himself uses 'what it is like' to be a creature to define 'the subjective character of experience' (526). Chalmers defines qualia as 'those properties that characterize conscious states according to what it is like to have them (2002, 268). Prinz (2003) refers to 'qualitative, phenomenal feelings – Nagel's famous what-it's-likes' (112). According to Tye (2013), the phenomenal character of experience is 'what it is like subjectively to undergo the experience' (§1). These connections may leave us with the impression that the concept of phenomenal consciousness has a unified sense, which defines *the* subjective character of experience.

But this impression is misleading. In fact, many philosophers concede that, at least conceptually, there seems to be more than one kind associated with the concept of phenomenal consciousness. Let me highlight this concession from three influential philosophers, who nevertheless disagree on the nature of phenomenal consciousness. They are Chalmers (1996), who claims that phenomenal consciousness is non-physical in nature; Searle (1997), who believes that it is biological; and Block (1995a, b), who believes that it is likely to be partly neural.

For Chalmers (1996), the term 'consciousness' refers to the 'subjective quality of experience' (6). It is connected to a list of other terms in the philosophical literature:

> A number of alternative terms and phrases pick out approximately the same class of phenomena as "consciousness" in its central sense. These include

> "experience," "qualia," "phenomenology," "phenomenal," "subjective experience," and "what it is like." Apart from grammatical differences, the differences among these terms are mostly subtle matters of connotation. "To be conscious" in this sense is roughly synonymous with "to have qualia," "to have subjective experience," and so on. Any differences in the class of phenomena picked out are insignificant. (6)

Here Chalmers accepts that these terms, which are currently associated with phenomenal consciousness, have different connotations and denotations. They only denote 'approximately the same class of phenomena'. He does not spell out the differences between the relevant classes of phenomena. It is not surprising that he finds these differences to be 'insignificant', since his aim is to provoke a metaphysical debate on the nature of phenomenal consciousness. Yet such differences are bound to be more significant to scientists trying to discover the neural nature of phenomenal consciousness.

Unlike Chalmers, Searle (1997) believes that consciousness is '*a feature of* the brain' (8, original italics). In common sense terms, he defines consciousness as 'those states of sentience and awareness that typically begin when we awake from a dreamless sleep and continue until we go to sleep again, or fall into a coma or die or otherwise become "unconscious"' (5). It is, Searle emphasises, an 'inner, first-person, qualitative phenomenon.' But, elsewhere in the same book, he allows that consciousness may not be one unified phenomenon. For instance, when he evaluates Dennett's theory of consciousness, he claims that Dennett thinks 'there are no such things as qualia, subjective experiences, first-person phenomena, or any of the rest of it' (99). This at least hints that qualia, subjective experiences, and first-person phenomena may be different kinds of things.

What about Block? I focus on his remarks in Block (1995a, b), because this is where he introduces and defends the distinction between phenomenal consciousness and access-consciousness. Block denies that he advocates 'one true taxonomy' of consciousness (1995b: 238). Indeed he acknowledges that there are 'somewhat different notions of phenomenal consciousness that are legitimate for some purposes, for example, the limitation to bodily sensations'. Here he cites the criticisms made by Humphrey (1995) and Katz (1995) to his concept of phenomenal consciousness. Humphrey (1995) claims that Block's concept is 'itself something of a mongrel' (257). He prefers to use the concept to refer only to one kind: the consciousness of bodily sensations, 'widely interpreted' to include sensory stimulations in the eyes, ears, nose, and skin. On the other hand, Katz (1995) distinguishes two kinds of phenomenal consciousness (258). The first is the 'phenomena of qualitative consciousness' in sensory experience. The second is whatever impoverished kind of phenomenal consciousness that 'accompanies representational thought'; this reflects 'how it's often only very barely "like anything" at all to think'.

More recently, two other philosophers – De Sousa (2004) and Lycan (2004) – offer more thorough conceptual surveys that help to distinguish the multiple kinds associated with the concept of phenomenal consciousness. Both recommend the same 'divide and conquer' strategy: by distinguishing the kinds conceptually, we can investigate them individually. De Sousa (2004) focuses on the concept of subjectivity. He complains that the term 'covers many things, and the word sounds all the more impressive for the fact that the things it purportedly designates are lumped into a

very mixed bag' (147). Sorting out this bag, he finds twelve 'senses' or 'aspects' of subjectivity, though he concedes that some of them may eventually be redundant and reducible to the others (161). Out of his twelve varieties, three are clearly cited in the philosophical literature in relation to phenomenal consciousness. They are: having a *perspective* from which to apprehend the world (150); having a sense of *self* and experiencing oneself 'as having the power to choose and act' (151); and having *qualia* (159).

On the other hand, Lycan (2004) focuses on the concept of consciousness. He identifies six problems of 'phenomenal experience' that have 'most greatly exercised philosophers' (37). These are: *qualia (strictly so called)*; the *homogeneity or grainlessness* of qualia; the *intrinsic perspectivalness, point-of-viewness, and/or first-personishness of experience*; *funny facts*, or special phenomenal knowledge; the *ineffability of 'what it's like'*; the *'explanatory gap'.* According to his analysis, these problems are about two distinct kinds of experience. The first kind consists of qualia. They are 'introspectible qualitative phenomenal features that characteristically inhere in sensory experiences', such as seeing a patch of colour, hearing a sound or smelling an odour (6). The second is the 'intrinsic perspectivalness of the mental' (9). Experience of this kind comes with points of view. Lycan goes on to sketch a different theory to explain each kind.

Drawing on the above analyses, I propose that there are at least three kinds of consciousness that philosophers associate with being a subject of experience. I call them modes of subjectivity. The first mode is the *qualitative consciousness* of sensations and perceptions. It involves qualia of the kind described by Humphrey,

Katz, and Lycan – for instance, 'what it is like' to feel an intense pain or a persistent itch, to see a patch of red, or to hear a loud sound. This mode of subjectivity is passive. For instance, qualia are produced when a creature's sensory organs are stimulated; but the creature is not required to conceptualise, classify, or otherwise reflect on the qualia. This accounts for the intuitive sense in which qualia are 'given', 'present', or 'immediate' to the subject. Infants and dogs are subjects in this qualitative mode; if any late-stage human foetuses can feel pain, then they are conscious in this way too.

The second mode of subjectivity is *perspectival consciousness*. Creatures which are conscious in this way have a perspective that orientates their mental states. They develop a perspective when they are aware of their own mental states. This awareness affects what they go on to feel and do. As infants grow up, they turn into subjects in this perspectival mode. They automatically become aware of some of their own sensations and perceptions. Eventually awareness extends to their own affective and cognitive states. If Lycan (2004) is right, then these subjects conceptualise and classify their own mental states, at least in a rudimentary way. They are aware of what the states are and whether they are new or familiar, welcome or worrisome, etc. This process is aided by their attention to the physical environment, and the intervention of their social community.

The third mode is *first-personal consciousness*. Creatures with this kind of consciousness are able to examine their mental states 'from the inside'. Using introspection, they can attend to and reflect on these mental states as their own. They can also control some of these mental states. In the first-personal mode,

subjects have experiences that are imbued with a sense of self. However, first-personal consciousness is not the same as self-consciousness. Recall, from §2.2, that Block (2002a) defines self-consciousness as the 'possession of the concept of the self and the ability to use this concept in thinking about oneself' (287). First-personal consciousness also requires a concept of the self, but it is used by the subject in attending to, reflecting on, or controlling his own mental states.

I will not defend this provisional classification here. My aim is only to show that, in conceptual terms, we can plausibly distinguish more than one mode of subjectivity. For now, let me indicate two other sources of support for my classification.[15] First, Nagel (1974) provides some indirect support. Unlike other philosophers who invoke his 'what it is like' phrase, Nagel never speaks of the qualitative, perspectival, and first-personal in the same breath. Instead he focuses on what I call perspectival consciousness: 'every subjective phenomenon is essentially connected with a single point of view' (520). According to him, facts about 'what it is like' to be a creature are facts that 'embody a particular point of view' belonging to that creature (522).

Elsewhere Nagel (1974) briefly refers to the other two modes of subjectivity that I distinguish. Nagel mentions the 'quality' of another person's experience, but does not equate it with the way that some facts about his experience seem accessible only from his 'point of view' (522). Later, in a footnote, he refers more specifically to the 'quality' of his visual experience when he looks at the portrait of Mona Lisa

[15] A third, more distant, source is the analyses of Descartes' theory of consciousness in Brown (2007) and Lähteenmäki (2007). The latter distinguishes three orders of consciousness in Descartes' theory: (a) the *rudimentary consciousness* of occurrent experiences; (b) the *reflexive consciousness* of experiences which contain the 'intellectual perception of an initial perception or act of will'; and (c) the *reflective consciousness* 'acquired through deliberate attentive reflection' on our thoughts (194).

(526). In his arguments on the subjective character of experience, Nagel only invokes the 'first person' twice (522 and 525). For him, it is a stance that we can adopt towards our own mental states, which is to be contrasted with the third-person stance that we use to describe the world. We can learn concepts 'in the first person' for describing our own mental states; the use of such concepts depend on our point of view and perceptual apparatus.

Second, some philosophers and scientists who investigate borderline conscious creatures broach similar conceptual distinctions. I cite two examples here. In his discussion of cephalopods, Godfrey-Smith (2013) argues that the 'subjective side of our mental lives' consists of at least three phenomena, which are 'probably blended together into each other in complicated ways' (8). These are: the kind of experiences that are 'undeniably *felt*', such as pain[16]; the kind of experiences in which 'many different kinds of sensory information are combined, and the present is experienced as related to a recent past'; and reflective or higher-order consciousness, through which 'we turn our gaze on our own thoughts' (7-8). Godfrey-Smith sees no evidence currently for thinking that a creature without the more integrated kind of experiences cannot feel pain.

Derbyshire and Raja (2011) discuss the development of pain in late-stage human foetuses and infants. Their classification is less clear-cut, but it bears some resemblance to mine. They distinguish qualia from the experience of qualia.[17] For

[16] He mistakenly attributes to Nagel (1974) the question of 'what it *feels* like to be' a creature (7, original italics).

[17] See also Lycan (1995), who introduces a similar distinction. He defines qualia as the phenomenal characters of sensory states: 'One registers such a quale whenever one perceives a colored object as such' (263). Then he adds: 'There is certainly a sense in which one has not experienced phenomenal

instance, the qualia of pain may be produced by neuronal activity in late-stage human foetuses or newborns. But, according to Derbyshire and Raja, the presence of these qualia is not sufficient for painful experience.

> To be experienced, however, qualia require a subjective element, a person occupying a point of view from where qualia can be 'observed'. If qualia require a point of view to be experienced, then qualia cannot exist in the fetus because there is no means by which a point of view can be gained or shared from within the womb. (246)

The *non sequitur* here is only apparent. Derbyshire and Raja go on to clarify that qualia 'do exist' even in late-stage foetuses, 'but they exist in the form of non-existence' (247). By this, they mean that the qualia are not experienced by the foetuses since the foetuses lack a point of view. Derbyshire and Raja suggest that infants gradually develop a point of view as a result of their social interactions with adults. This point of view enables them to discriminate and experience the qualia of pain amidst the 'cacophony of possible experiences' produced in the brain. They thereby gain 'basic sentient experience' (250). In Derbyshire and Raja's account, these infants attain first-person knowledge of their own painful experiences only after further development.

Whether there exist three or more modes of subjectivity remains to be discovered by consciousness science. I have already argued that, in the impasse on the neural structure of phenomenal consciousness, scientists can legitimately stipulate two kinds to be the referent of the concept of phenomenal consciousness: one kind defined by transient neural synchrony and local recurrent processing, another

---

red, or *felt* pain, unless one is aware of the redness or the pain. To experience a sensation in that fuller sense, one must both have the relevant quale and notice it introspectively.'

defined by the sustained neural synchrony and global recurrent processing studied by global workspace theorists. My conceptual analysis may help to make sense of these two kinds, which were found through empirical investigation. Might these kinds approximate what I am calling qualitative consciousness and perspectival consciousness? And might there exist another kind that is first-personal consciousness?

Of course, there might even be a more fundamental experiential essence underlying the qualitative, perspectival, and first-personal kinds of consciousness – just as there might be a more fundamental basis of biodiversity uniting the three different kinds of species. Nothing in my conceptual analysis can rule this out. I anticipate that many more cycles of conceptual clarification and empirical investigation are needed before we get any clear understanding of the multiple kinds associated with our concept of phenomenal consciousness.

6.7 Conclusion

In this final chapter, I defended my unified approach to borderline consciousness, phenomenal consciousness, and artificial consciousness. I addressed four key objections against my proposal for stipulating consciousness. These objections claim that stipulation is naturalistically unsound, metaphysically presumptuous about vagueness, semantically shifty, and subjectively absurd.

Let me sum up my defence. First, I drew on my arguments about species to show

that empirically constrained stipulation is naturalistically sound on both methodological and metaphysical grounds. Following the norms of stipulation that I derived in Chapter 5, I argued that the conditions of stipulation are already in place in the impasse on the neural structure of phenomenal consciousness. They are, however, not so in the impasse on the possibility of artificial consciousness. Second, I emphasised that my proposal does not draw support from any controversial theory of vagueness. Yet it can be interpreted to be consistent with all four theories of vagueness in the philosophical literature.

Third, I distinguished my proposal for stipulating consciousness from arbitrary stipulations that obviously change the subject. Then, drawing an analogy with stipulating species, I denied that stipulating consciousness abruptly changes the subject of consciousness science. I also used two models of reference to show that scientists need not abruptly or arbitrarily change the subject even if their stipulations shift the reference of the term 'phenomenal consciousness'. Fourth, I distinguished the introspective and intuitive bases for the claim that phenomenal consciousness is metaphysically determinate. I interpreted my proposal so that it is consistent with our introspective and intuitive views. Then I offered some methodological and conceptual challenges to these views.

Finally, I urged that a more thorough conceptual analysis of phenomenal consciousness can help to make sense of the multiplicity of subjectivity. I showed that a significant range of philosophers and scientists acknowledge that there may be more than one kind associated with the concept of phenomenal consciousness. Drawing on their analyses, I proposed that there are at least three kinds of

consciousness that philosophers associate with being a subject of experience: qualitative consciousness, perspectival consciousness, and first-personal consciousness.

Throughout this chapter, I highlighted weaknesses in my philosophical defence. These weaknesses point to areas that deserve more philosophical study. Let me highlight five areas that seem promising. First, we need to assess other examples of stipulation in biological and chemical classification. These examples may indicate other conditions for stipulation and other bases of stipulation that are relevant to consciousness science. Second, we need to compare empirically constrained stipulation with other possible solutions to the multiple kinds problem. This will help to determine if my proposal for stipulating consciousness is the best solution to the multiple kinds problem in consciousness science. Third, we should examine the fate of more borderline cases in scientific classification. This will help to distinguish cases that can be re-classified by stipulation from cases that cannot. Fourth, we should trace the history of the term 'phenomenal consciousness' in more detail. By studying how the term is introduced into consciousness science and spread among scientists and philosophers, we will be able to distinguish the roles of partial reference and contextual reference in the semantic structures associated with the term. Fifth, we should test my provisional classification of qualitative consciousness, perspectival consciousness, and first-personal consciousness. It ought to be revised in the light of any new empirical insight on phenomenal consciousness. This may, in turn, guide further investigation into the multiplicity that is subjectivity.

Bibliography

Aizawa, Kenneth, and Carl Gillett. 2009a. 'Levels, Individual Variation, and Massive Multiple Realization in Neurobiology'. In *Oxford Handbook of Philosophy and Neuroscience*, edited by John Bickle, 529–81. New York: Oxford University Press.

———. 2009b. 'The (Multiple) Realization of Psychological and Other Properties in the Sciences'. *Mind and Language* 24 (2): 181–208.

———. 2011. 'The Autonomy of Psychology in the Age of Neuroscience'. In *Causality in the Sciences*, edited by Phyllis Illari and Federica Russo, 202–23. New York: Oxford University Press.

Åkerman, Jonas. 2012. 'Contextualist Theories of Vagueness'. *Philosophy Compass* 7 (7): 470–80.

Allen, Colin. 2011. 'Animal Consciousness'. In *The Stanford Encyclopedia of Philosophy* (Winter 2011 Edition), edited by Edward N. Zalta. http://plato.stanford.edu/archives/win2011/entries/consciousness-animal/.

Allen, Colin, and Michael Trestman. 2015. 'Animal Consciousness'. In *The Stanford Encyclopedia of Philosophy* (Summer 2015 Edition), edited by Edward N. Zalta. http://plato.stanford.edu/archives/sum2015/entries/consciousness-animal/.

Alston, William P. 1964. *Philosophy of Language*. New Jersey: Prentice-Hall.

Alter, Torin A., and Sven Walter (eds.). 2007. *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism*. Oxford: Oxford University Press.

Anand, K.J.S. 1998. 'Clinical Importance of Pain and Stress in Preterm Neonates'. *Biology of the Neonate* 73 (1): 1–9.

———. 2006. 'Fetal Pain?' *Pain: Clinical Updates* XIV (2).

———. 2007. 'Consciousness, Cortical Function, and Pain Perception in Nonverbal Humans'. *Behavioral and Brain Sciences* 30 (1): 82–83.

Anand, K.J.S., and K.D. Craig. 1996. 'New Perspectives on the Definition of Pain'. *Pain* 67 (1): 3–6.

Anand, K.J.S., and P.R. Hickey. 1987. 'Pain and Its Effects in the Human Neonate and Fetus'. *The New England Journal of Medicine* 317 (21): 1321–29.

Andreas, Holger. 2013. 'Theoretical Terms in Science'. In *The Stanford Encyclopedia of Philosophy* (Summer 2013 Edition), edited by Edward N. Zalta. http://plato.stanford.edu/archives/sum2013/entriesheoretical-terms-science/.

Andrews, K., and M. Fitzgerald. 1994. 'The Cutaneous Withdrawal Reflex in Human Neonates: Sensitization, Receptive Fields, and the Effects of Contralateral Stimulation'. *Pain* 56 (1): 95–101.

Antony, Michael V. 2006a. 'Papineau on the Vagueness of Phenomenal Concepts'. *Dialectica* 60 (4): 475–83.

———. 2006b. 'Vagueness and the Metaphysics of Consciousness'. *Philosophical Studies* 128 (3): 515–38.

———. 2008. ‘Are Our Concepts Conscious State and Conscious Creature Vague?’ *Erkenntnis* 68 (2): 239–63.
Appel, Mirjam, and Robert W. Elwood. 2009. ‘Motivational Trade-Offs and Potential Pain Experience in Hermit Crabs’. *Applied Animal Behaviour Science* 119 (1-2): 120–24.
Avramides, Anita. 2000. *Other Minds*. Oxford: Routledge.
Baars, Bernard J. 1997. ‘In the Theatre of Consciousness: Global Workspace Theory, a Rigorous Scientific Theory of Consciousness’. *Journal of Consciousness Studies* 4 (4): 292–309.
———. 2005. ‘Subjective Experience Is Probably Not Limited to Humans: The Evidence from Neurobiology and Behavior’. *Consciousness and Cognition* 14 (1): 7–21.
Baker, Alan. 2013. ‘Simplicity’. In *The Stanford Encyclopedia of Philosophy* (Fall 2013 Edition), edited by Edward N. Zalta. http://plato.stanford.edu/archives/fall2013/entries/simplicity/.
Barnes, Elizabeth. 2009. ‘Vagueness’. In *The Routledge Companion to Metaphysics*, edited by Robin Le Poidevin, 370–81. New York: Routledge.
———. 2010. ‘Arguments Against Metaphysical Indeterminacy and Vagueness’. *Philosophy Compass* 5 (11): 953–64.
Bayne, Tim. 2007. ‘Conscious States and Conscious Creatures: Explanation in the Scientific Study of Consciousness’. *Philosophical Perspectives* 21 (1): 1–22.
———. 2009a. ‘Consciousness’. In *The Routledge Companion to Philosophy of Psychology*, edited by John Symons and Paco Calvo, 477–94. Oxford: Routledge.
———. 2009b. ‘Other Minds’. In *The Oxford Companion to Consciousness*, edited by Tim Bayne, Axel Cleeremans, and Patrick Wilken, 490-2. Oxford: Oxford University Press.
———. 2009c. ‘Perception and the Reach of Phenomenal Content’. *Philosophical Quarterly* 59 (236): 385–404.
Bechtel, William P., and Robert N. McCauley. 1999. ‘Heuristic Identity Theory (or Back to the Future): The Mind-Body Problem Against the Background of Research Strategies in Cognitive Neuroscience’. In *Proceedings of the 21st Annual Meeting of the Cognitive Science Society*, edited by Martin Hahn and S. C. Stoness, 67–72. New Jersey: Lawrence Erlbaum.
Bechtel, William P., and Jennifer Mundale. 1999. ‘Multiple Realizability Revisited: Linking Cognitive and Neural States’. *Philosophy of Science* 66 (2): 175–207.
Beebee, Helen, and Nigel Sabbarton-Leary. 2010. ‘Introduction’. In *The Semantics and Metaphysics of Natural Kinds*, edited by Helen Beebee and Nigel Sabbarton-Leary, 1–24. New York: Routledge.
———. (eds.). 2010. *The Semantics and Metaphysics of Natural Kinds*. New York: Routledge.
Benatar, D., and M. Benatar. 2001. ‘A Pain in the Fetus: Toward Ending Confusion about Fetal Pain’. *Bioethics* 15: 57–76.
Bermudez, Jose Luis. 2004. ‘Vagueness, Phenomenal Concepts and Mind-Brain Identity’. *Analysis* 64 (2): 131–39.

Besson, Corine. 2010. ‘Rigidity, Natural Kind Terms, and Metasemantics’. In *The Semantics and Metaphysics of Natural Kinds*, edited by Helen Beebee and Nigel Sabbarton-Leary, 25–44. New York: Routledge.

Bickle, John. 2013. ‘Multiple Realizability’. In *The Stanford Encyclopedia of Philosophy* (Spring 2013 Edition), edited by Edward N. Zalta. http://plato.stanford.edu/archives/spr2013/entries/multiple-realizability/.

Bird, Alexander. 2007. ‘A Posteriori Knowledge of Natural Kind Essences: A Defence’. *Philosophical Topics* 35: 293–312.

———. 2009. ‘Essences and Natural Kinds’. In *The Routledge Companion to Metaphysics*, edited by Robin Le Poidevin, 497–506. Oxford: Routledge.

———. 2010. ‘Discovering the Essences of Natural Kinds’. In *The Semantics and Metaphysics of Natural Kinds*, edited by Helen Beebee and Nigel Sabbarton-Leary, 124–36. New York: Routledge.

Bird, Alexander, and Emma Tobin. 2015. ‘Natural Kinds’, *The Stanford Encyclopedia of Philosophy* (Spring 2015 Edition), edited by Edward N. Zalta. http://plato.stanford.edu/archives/spr2015/entries/natural-kinds/.

Black, Max. 1937. ‘Vagueness: An Exercise in Logical Analysis’. *Philosophy of Science* 4: 427–55.

Block, Ned. 1978. ‘Troubles with Functionalism’. *Minnesota Studies in the Philosophy of Science* 9: 261–325. All references are to the version in Block (2007), 63–102.

———. 1980. ‘What Intuitions About Homunculi Don’t Show’. *Behavioral and Brain Sciences* 3 (3): 425. All references are to the version in Block (2007), 103–8.

———. 1992. ‘Begging the Question Against Phenomenal Consciousness’. *Behavioral and Brain Sciences* 15 (2): 205–6. All references are to the version in Block, Flanagan, and Güzeldere eds. (1997), 175–83.

———. 1995a. ‘On a Confusion about the Function of Consciousness’. *Behavioral and Brain Sciences* 18: 227–47. All references are to the version in Block (2007), 159–214.

———. 1995b. ‘How Many Concepts of Consciousness?’ *Behavioral and Brain Sciences* 18 (2): 227–47. All references are to the version in Block (2007), 215–48 .

———. 2001. ‘Paradox and Cross-Purposes in Recent Work on Consciousness’. *Cognition* 79 (1-2): 197–219. All references are to the version in Block (2007), 311–38.

———. 2002a. ‘Concepts of Consciousness’. In *Philosophy of Mind: Classical and Contemporary Readings*, edited by David J. Chalmers. Oxford: Oxford University Press. All references are to the version in Block (2007), 275–96.

———. 2002b. ‘The Harder Problem of Consciousness’. *The Journal of Philosophy* 99 (8): 391–425. All references are to the version in Block (2007), 397–434.

———. 2005. ‘Two Neural Correlates of Consciousness’. *Trends in Cognitive Sciences* 9 (2): 46–52. All references are to the version in Block (2007), 343–62.

———. 2007. *Consciousness, Function, and Representation: Collected Papers Volume 1*. Cambridge, MA: MIT Press.

———. 2007a. ‘Consciousness, Accessibility, and the Mesh Between Psychology and Neuroscience’. *Behavioral and Brain Sciences* 30 (5): 481–548.

———. 2007b. ‘Overflow, Access, and Attention’. *Behavioral and Brain Sciences* 30 (5-6): 530–48.
———. 2008. ‘Consciousness and Cognitive Access’. *Proceedings of the Aristotelian Society* 108 (1): 289–317.
———. 2011. ‘Perceptual Consciousness Overflows Cognitive Access’. *Trends in Cognitive Sciences* 15 (12): 567–75.
Block, Ned, Owen Flanagan, and Güven Güzeldere (eds.). 1997. *The Nature of Consciousness: Philosophical Debates*. Cambridge, MA: MIT Press.
Block, Ned, and Robert Stalnaker. 1999. ‘Conceptual Analysis, Dualism, and the Explanatory Gap’. *Philosophical Review* 108 (1): 1–46.
Boyd, Richard. 1980. ‘Scientific Realism and Naturalistic Epistemology’. *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association* 1980 (2): 613–62.
———. 1988. ‘How to Be a Moral Realist’. In *Essays on Moral Realism*, edited by Geoffrey Sayre-McCord. New York: Cornell University Press.
———. 1991. ‘Realism, Anti-Foundationalism and the Enthusiasm for Natural Kinds’. *Philosophical Studies* 61 (1/2): 127–48.
———. 1993. ‘Metaphor and Theory Change: What Is “Metaphor” a Metaphor For?’ In *Metaphor and Thought* (2nd Edition), edited by A. Ortony, 481–532. New York: Cambridge University Press.
———. 1999a. ‘Homeostasis, Species, and Higher Taxa’. In *Species: New Interdisciplinary Essays*, edited by Robert A. Wilson, 141–85. Cambridge, MA: MIT Press.
———. 1999b. ‘Kinds as the “Workmanship of Men”: Realism, Constructivism, and Natural Kinds’. In *Rationality, Realism, Revision: Proceedings of the 3rd International Congress of the Society for Analytical Philosophy*, edited by Julian Nida-Rümelin, 52–89. Berlin: de Gruyter.
———. 1999c. ‘Kinds, Complexity and Multiple Realization’. *Philosophical Studies* 95 (1-2): 67–98.
———. 2010. ‘Realism, Natural Kinds, and Philosophical Methods’. In *The Semantics and Metaphysics of Natural Kinds*, edited by Helen Beebee and Nigel Sabbarton-Leary, 212–34. New York: Routledge.
———. 2013. ‘What of Pragmatism with the World Here?’ In *Reading Putnam*, edited by Maria Baghraman, 39–94. New York: Routledge.
Braithwaite, V.A. 2010. *Do Fish Feel Pain?* Oxford: Oxford University Press.
Braithwaite, V. A., and P. Boulcott. 2007. ‘Pain Perception, Aversion and Fear in Fish’. *Diseases of Aquatic Organisms* 75 (2): 131–38.
Brigandt, Ingo. 2003a. ‘Species Pluralism Does Not Imply Species Eliminativism’. *Philosophy of Science* 70 (5): 1305–16.
———. 2003b. ‘Species Pluralism Does Not Imply Species Eliminativism’. *Philosophy of Science* 70 (5): 1305–16.
Brown, Culum. 2014. ‘Fish Intelligence, Sentience and Ethics’. *Animal Cognition*, June, 1–17.
Brown, Deborah. 2007. ‘Augustine and Descartes on the Function of Attention in Perceptual Awareness’. In *Consciousness: From Perception to Reflection in the History of Philosophy*, edited by Sara Heinämaa, Vili Lähteenmäki, and Pauliina Remes, 153–76. Dordrecht: Springer.

Burge, Tyler. 1993. ‘Concepts, Definitions, and Meaning’. *Metaphilosophy* 24 (4): 309–25.
———. 1997. ‘Two Kinds of Consciousness’. In *The Nature of Consciousness: Philosophical Debates*, edited by Ned Block, Owen Flanagan, and Güven Güzeldere, 427–33. Cambridge, MA: MIT Press.
———. 2007. *Foundations of Mind: Philosophical Essays (Volume 2)*. New York: Oxford University Press.
Cabanac, Michel, Arnaud J. Cabanac, and André Parent. 2009. ‘The Emergence of Consciousness in Phylogeny’. *Behavioural Brain Research* 198 (2): 267–72.
Chalmers, David J. 1996. *The Conscious Mind*. Oxford: Oxford University Press.
———. (ed.). 2002. *Philosophy of Mind: Contemporary and Classical Readings*. Oxford: Oxford University Press.
———. 2003. ‘Consciousness and Its Place in Nature’. In *Blackwell Guide to the Philosophy of Mind*, edited by Stephen P. Stich and Ted A. Warfield, 102–42. Blackwell. All references are to the version in Chalmers ed. (2002), 247–72.
———. 2007. ‘Phenomenal Concepts and the Explanatory Gap’. In *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism*, edited by Torin A. Alter and Sven Walter, 167-94. New York: Oxford University Press.
Chandroo, Kristopher Paul, Stephanie Yue, and Richard David Moccia. 2004. ‘An Evaluation of Current Perspectives on Consciousness and Pain in Fishes’. *Fish and Fisheries* 5 (4): 281–95.
Chin, Chuanfei. 2007. ‘Predicting and Projecting Pain: A Philosophical Study of Pain Science’. Oxford: University of Oxford.
———. 2011. ‘Models as Interpreters (with a Case Study from Pain Science)’. *Studies In History and Philosophy of Science Part A* 42 (2): 303–12.
Church, Jennifer. 1995. ‘Fallacies or Analyses?’ *Behavioral and Brain Sciences* 18 (02): 251–52. All references are to the version in Block, Flanagan, and Güzeldere eds. (1997), 425–6.
Claridge, M.F., H.A. Dawah, and M.R. Wilson (eds.). 1997. *Species: The Units of Biodiversity*. London: Chapman & Hall.
Coyne, Jerry A., and H. Allen Orr. 2004. ‘Speciation: A Catalogue and Critique of Species Concepts’. In *Philosophy of Biology: An Anthology*, edited by Alex Rosenberg and Robert Arp, 272–92. Oxford: Wiley-Blackwell.
Cracraft, Joel. 1983. ‘Species Concepts and Speciation Analysis’. In *Current Ornithology*, edited by Richard F. Johnston, 159–87. New York: Springer.
———. 2000. ‘Species Concepts in Theoretical and Applied Biology: A Systematic Debate with Consequences’. In *Species Concepts and Phylogenetic Theory: A Debate*, edited by Q.D. Wheeler and R. Meier, 3–14. New York: Columbia University Press.
Craig, K.D., M.F. Whitfield, R.V. Grunau, J. Linton, and H.D. Hadjistavropoulos. 1993. ‘Pain in the Preterm Neonate: Behavioural and Physiological Indices’. *Pain* 52 (3): 287–99.
Craver, Carl F. 2009. *Explaining the Brain*. Oxford: Oxford University Press.
Crick, Francis, and Christof Koch. 1990. ‘Toward a Neurobiological Theory of Consciousness’. *Seminars in the Neurosciences* 2: 263–75.
———. 2003. ‘A Framework for Consciousness’. *Nature Neuroscience* 6: 119–26.

De Brigard, Felipe, and Jesse Prinz. 2010. ‘Attention and Consciousness’. *Wiley Interdisciplinary Reviews: Cognitive Science* 1 (1): 51–59.
Dehaene, Stanislas, Jean-Pierre Changeux, Lionel Naccache, Jérôme Sackur, and Claire Sergent. 2006. ‘Conscious, Preconscious, and Subliminal Processing: A Testable Taxonomy’. *Trends in Cognitive Sciences* 10 (5): 204–11.
Dehaene, Stanislas, and Lionel Naccache. 2001. ‘Towards a Cognitive Neuroscience of Consciousness: Basic Evidence and a Workspace Framework’. *Cognition* 79 (1-2): 1–37.
Del Cul, Antoine, Sylvain Baillet, and Stanislas Dehaene. 2007. ‘Brain Dynamics Underlying the Nonlinear Threshold for Access to Consciousness’. *PLoS Biol* 5 (10): e260.
Dennett, Daniel C. 1988. ‘Quining Qualia’. In *Consciousness in Contemporary Science*, edited by A. Marcel and E. Bisiach, 43–77. Oxford: Oxford University Press. All references are to the version in Block, Flanagan, and Güzeldere eds. (1997), 619–42.
———. 1991. *Consciousness Explained*. New York: Little, Brown & Co.
———. 1993. ‘The Message Is: There Is No Medium’. *Philosophy and Phenomenological Research* 53 (4): 919–31.
———. 1995. ‘Animal Consciousness: What Matters and Why’. *Social Research* 62 (3): 690.
———. 1996. *Kinds of Minds: Toward an Understanding of Consciousness*. New York: Basic Books.
de Queiroz, Kevin. 1999. ‘The General Lineage Concept of Species and the Defining Properties of the Species Category’. In *Species: New Interdisciplinary Essays*, edited by Robert A. Wilson, 49–89. Cambridge, MA: MIT Press.
———. 2007. ‘Species Concepts and Species Delimitation’. *Systematic Biology* 56 (6): 879–86.
De Sousa, Ronald. 2004. ‘Twelve Varieties of Subjectivity’. In *Language, Knowledge, and Representation: Proceedings of the Sixth International Colloquium on Cognitive Science*, edited by Jésus M. Larrazabal and Luis A. Pérez Miranda, 93–102. Dordrecht: Springer Science.
Derbyshire, Stuart. 2001. ‘Fetal Pain: An Infantile Debate’. *Bioethics* 15 (1): 77–84.
———. 2006. ‘Can Fetuses Feel Pain?’ *BMJ (Clinical Research Ed.)* 332 (7546): 909–12.
Derbyshire, Stuart, and Anand Raja. 2011. ‘On the Development of Painful Experience’. *Journal of Consciousness Studies* 18 (9-10): 9–10.
Devitt, Michael. 2008. ‘Resurrecting Biological Essentialism’. *Philosophy of Science* 75 (3): 344–82.
———. 2011. ‘Natural Kinds and Biological Realisms’. In *Carving Nature at Its Joints*, edited by Michael O’Rourke, Joseph Keim Campbell, and Matthew H. Slater, 155-74. Cambridge, MA: MIT Press.
Devitt, Michael, and Kim Sterelny. 1999. *Language and Reality: An Introduction to the Philosophy of Language* (Second Edition). Cambridge, MA: MIT Press.
Devor, Marshall. 2007. ‘Pain, Cortex, and Consciousness’. *Behavioral and Brain Sciences* 30 (01): 89–90.

Donnellan, Keith. 1983. ‘Kripke and Putnam on Natural Kind Terms’. In *Knowledge and Mind*, edited by C. Ginet and S. Shoemaker, 84–104. Oxford: Oxford Univresity Press.
Douven, Igor. 2011. ‘Abduction’. In *The Stanford Encyclopedia of Philosophy* (Spring 2011 Edition), edited by Edward N. Zalta. http://plato.stanford.edu/archives/spr2011/entries/abduction/.
Dretske, Fred. 2004. ‘Change Blindness’. *Philosophical Studies* 120 (1-3): 1–18.
———. 2007. ‘What Change Blindness Teaches About Consciousness’. *Philosophical Perspectives* 21 (1): 215–20.
Dunlop, Rebecca, Sarah Millsopp, and Peter Laming. 2006. ‘Avoidance Learning in Goldfish (*Carassius Auratus*) and Trout (*Oncorhynchus Mykiss*) and Implications for Pain Perception’. *Applied Animal Behaviour Science* 97 (2-4): 255–71.
Dupré, John. 1981. ‘Natural Kinds and Biological Taxa’. *The Philosophical Review* 90 (1): 66–90.
———. 1993. *The Disorder of Things: Metaphysical Foundations of the Disunity of Science*. Cambridge, MA: Harvard University Press.
———. 2004. ‘Review: Joseph LaPorte’s *Natural Kinds and Conceptual Change*’. *Notre Dame Philosophical Reviews*.
Earman, John, and Arthur Fine. 1977. ‘Against Indeterminacy’. *The Journal of Philosophy* 74 (9): 535–38.
Edelman, David B., Bernard J. Baars, and Anil K. Seth. 2005. ‘Identifying Hallmarks of Consciousness in Non-Mammalian Species’. *Consciousness and Cognition* 14 (1): 169–87.
Edelman, David B., and Anil K. Seth. 2009. ‘Animal Consciousness: A Synthetic Approach’. *Trends in Neurosciences* 32 (9): 476–84.
Ellis, Brian. 2001. *Scientific Essentialism*. Cambridge: Cambridge University Press.
———. 2008. ‘Essentialism and Natural Kinds’. In *The Routledge Companion to Philosophy of Science*, edited by Stathis Psillos and Martin Curd, 139–48. Oxford: Routledge.
Elwood, Robert W. 2011. ‘Pain and Suffering in Invertebrates?’ *ILAR Journal* 52 (2): 175–84.
Elwood, Robert W., and Mirjam Appel. 2009. ‘Pain Experience in Hermit Crabs?’ *Animal Behaviour* 77 (5): 1243–46.
Ereshefsky, Marc. 1992. ‘Eliminative Pluralism’. *Philosophy of Science* 59 (4): 671–90.
———. 1998. ‘Species Pluralism and Anti-Realism’. *Philosophy of Science* 65 (1): 103–20.
———. 2010a. ‘Darwin’s Solution to the Species Problem’. *Synthese* 175 (3): 405–25.
———. 2010b. ‘Species’. In *The Stanford Encyclopedia of Philosophy* (Spring 2010 Edition), edited by Edward N. Zalta. http://plato.stanford.edu/archives/spr2010/entries/species/.
———. 2010c. ‘Species, Taxonomy, and Systematics’. In *Philosophy of Biology: An Anthology*, edited by Alex Rosenberg and Robert Arp, 255–71. Oxford: Wiley-Blackwell.
———. 2010d. ‘What’s Wrong with the New Biological Essentialism’. *Philosophy of Science* 77 (5): 674–85.

———. 2010e. ‘Microbiology and the Species Problem’. *Biology & Philosophy* 25 (4): 553–68.
Farah, Martha J. 2002. ‘Emerging Ethical Issues in Neuroscience’. *Nature Neuroscience* 5 (11): 1123–29.
———. 2005. ‘Neuroethics: The Practical and the Philosophical’. *Trends in Cognitive Sciences* 9 (1): 34–40.
———. 2008. ‘Neuroethics and the Problem of Other Minds: Implications of Neuroscience for the Moral Status of Brain-Damaged Patients and Nonhuman Animals ’. *Neuroethics* 1 (1): 9-18.
Feinberg, Todd E., and Jon Mallatt. 2013. ‘The Evolutionary and Genetic Origins of Consciousness in the Cambrian Period over 500 Million Years Ago’. *Consciousness Research* 4: 667.
Field, Hartry. 1973. ‘Theory Change and the Indeterminacy of Reference’. *The Journal of Philosophy* 70 (14): 462–81.
Fodor, Jerry A. 1974. ‘Special Sciences (Or: The Disunity of Science as a Working Hypothesis)’. *Synthese* 28 (2): 97–115.
French, Steven, and Michela Massimi. 2013. ‘Philosophy of Science A Personal Peek into the Future’. *Metaphilosophy* 44 (3): 230–40.
Friedman, Michael. 1997. ‘Philosophical Naturalism’. *Proceedings and Addresses of the American Philosophical Association* 71 (2): 5–21.
Gherardi, Francesca. 2009. ‘Behavioural Indicators of Pain in Crustacean Decapods’. *Annali dell’Istituto Superiore Di Sanità* 45 (4): 432–38.
Ghiselin, Michael T. 1974. ‘A Radical Solution to the Species Problem’. *Systematic Zoology* 23: 536–44.
———. 1997. *Metaphysics and the Origin of Species*. Albany: State University of New York Press.
Gillett, Carl. 2003. ‘The Metaphysics of Realization, Multiple Realizability, and the Special Sciences’. *The Journal of Philosophy* 100 (11): 591–603.
Glover, Vivette, and Nicholas M. Fisk. 1996. ‘Do Fetuses Feel Pain? We Don’t Know; Better to Err on the Safe Side from Mid-Gestation’. *BMJ (Clinical Research Ed.)* 313 (7060): 796.
———. 1999. ‘Fetal Pain: Implications for Research and Practice’. *British Journal of Obstetrics and Gynaecology* 106 (9): 881–86.
Godfrey-Smith, Peter. 2003. *Theory and Reality: An Introduction to the Philosophy of Science*. Chicago, ILL: University Of Chicago Press.
———. 2013. ‘Cephalopods and the Evolution of the Mind.’ *Pacific Conservation Biology* 19 (1): 4–9.
Griffiths, Paul E. 1997. *What Emotions Really Are: The Problem of Psychological Categories*. Chicago, ILL: University of Chicago Press.
Griffiths, Paul E. 2004. ‘Emotions as Natural and Normative Kinds’. *Philosophy of Science* 71 (5): 901–11.
Hacking, Ian. 1991. ‘A Tradition of Natural Kinds’. *Philosophical Studies* 61 (1-2): 109–26.
———. 2007a. ‘Natural Kinds: Rosy Dawn, Scholastic Twilight’. *Royal Institute of Philosophy Supplements* 82 (Supplement 61): 203–39.
———. 2007b. ‘Putnam’s Theory of Natural Kinds and Their Names Is Not the Same as Kripke’s’. *Principia: Revista Internacional de Epistemologica* 11: 1–24.

———. 2007c. ‘The Contingencies of Ambiguity’. *Analysis* 67 (296): 269–77.
Hall, Ned. 2012a. ‘David Lewis’s Metaphysics’. In *The Stanford Encyclopedia of Philosophy* (Fall 2012 Edition), edited by Edward N. Zalta. http://plato.stanford.edu/archives/fall2012/entries/lewis-metaphysics/.
———. 2012b. ‘The Natural/Non-Natural Distinction: Supplement to David Lewis’ Metaphysics’. In *The Stanford Encyclopedia of Philosophy* (Fall 2012 Edition), edited by Edward N. Zalta. http://plato.stanford.edu/archives/fall2012/entries/lewis-metaphysics/.
Hart, W.D. 1992. ‘Hat-Tricks and Heaps’. *Philosophical Studies* 33 (1): 1–24.
Hawley, Katherine, and Alexander Bird. 2011. ‘What Are Natural Kinds?’. *Philosophical Perspectives* 25 (1): 205–21.
Hempel, Carl G. 1965. ‘Fundamentals of Taxonomy’. In *Aspects of Scientific Explanation, And Other Essays in the Philosophy of Science*, 137–55. New York: The Free Press.
Hendry, Robin. 2010. ‘The Elements and Conceptual Change’. In *The Semantics and Metaphysics of Natural Kinds*, edited by Helen Beebee and Nigel Sabbarton-Leary, 137-158. Abingdon: Routledge.
———. 2012. ‘Chemical Substances and the Limits of Pluralism’. *Foundations of Chemistry* 14 (1): 55–68.
Hohwy, Jacob. 2004. ‘Evidence, Explanation, and Experience: On the Harder Problem of Consciousness’. *The Journal of Philosophy* 101 (5): 242–54.
Huebner, Bryce. 2010. ‘Commonsense Concepts of Phenomenal Consciousness: Does Anyone Care about Functional Zombies?’ *Phenomenology and the Cognitive Sciences* 9 (1): 133–55.
Hull, David. 1976. ‘Are Species Really Individuals?’ *Systematic Zoology* 25: 174–91.
———. 1978. ‘A Matter of Individuality’. *Philosophy of Science* 45 (3): 335–60.
———. 1987. ‘Genealogical Actors in Ecological Roles’. *Biology and Philosophy* 2 (2): 168.
Humphrey, Nicholas. 1995. ‘Blocking out the Distinction between Sensation and Perception: Superblindsight and the Case of Helen’. *Behavioral and Brain Sciences* 18 (02): 257–58.
Hyslop, Alec. 2015. ‘Other Minds’. *Stanford Encyclopedia of Philosophy* (Fall 2015 Edition), edited by Edward N. Zalta. http://plato.stanford.edu/archives/fall2015/entries/other-minds/.
Irvine, Elizabeth. 2013. *Consciousness as a Scientific Concept: A Philosophy of Science Perspective*. Dordrecht: Springer.
Jacobs, Jon. 2009. ‘Naturalism’. *Internet Encyclopedia of Philosophy*. http://www.iep.utm.edu/naturali/.
Katz, Leonard D. 1995. ‘On Distinguishing Phenomenal Consciousness from the Representational Functions of Mind’. *Behavioral and Brain Sciences* 18 (2): 258–59.
Keefe, Rosanna. 2000. *Theories of Vagueness*. Cambridge: Cambridge University Press.
Keefe, Rosanna, and Peter Smith (eds.). 1996. *Vagueness: A Reader*. Cambridge, MA: MIT Press.
Khalidi, Muhammad Ali. 2013. *Natural Categories and Human Kinds: Classification in the Natural and Social Sciences*. Cambridge: Cambridge University Press.

Kim, Jaegwon. 1992. ‘Multiple Realization and the Metaphysics of Reduction’. *Philosophy and Phenomenological Research* 52 (1): 1–26.
Kirk, Robert. 1996. *Raw Feeling: A Philosophical Account of the Essence of Consciousness*. Oxford: Oxford University Press.
———. 2015. ‘Zombies’. In *The Stanford Encyclopedia of Philosophy* (Summer 2015 Edition), edited by Edward N. Zalta. http://plato.stanford.edu/archives/sum2015/entries/zombies/.
Kitcher, Philip. 1978. ‘Theories, Theorists and Theoretical Change’. *The Philosophical Review* 87 (4): 519–47.
———. 1984. ‘Species’. *Philosophy of Science* 51 (2): 308–33.
———. 1987. ‘Ghostly Whispers: Mayr, Ghiselin, and the “Philosophers” on the Ontological Status of Species’. *Biology and Philosophy* 2 (2): 184.
———. 1989. ‘Some Puzzles about Species’. In *What the Philosophy of Biology Is: Essays Dedicated to David Hull*, edited by Michael Ruse. Dordrecht: Kluwer Academic Publishers.
———. 1992. ‘The Naturalists Return’. *Philosophical Review* 101 (1): 53–114.
Knobe, Joshua. 2008. ‘Can a Robot, an Insect or God Be Aware?’ *Scientific American*, June 24.
Koch, Christof, and Naotsugu Tsuchiya. 2007. ‘Attention and Consciousness: Two Distinct Brain Processes’. *Trends in Cognitive Sciences* 11 (1): 16–22.
———. 2012. ‘Attention and Consciousness: Related yet Different’. *Trends in Cognitive Sciences* 16 (2): 103–5.
Kouider, Sid, Stanislas Dehaene, Antoinette Jobert, and Denis Le Bihan. 2007. ‘Cerebral Bases of Subliminal and supraliminal Priming During Reading’. *Cerebral Cortex* 17: 2019-29.
Kornblith, Hilary. 1993. *Inductive Inference and Its Natural Ground: An Essay in Naturalistic Epistemology*. Cambridge, MA: MIT Press.
Koslicki, Kathrin. 2008. ‘Natural Kinds and Natural Kind Terms’. *Philosophy Compass* 3 (4): 789–802.
Kozuch, Benjamin. 2014. ‘Review of Elizabeth Irvine’s *Consciousness as a Scientific Concept: A Philosophy of Science Perspective*’. *The British Journal for the Philosophy of Science* 65 (3): 651–55.
Kripke, Saul A. 1980. *Naming and Necessity*. Cambridge, MA: Harvard University Press.
———. 2013. *Reference and Existence: The John Locke Lectures*. Oxford: Oxford University Press.
Kukla, Andre. 1994. ‘Non-Empirical Theoretical Virtues and the Argument From Underdetermination’. *Erkenntnis* 41 (2): 157–70.
Lagercrantz, Hugo. 2014. ‘The Emergence of Consciousness: Science and Ethics’. *Seminars in Fetal and Neonatal Medicine*. Accessed September 4.
Lähteenmäki, Vili. 2007. ‘Orders of Consciousness and Forms of Reflexivity in Descartes’. In *Consciousness: From Perception to Reflection in the History of Philosophy*, edited by Sara Heinämaa, Vili Lähteenmäki, and Pauliina Remes, 177–202. Dordrecht: Springer.
Lamme, Victor A.F. 2004. ‘Separate Neural Definitions of Visual Consciousness and Visual Attention; a Case for Phenomenal Awareness’. *Neural Networks* 17 (5–6): 861–72.

———. 2006. ‘Towards a True Neural Stance on Consciousness’. *Trends in Cognitive Sciences* 10 (11): 494–501.
———. 2010. ‘How Neuroscience Will Change Our View on Consciousness’. *Cognitive Neuroscience* 1 (3): 204–20.
LaPorte, Joseph. 2004. *Natural Kinds and Conceptual Change*. Cambridge: Cambridge University Press.
———. 2010. ‘Theoretical Identity Statements, Their Truth, and Their Discovery’. In *The Semantics and Metaphysics of Natural Kinds*, edited by Helen Beebee and Nigel Sabbarton-Leary, 104–24. London: Routledge
Laureys, Stephen, and Melanie Boly. 2007. ‘What Is It like to Be Vegetative or Minimally Conscious?’ *Current Opinion in Neurology* 20 (6): 609–13.
———. 2009. ‘Brain Damage’. In *The Oxford Companion to Consciousness*, edited by Tim Bayne, Axel Cleeremans, and Patrick Wilken, 121–26. Oxford: Oxford University Press.
Lee, S.J., H.J.P. Ralston, E.A. Drey, J.C. Partridge, and M.A. Rosen. 2005. ‘Fetal Pain: A Systematic Multidisciplinary Review of the Evidence’. *JAMA: The Journal of the American Medical Association* 294 (8): 947.
Lewis, David. 1969. ‘Review of W. H. Capitan and D. D. Merrill (eds.). *Art, Mind and Religion*’. *The Journal of Philosophy* 66 (1): 22–27.
———. 1972. ‘Psychophysical and Theoretical Identifications’. *Australasian Journal of Philosophy* 50: 249–58.
———. 1983. ‘New Work for a Theory of Universals’. *Australasian Journal of Philosophy* 61: 343–77.
———. 1984. ‘Putnam’s Paradox’. *Australasian Journal of Philosophy* 62 (3): 221–36.
———. 1986. *On the Plurality of Worlds*. Oxford: Blackwell.
———. 1993. ‘Many, but Almost One’. In *Ontology, Causality, and Mind: Essays on the Philosophy of D. M. Armstrong*, edited by John Bacon, Keith Campbell, and Lloyd Reinhardt, 23–38. Cambridge: Cambridge University Press.
———. 1994. ‘David Lewis: Reduction of Mind’. In *A Companion to the Philosophy of Mind*, edited by Samuel Guttenplan, 412–31. Oxford: Blackwell Publishers.
Lipton, Peter. 2004. *Inference to the Best Explanation* (Second Edition). Oxford: Routledge.
Lloyd-Thomas, Adrian R., and Maria Fitzgerald. 1996. ‘Do Fetuses Feel Pain? Reflex Responses Do Not Necessarily Signify Pain’. *BMJ (Clinical Research Ed.)* 313 (7060): 797–98.
Loar, Brian. 1997. ‘Phenomenal States’. In *The Nature of Consciousness: Philosophical Debates*, edited by Ned Block, Owen Flanagan, and Güven Güzeldere, 597–616. Cambridge, MA: MIT Press.
Longino, Helen. 2002. *The Fate of Knowledge*. New Jersey: Princeton University Press.
———. 2004. ‘How Values Can Be Good for Science’. In *Science, Values and Objectivity*, edited by Peter Machamer and Gereon Wolters, 127–42. Pittsburgh: University of Pittsburgh Press.
———. 2015. ‘The Social Dimensions of Scientific Knowledge’. In *The Stanford Encyclopedia of Philosophy* (Spring 2015 Edition), edited by Edward N. Zalta. http://plato.stanford.edu/archives/spr2015/entries/scientific-knowledge-social/.

Lowery, C.L., M.P. Hardman, N. Manning, R.W. Hall, and K.J.S. Anand. 2007. 'Neurodevelopmental Changes of Fetal Pain'. *Seminars in Perinatology*, 31:275–82.
Lurz, Robert. 2009. 'Animal Minds'. *The Internet Encycloypedia of Philosophy*. http://www.iep.utm.edu/ani-mind/.
Lycan, William G. 2015. 'Representational Theories of Consciousness'. In *The Stanford Encyclopedia of Philosophy* (Summer 2015 Edition), edited by Edward N. Zalta. http://plato.stanford.edu/archives/sum2015/entries/consciousness-representational/.
———. 1995. 'We've Only Just Begun'. *Behavioral and Brain Sciences* 18 (2): 262.
———. 2004. 'The Plurality of Consciousness'. In *Language, Knowledge, and Representation: Proceedings of the Sixth International Colloquium on Cognitive Science*, edited by Jésus M. Larrazabal and Luis A. Pérez Miranda, 93–102. Dordrecht: Springer Science.
Machamer, Peter, and Lisa Osbeck. 2004. 'The Social in the Epistemic'. In *Science, Values and Objectivity*, edited by Peter Machamer and Gereon Wolters, 78–89. Pittsburgh: University of Pittsburgh Press.
Machery, Edouard. 2009. *Doing without Concepts*. Oxford: Oxford University Press.
Mack, Arien, and Irvin Rock. 1998. *Inattentional Blindness*. Cambridge, MA: MIT Press.
MacLeod, Miles, and Thomas A. C. Reydon. 2013. 'Natural Kinds in Philosophy and in the Life Sciences: Scholastic Twilight or New Dawn?' *Biological Theory* 7 (2): 89–99.
Magnus, P.D. 2012. *Scientific Enquiry and Natural Kinds: From Planets to Mallards*. New York: Palgrave Macmillan.
———. 2014. 'No Grist for Mill on Natural Kinds'. *Journal for the History of Analytical Philosophy* 2 (4).
Mallet, James. 2007. 'Concepts of Species'. In *Encyclopedia of Biodiversity*, edited by Simon Asher Levin, 1–15. New York: Elsevier.
Manson, Neil. 2007. 'Contemporary Naturalism and the Concept of Consciousness'. In *Consciousness: From Perception to Reflection in the History of Philosophy*, edited by Sara Heinämaa, Vili Lähteenmäki, and Pauliina Remes, 287–310. Dordrecht: Springer.
Marr, David. 1982. *Vision: A Computational Investigation Into the Human Representation and Processing of Visual Information*. New York: W.H. Freeman.
Martí, Genoveva, and José Martínez-Fernández. 2010. 'General Terms as Designators : A Defence of the View'. In *The Semantics and Metaphysics of Natural Kinds*, edited by Helen Beebee and Nigel Sabbarton-Leary, 46–63. Oxford: Routledge.
Mason, Georgia J. 2011. 'Invertebrate Welfare: Where Is the Real Evidence for Conscious Affective States?' *Trends in Ecology & Evolution* 26 (5): 212–13.
Massimi, Michela. 2012. 'Dwatery Ocean'. *Philosophy* 87 (04): 531–55.
Mather, Jennifer A. 2008. 'Cephalopod Consciousness: Behavioural Evidence'. *Consciousness and Cognition* 17 (1): 37–48.

Mayden, R. L. 1997. ‘A Hierarchy of Species Concepts: The Denouement in the Saga of the Species Problem’. In *Species: The Units of Diversity,* edited by M.F. Claridge, H.A. Dawah, and M.R. Wilson, 381–423. London: Chapman & Hall.
Mayr, Ernst. 1969. *Principles of Systematic Zoology*. New York: McGraw-Hill.
McCullagh, P. 1997. ‘Do Fetuses Feel Pain? Can Fetal Suffering Be Excluded beyond Reasonable Doubt?’ *BMJ (Clinical Research Ed.)* 314 (7076): 302–3.
McGinn, Colin. 1989. ‘Can We Solve the Mind-Body Problem?’ *Mind* 98 (391): 349–66.
———. 1991. *The Problem of Consciousness: Essays Towards a Resolution*. Oxford: Blackwell.
———. 1996. *The Character of Mind: An Introduction to the Philosophy of Mind* (Second Edition). New York: Oxford University Press.
———. 1999. *The Mysterious Flame: Conscious Minds in a Material World*. New York: Basic Books.
McLaughlin, Brian P. 2003. ‘A Naturalist-Phenomenal Realist Response to Block’s Harder Problem’. *Philosophical Issues* 13 (1): 163–204.
McMullin, Ernan. 2008. ‘The Virtues of a Good Theory’. In *The Routledge Companion to Philosophy of Science*, edited by Stathis Psillos and Martin Curd, 498–508. Oxford: Routledge.
Mellor, D.J., T.J. Diesch, A.J. Gunn, and L. Bennet. 2005. ‘The Importance of “Awareness” for Understanding Fetal Pain’. *Brain Research Reviews* 49 (3): 455–71.
Merker, Bjorn. 2007. ‘Consciousness without a Cerebral Cortex: A Challenge for Neuroscience and Medicine’. *Behavioral and Brain Sciences* 30 (1): 63–81.
Mill, John Stuart. 1874. *A System of Logic, Ratiocinative and Inductive: Being a Connected View of the Principles of Evidence and the Methods of Scientific Inverstigation* (8th Edition). London: Harper & Brothers.
Nagel, Thomas. 1974. ‘What Is It like to Be a Bat?’ *The Philosophical Review* 83 (4): 435–50. All references are to the version in Block, Flanagan, and Güzeldere eds. (1997), 519–28.
———. 1986. *The View from Nowhere*. New York: Oxford University Press.
Norton, John D. 2008. ‘Must Evidence Underdetermine Theory?’ In *The Challenge of the Social and the Pressure of Practice: Science and Values Revisited*, edited by Martin Carrier, Don Howard, and Janet Kourany, 17–44. Pittsburgh: University of Pittsburgh Press.
Okasha, Samir. 2002. ‘Darwinian Metaphysics: Species and the Question of Essentialism’. *Synthese* 131 (2): 191–213.
O’Malley, Maureen A., and John Dupré. 2007. ‘Size Doesn’t Matter: Towards a More Inclusive Philosophy of Biology’. *Biology and Philosophy* 22 (2): 155–91.
Panksepp, Jaak. 2005. ‘Affective Consciousness: Core Emotional Feelings in Animals and Humans’. *Consciousness and Cognition* 14 (1): 30–80.
Papineau, David. 1993. *Philosophical Naturalism*. London: Blackwell Publishing.
———. 1996. ‘Reply to Commentators’. *Philosophy and Phenomenological Research* 56 (3): 687–97.
———. 2002. *Thinking about Consciousness*. Oxford: Clarendon Press.
———. 2003. ‘Could There Be A Science of Consciousness?’ *Philosophical Issues* 13 (1): 205–20.

———. 2006. ‘Phenomenal and Perceptual Concepts’. In *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism*, edited by Torin A. Alter and Sven Walter, 111–44. Oxford: Oxford University Press.
———. 2007. ‘Reuniting (Scene) Phenomenology with (Scene) Access’. *Behavioral and Brain Sciences* 30 (5-6): 521–521.
———. 2015. ‘Naturalism’. In *The Stanford Encyclopedia of Philosophy* (Fall 2015 Edition), edited by Edward N. Zalta. http://plato.stanford.edu/archives/fall2015/entries/naturalism/.
Pashler, Harold E. 1998. *The Psychology of Attention*. Cambridge, MA: MIT Press.
Peacocke, Christopher. 1984. ‘Consciousness and Other Minds I’. *Proceedings of the Aristotelian Society,* Supplementary Volumes 58 (January): 97–117.
Pedroso, Makmiller. 2012. ‘Essentialism, History, and Biological Taxa’. *Studies in History and Philosophy of Science Part C* 43 (1): 182–90.
Piccinini, Gualtiero. 2007. ‘The Ontology of Creature Consciousness: A Challenge for Philosophy’. *Behavioral and Brain Sciences* 30 (1): 103–4.
Polger, Thomas W. 2006. *Natural Minds*. Cambridge, MA: MIT Press.
Polger, Thomas W., and Lawrence Shapiro. 2008. ‘Understanding the Dimensions of Realization’. *The Journal of Philosophy* 105 (4): 213–22.
Prinz, Jesse. 1998. ‘Vagueness, Language, and Ontology’. *The Electronic Journal of Analytic Philosophy* 6.
———. 2003. ‘Level-Headed Mysterianism and Artificial Experience’. *Journal of Consciousness Studies* 10 (4-5): 111–32. In *Machine Consciousness,* edited by Owen Holland, 111-32. Exeter: Imprint Academic.
———. 2005. ‘A Neurofunctional Theory of Consciousness’. In *Cognition and the Brain: The Philosophy and Neuroscience Movement*, edited by Andrew Brook and Kathleen Akins, 381–96. Cambridge: Cambridge University Press.
———. 2012. *The Conscious Brain: How Attention Engenders Experience*. New York: Oxford University Press.
Putnam, Hilary. 1964. ‘Robots: Machines or Artificially Created Life?’ *The Journal of Philosophy* 61: 668–91. All references are to the version in Putnam (1975), 386–407.
———. 1967. ‘The Nature of Mental States’. Originally ‘Psychological predicates’. In *Art, Mind, and Religion*, edited by W.H. Capitan and D.D. Merrill. Pittsburgh: University of Pittsburgh Press. All references are to the version in Putnam (1975), 429–40.
———. 1970. ‘Is Semantics Possible’. In *Languages, Belief and Metaphysics*, edited by H. Kiefer and M.K. Munitz. New York: SUNY Press. All references are to the version in Putnam (1975), 139–52.
———. 1973. ‘Explanation and Reference’. In *Conceptual Change,* edited by G. Pearce and P. Maynard. Dordrecht-Reidel. All references are to the version in Putnam (1975), 196–214.
———. 1975. *Mind, Language and Reality: Philosophical Papers, Volume 2*. Cambridge: Cambridge University Press.
———. 1975a. ‘The Meaning of “Meaning”’. In *Language, Mind, and Knowledge: Minnesota Studies in the Philosophy of Science VII*, edited by K. Gunderson.

Minnesota: University of Minnesota Press. All references are to the version in Putnam (1975), 215–71.
———. 1988. *Representation and Reality*. MIT Press.
———. 1990. 'Is Water Necessarily H_2O?' In *Realism with a Human Face*, 54–79. Cambridge, MA: Harvard University Press.
Quine, W.V.O. 1969. 'Natural Kinds'. In *Ontological Relativity and Other Essays*, 114–38. New York: Columbia University Press.
Raffman, Diana. 1996. 'Vagueness and Context-Relativity'. *Philosophical Studies* 81 (2-3): 175–92.
Reissland, Nadja, Brian Francis, and James Mason. 2013. 'Can Healthy Fetuses Show Facial Expressions of "Pain" or "Distress"?' *PLoS ONE* 8 (6): e65530.
Rescher, Nicholas. 2001. *Paradoxes: Their Roots, Range, and Resolution*. Chicago, IL: Open Court.
Richards, Richard A. 2010. *The Species Problem: A Philosophical Analysis*. Cambridge: Cambridge University Press.
Robinson, Howard. 2012. 'Dualism'. In *The Stanford Encyclopedia of Philosophy* (Winter 2012 Edition), edited by Edward N. Zalta. http://plato.stanford.edu/archives/win2012/entries/dualism/.
Rose, J.D. 2002. 'The Neurobehavioural Nature of Fishes and the Question of Awareness and Pain'. *Reviews in Fisheries Science* 10 (1): 1–38.
———. 2007. 'Anthropomorphism and "Mental Welfare" of Fishes'. *Diseases of Aquatic Organisms* 75 (2): 139–54.
Rose, J.D., R. Arlinghaus, S.J. Cooke, B.K. Diggles, W. Sawynok, E.D. Stevens, and C.D.L. Wynne. 2014. 'Can Fish Really Feel Pain?' *Fish and Fisheries* 15 (1): 97–133.
Salmon, Nathan. 2003. 'Naming, Necessity, and Beyond'. *Mind* 112 (447): 475–92.
Sanford, David H. 2014. 'Determinates vs. Determinables'. In *The Stanford Encyclopedia of Philosophy* (Summer 2014 Edition), edited by Edward N. Zalta. http://plato.stanford.edu/archives/sum2014/entries/determinate-determinables/.
Schwartz, Stephen P. 2002. 'Kinds, General Terms, and Rigidity: A Reply to LaPorte'. *Philosophical Studies* 109 (3): 265–77.
Searle, John R. 1992. *The Rediscovery of the Mind*. Cambridge, MA: MIT Press.
———. 1997. *The Mystery of Consciousness*. New York: The New York Review of Books.
Seth, Anil K., Bernard J. Baars, and David B. Edelman. 2005. 'Criteria for Consciousness in Humans and Other Mammals'. *Consciousness and Cognition* 14 (1): 119–39.
Shapin, Steven, and Simon Schaffer. 1986. *Leviathan and the Air Pump: Hobbes, Boyle, and the Experimental Life*. New Jersey: Princeton University Press.
Shapiro, Lawrence. 2000. 'Multiple Realizations'. *The Journal of Philosophy* 97 (12): 635–54.
Shapiro, Lawrence, and Thomas W. Polger. 2012. 'Identity, Variability, and Multiple Realization in the Special Sciences'. In *New Perspectives on Type Identity: The Mental and the Physical*, edited by Simone Gozzano and Christopher S. Hill, 264–86. Cambridge: Cambridge University Press.
Shapiro, Stewart. 2006. *Vagueness in Context*. New York: Oxford University Press.

Shea, Nicholas. 2012. ‘Methodological Encounters with the Phenomenal Kind’. *Philosophy and Phenomenological Research* 84 (2): 307–44.
———. 2014. ‘Using Phenomenal Concepts to Explain Away the Intuition of Contingency’. *Philosophical Psychology* 27 (4): 553-70.
Shea, Nicholas, and Tim Bayne. 2010. ‘The Vegetative State and the Science of Consciousness’. *British Journal for the Philosophy of Science* 61 (3): 459–84.
Slater, Matthew H. 2005. ‘Monism on the One Hand, Pluralism on the Other’. *Philosophy of Science* 72 (1): 22–42.
———. 2015. ‘Natural Kindness’. *British Journal for the Philosophy of Science* 66 (2): 375–411.
Smart, J. J. C. 1959. ‘Sensations and Brain Processes’. *The Philosophical Review* 68 (2): 141–56.
———. 2007. ‘The Mind/Brain Identity Theory’. In *The Stanford Encyclopedia of Philosophy* (Winter 2012 Edition), edited by Edward N. Zalta. http://plato.stanford.edu/archives/win2012/entries/mind-identity/.
Smith, R. P., R. Gitau, V. Glover, and N. Fisk. 2000. ‘Pain and Stress in the Human Fetus’. *European Journal of Obstetrics and Gynecology* 92 (1): 161–65.
Sneddon, L.U. 2009. ‘Pain Perception in Fish: Indicators and Endpoints’. *ILAR Journal* 50 (4): 338–42.
Soames, Scott. 1999. *Understanding Truth*. New York: Oxford University Press.
Sober, Elliott. 1988. *Reconstructing the Past: Parsimony, Evolution, and Inference*. Cambridge, MA: MIT Press.
———. 1999. ‘The Multiple Realizability Argument Against Reductionism’. *Philosophy of Science* 66 (4): 542–64.
Sorensen, Roy. 2005. ‘Précis of “Vagueness and Contradiction”’. *Philosophy and Phenomenological Research* 71 (3): 678–85.
———. 2012. ‘Vagueness’. In *The Stanford Encyclopedia of Philosophy* (Summer 2012 Edition), edited by Edward N. Zalta. http://plato.stanford.edu/archives/sum2012/entries/vagueness/.
Stanford, P. Kyle. 1995. ‘For Pluralism and against Realism about Species’. *Philosophy of Science* 62: 70–79.
———. 2013. ‘Underdetermination of Scientific Theory’. In *The Stanford Encyclopedia of Philosophy* (Winter 2013 Edition), edited by Edward N. Zalta. http://plato.stanford.edu/archives/win2013/entries/scientific-underdetermination/.
Stanford, P. Kyle, and Philip Kitcher. 2000. ‘Refining the Causal Theory of Reference for Natural Kind Terms’. *Philosophical Studies* 97 (1): 97–127.
Stich, Stephen P. 1998. *Deconstructing the Mind*. New York: Oxford University Press.
Stoljar, Daniel. 2015. ‘Physicalism’. In *The Stanford Encyclopedia of Philosophy* (Spring 2015 Edition), edited by Edward N. Zalta. http://plato.stanford.edu/archives/spr2015/entries/physicalism/.
Strawson, Galen. 2009. *Mental Reality* (Second Edition). Cambridge, MA: MIT Press.
Sturgeon, Scott. 1994. ‘The Epistemic View of Subjectivity’. *The Journal of Philosophy* 91 (5): 221–35.
Sytsma, Justin. 2010. ‘Folk Psychology and Phenomenal Consciousness’. *Philosophy Compass* 5 (8): 700–711.

Templeton, A., R. Anderson, T. Belfield, S. Derbyshire, K. Ellis, and J. Fisher. 2010. *Fetal Awareness: Review of Research and Recommendations for Practice. Report of a Working Party*. London: Royal College of Obstetricians and Gynaecologists.

Tononi, Giulio. 2004. 'An Information Integration Theory of Consciousness'. *BMC Neuroscience* 5 (1): 42.

———. 2008. 'Consciousness as Integrated Information: A Provisional Manifesto'. *The Biological Bulletin* 215 (3): 216–42.

Turner, Derek. 2007. *Making Prehistory: Historical Science and the Scientific Realism Debate*. Cambridge: Cambridge University Press.

Tye, Michael. 1996. 'Is Consciousness Vague or Arbitrary?' *Philosophy and Phenomenological Research* 56 (3): 679–85.

———. 1997. 'The Problem of Simple Minds: Is There Anything It Is like to Be a Honey Bee?' *Philosophical Studies* 88 (3): 289–317. All references are to the version in Tye (2000), 171-85.

———. 2000. *Consciousness, Color, and Content*. Cambridge, MA: MIT Press.

———. 2009. *Consciousness Revisited: Materialism Without Phenomenal Concepts*. Cambridge, MA: MIT Press.

———. 2015. 'Qualia'. In *The Stanford Encyclopedia of Philosophy* (Fall 2015 Edition), edited by Edward N. Zalta. ttp://plato.stanford.edu/archives/fall2015/entries/qualia/.

Unger, Peter. 1988. 'Conscious Beings in a Gradual World'. *Midwest Studies In Philosophy* 12 (1): 287–333.

Van Gulick, Robert. 1997a. 'Understanding the Phenomenal Mind: Are We All Just Armadillos? Part II: The Absent Qualia Argument'. In *The Nature of Consciousness: Philosophical Debates*, edited by Ned Block, Owen Flanagan, and Güven Güzeldere, 435–42. Cambridge, MA: MIT Press.

———. 1997b. 'Understanding the Phenomenal Mind: Are We All Just Armadillos? Part I: Phenomenal Knowledge and Explanatory Gaps'. In *The Nature of Consciousness: Philosophical Debates*, edited by Ned Block, Owen Flanagan, and Güven Güzeldere, 559–66. Cambridge, MA: MIT Press.

———. 2014. 'Consciousness'. *The Stanford Encyclopedia of Philosophy* (Spring 2014 Edition), edited by Edward N. Zalta. http://plato.stanford.edu/archives/spr2014/entries/consciousness/.

Van Scheltema, P., S. Bakker, F. Vandenbussche, and D. Oepkes. 2008. 'Fetal Pain'. *Fetal and Maternal Medicine Review* 19 (04): 311–24.

Watt, Douglas F. 2005. 'Panksepp's Common Sense View of Affective Neuroscience Is Not the Commonsense View in Large Areas of Neuroscience'. *Consciousness and Cognition* 14 (1): 81–88.

Weiner, Joan. 2007. 'Science and Semantics: The Case of Vagueness and Supervaluation'. *Pacific Philosophical Quarterly* 88 (3): 355–74.

Wheeler, Q.D., and R. Meier (eds.). 2000. *Species Concepts and Phylogenetic Theory*. New York: Columbia University Press.

Wilkes, Kathleen V. 1984. 'Is Consciousness Important?' *British Journal for the Philosophy of Science* 35: 223–43.

Wilkes, Kathleen V. 1988. '*Yishi, Duh, Um* and Consciousness'. In *Consciousness in Contemporary Science*, edited by A. Marcel and E. Bisiach. Oxford: Oxford University Press.
Williams, J. Robert G. 2008. 'Ontic Vagueness and Metaphysical Indeterminacy'. *Philosophy Compass* 3 (4): 763–88.
Williams, J. Robert G., and Elizabeth Barnes. 2011. 'A Theory of Metaphysical Indeterminacy'. In *Oxford Studies in Metaphysics Volume 6*. Oxford: Oxford University Press.
Williamson, Timothy. 1994. *Vagueness*. Oxford: Routledge.
———. 2005. 'Vagueness in Reality'. In *The Oxford Handbook of Metaphysics*, edited by Michael J. Loux and Dean W. Zimmerman, 690–715. New York: Oxford University Press.
Wilson, Robert A. 2004. 'Review of LaPorte on Natural Kinds'. *Philosophy in Review* 24: 423–26.
Wimsatt, William C. 1976. 'Reductionism, Levels of Organization, and the Mind-Body Problem'. In *Consciousness and the Brain: A Scientific and Philosophical Inquiry*, edited by Gordon G. Globus, Grover Maxwell, and Irwin Savodnik, 205–67. New York: Plenum Press.
Wisdom, J., J.L. Austin, and A.J. Ayer. 1946. 'Symposium: Other Minds'. *Proceedings of the Aristotelian Society, Supplementary Volumes* 20: 122–97.
Woolgar, Steve. 1988. *Science: The Very Idea*. Oxford: Routledge.
Wright, Crispin. 1995. 'The Epistemic Conception of Vagueness'. *Southern Journal of Philosophy* 33 (S1): 133-160.
———. 2001. 'On Being in a Quandary: Relativism, Vagueness, Logical Revisionism'. *Mind* 110 (437): 45–97.
Wynne, C.D.L. 2005. 'Animal Consciousness: How Can We Know?' *Trends in Cognitive Sciences* 9 (12): 562–63.