Emergence and Experience: Systemic Emergence and the Prospects for a Mechanistic Explanation of the Existence of Experience

Andy McKilliam

Department of Philosophy, School of Humanities

The University of Adelaide

A thesis submitted in fulfilment of the requirements for the degree of Master of Philosophy

March 2017

Abstract

The dominant view among philosophers and scientists today is that the world, and everything in it, is constructed from a relatively small set of fundamental entities: roughly those picked out by physics. This view is known as materialism. Materialism is not the view that only these fundamentals exist. The world contains many wondrous things that are not themselves fundamental physical entities: things such as flowers, organisms, families, feelings of joy. Rather, materialism, as I shall defend it, is the view that only the fundamental microphysical entities are instantiated in a *basic* way. Everything else *emerges,* in a non-mysterious fashion, as a result of intricately organized collections of more basic entities.

Materialism has a lot going for it but it also faces a number of major challenges. One of those challenges is to account for conscious experience. Consciousness is an undeniable feature of the world. And yet, we currently have no idea as to how something like a subjective conscious experience could be a non-mysteriously emergent feature of material systems. So puzzled are we on this front that a number of philosophers think that ultimately materialism cannot be correct. They think that somewhere along the way, consciousness must be taken as a fundamental (or basic) feature of the world.

As it stands we have two intuitively appealing, yet hard to reconcile theses:

1. *Materialism*: Only the fundamental entities described by physics are instantiated in a basic way. Everything else emerges with organized collections of these fundamentals.
2. *Conscious Realism*: Conscious experiences are real, causally potent, and in need of explanation.

This thesis will work towards their reconciliation.

I develop and defend a conception of emergence—emergence as systemic novelty—that is in keeping with discussions in systems biology and the other sciences of the mind. This picture allows us to understand how causally potent systems can emerge without jeopardizing the core tenets of materialism. Consciousness still poses a serious problem for this view as there are a number of intuitively powerful reasons to think that, unlike other systemically emergent phenomena, conscious experience cannot be accounted for in terms of the organized

interactions of the system's constituents. A number of thinkers have argued that this entails the falsity of the materialism in all its forms. In addressing this concern, I argue that there are in fact two problems associated with consciousness: there is the problem of accounting for the existence of *experience in general*, and there is the problem of accounting for the *qualitative character* of experience. While the second of these problems may indeed be intractable, there is reason to be optimistic about the prospects of solving the first. If *experience in general* is not itself something we experience, then there may be space for a conceptual renovation that allows for an illuminating explanation of the existence of experience. Further, I argue that a solution to the problem of accounting for the existence of *experience in general* is all that is needed to vindicate materialism.

Table of Contents

Abstract i
Declaration iv
Acknowledgements v
Introduction 1
Chapter 1: Materialism 7
1.1 What Materialism is Not 8
1.2 An Initial Concern: Hempel's Dilemma 13
1.3 Why Believe Materialism Is True? 14
1.4 A Note on Terminology: Materialism or Physicalism? 18
1.5 Levels 21
1.6 Two Failed Attempts at Relating Levels of Mechanism 25
Chapter 2: Emergence 30
2.1 Some Foundations 30
2.2 Naïve Secretion Emergence 37
2.3 Emergence as Systemic Novelty 40
2.4 Emergent Causation 44
2.5 Summary 48
Chapter 3: The Problem of Consciousness 50
3.1 The Hard Problem of Consciousness 51
3.2 Do Developments in Philosophy of Science Help? 58
3.3 Summary 65
Chapter 4: Arguments Against Materialism and the Outlines of a Novel Response 67
4.1 Arguments against materialism 68
4.2 The Two Problems of Consciousness 74
4.3 Why Materialists Need not Fear Qualia 82
4.4 Can we Reductively Explain the Existence of Experience? 90
4.5 Summary 95
Reference List 97

Declaration

I certify that this work contains no material which has been accepted for the award of any other degree or diploma in my name, in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. In addition, I certify that no part of this work will, in the future, be used in a submission in my name, for any other degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint-award of this degree.

I give consent to this copy of my thesis, when deposited in the University Library, being made available for loan and photocopying, subject to the provisions of the Copyright Act 1968.

I also give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library Search and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

I acknowledge the support I have received for my research through the provision of an Australian Government Research Training Program Scholarship.

Andy Mckilliam

Signature: Date:

Acknowledgements

I would like to thank the following people whose help has been invaluable during the process of completing this thesis: Dr. Jon Opie especially, and also Dr. Gerard O'Brien for expert guidance and helpful comments on early versions of this thesis; the University of Adelaide Philosophy Department for helpful feedback on early versions of this material; Ed Heddle for the unquenchable supply of Jazz piano and for being a first-rate sounding-board; my Dad for raising me on questions rather than answers; and finally, my Mum for her unwavering love and support.

Introduction

In the house in which I currently live there is a particular spot on a particular couch by the window where I like to do my thinking, reading, and much of my writing. One of the boons of this perch is that the path between it and the fridge is direct and generally obstacle free; making the procurement of refreshments, even while reading, a relatively unchallenging and typically uneventful task. However, on one particular trip to the fridge early in my candidature, my journey was interrupted by a stray dinning chair, not in its usual location tucked safely under the dining table off to the left. On this occasion, the second to last toe of my left foot had a rather intimate and wholly undesirable encounter with the rear left leg of the offending chair. A number of events followed.

Specialised cells under the skin and in the joints of my toe (nociceptors) sent a burst of signals to my brain where a whir of neural activity brought about a host of behavioral responses: signals from my motor cortex lead to a particular sequence of muscle contractions and relaxations, causing me to shift my weight to the right as I jumped up, clutched my left foot in my hands and inspected it for damage; activity in Brocca's area—an area of the brain known to be involved in speech production—was part of the causal process that lead to certain vocalizations unfit to be repeated in scholarly work such as this aspires to be; then, after realizing that no serious damage had been incurred, I returned the chair to its proper location and continued on my way to the fridge.

All of this was *objectively observable*. Anyone (or anything) possessing the appropriate sensory apparatus (and measuring instruments), would have been able to witness each of these phenomena take place. But, as anyone who has stubbed their toe will attest, these objectively observable events and processes do not exhaust the phenomena that follow an instance of toe-stubbing. Stubbing your toe, and the neural processes that follow, does not merely cause you to behave in certain ways, it also hurts. A lot! In addition to the *objectively observable* phenomena just mentioned, the processes that took place in my nervous system following the stubbing of my toe also, somehow, induced in me a *subjective conscious experience* of pain. Moreover, it is natural to think that this *pain* had a causal role to play in those processes. Intuitively, the reason I grabbed my foot and inspected it for damage is that it hurt. Presumably, if it hadn't hurt, I would

have simply returned the chair to its proper location and continued on my way to the fridge.

The first of the two assumptions that form the foundations of this thesis is that conscious experiences, such as the pain that I felt after stubbing my toe, are real and causally efficacious features of the world.

The second of these foundational assumptions is the combination of two claims: (1) the *natural world* is all that exists—there are no ghosts, souls, gods, supernatural powers, or anything of the sort; and (2) the natural world is ultimately constructed from a relatively small set of fundamental building blocks: roughly, those picked out by physics. This view is known as materialism. Materialism is not the view that only the microphysical fundamentals exist. The world contains many wondrous things that are not themselves fundamental microphysical entities—things such as flowers, organisms, families, feelings of joy and of pain. Rather, as I shall defend it, materialism is the view that only the fundamental microphysical entities are instantiated in a *basic* way, everything else *arises* (or emerges) in a non-mysterious fashion as a result of intricately organized systems composed of these fundamentals.

Materialism is by far the dominant view among philosophers and scientists today, but it faces a major challenge: how to account for conscious experience as a real and causally potent feature of the world. As it stands, we currently have no clear and illuminating idea how something like a subjective conscious experience could arise in a non-mysterious fashion as a result of intricately organized collections of material entities. To borrow a phrase from Sellars, we currently have no idea how subjective conscious experiences, such as the *pain* I felt after stubbing my toe, "hang together" with the kinds of processes that took place is my nervous system at that time (1963). So puzzled are we on this front that a number of philosophers take it to show that ultimately materialism must be false. These thinkers argue that somewhere along the line (either at the most basic level of microphysics, or at the level of complex 'information processing' systems), consciousness must be taken as a *fundamental* (or basic) feature of the world.

So, we have two attractive yet hard to reconcile theses:

1. *Materialism*: Only the fundamental entities described by physics are instantiated in a basic way.

2. *Conscious Realism*: Conscious experiences are real and causally potent features of the world.

This thesis will work towards reconciling these two claims.

Over the past decade or so, it has once again become fashionable to employ the term *emergence* to describe how conscious experiences (among other things) hang together with material systems. The basic idea of emergence is that when entities become organized in complex ways, collectively they can constitute a whole that can do entirely new kinds of things, and instantiate entirely new kinds of properties: properties unlike those had by the parts alone. To take a familiar example, consider the property of 'liquidity'. Intuitively, something is liquid if it flows in a particular way and has a relatively constant volume (i.e., it does not dissipate like a gas, but nor does it have a fixed shape like a solid). While no individual H_2O molecule is liquid—liquidity is not the kind of property that can be instantiated by individual molecules of any kind—if you happen to find a large number of H_2O molecules in close proximity, under the right conditions they can collectively constitute a body of liquid water. 'Liquidity' is an emergent property.

Despite its current trendiness, many philosophers are suspicious of emergence.[1] The reason emergence is met with such squinty-eyed suspicion by philosophers is that the term 'emergence' is used to refer to such a wide range of ideas (some quite metaphysically extravagant) that it is hard to know exactly what it is supposed to mean. More worrying still, in some of its uses, the term 'emergence' appears to be nothing more than a placeholder for 'and here the magic happens'. Despite this, a clear conception of emergence as a non-mysterious relation that holds between a system and its parts (taken individually) has a lot of philosophical value. In chapters 1

[1] At a conference dedicated to discussing the various ways in which neuroscience is reshaping our conception of the mind that took place at Macquarie University in June 2016, the central themes that drive emergence were implicitly discussed in a number of presentations, and explicitly so in others (viz, Michael Anderson's paper). But, at the closing of the conference when I asked what people had in mind when they used (or thought of) the term emergence—Did they think that it implied that there was something fundamentally mysterious about the way certain things hang together with certain others? Or, did they think of it merely as a useful term for describing the way in which parts of a system are related to the system as a whole (perhaps especially when the parts are interacting in a complex non-linear fashion)?—the tone in the room went cold. After an initial silence, broken only by a few splattered coughs and awkward collar adjustments, Colin Klein, half seriously, half-jokingly, addressed the speakers collectively and said "Ok, who brought up emergence?"

and 2 of this thesis I develop and defend a conception of emergence—emergence as systemic novelty—that is in keeping with discussions in systems biology and other sciences of the mind.

The picture that arises from this, emergent materialism, sees conscious experience as a real, causally potent, and yet non-mysterious feature of certain complex material systems. Naturally, accounting for consciousness as a non-mysterious feature of certain complex material systems presents a serious challenge. In chapters 3 and 4 I present the problem in more detail and offer the beginnings of a novel solution. The basic plan is as follows.

In chapter 1, I introduce and motivate the basic metaphysical thesis of materialism and provide several reasons for preferring materialism over the alternatives. Unfortunately, providing an account of the relationship between the basic entities at the level of microphysics and the non-basic entities that emerge as a result of the organized interactions of these is not straightforward. After rejecting both 'crude reductionism' and 'supervenience' I argue that a conception of emergence that sees emergent phenomena as both *non-mysterious* and *ontologically novel* can do the job. In chapter 2, I discuss the concept of emergence in some detail, and after rejecting 'epistemological', 'spooky', and 'naïve' notions of emergence, I offer an account of emergence in terms of 'systemic novelty'. This account has three central commitments: first, it recognizes emergence in any system (whole) that has at least one property that is of a *kind* not possessed by any of its constituent parts; second, it holds that a system's properties are *non-causally determined* by the organized interaction of its constituent parts; third, it holds that the environment in which a system is embedded plays an ineliminable role in determining how its parts will be organized and how they will interact.

After presenting this view, I briefly discuss Jaegwon Kim's causal exclusion argument, and why it does not threaten (and is not intended to threaten) this conception of emergence. I argue that 'levels of mechanism' is the most appropriate conception of levels for understanding emergence as systemic novelty, and that the causal exclusion argument does not threaten entities at higher levels of mechanism. However, while emergence as systemic novelty endorses emergent causation, it does not license claims of spooky 'downward' causation—wholes do not have any causal powers over the behaviour of their parts. Following Craver and Bechtel (2007, 2013), I

argue that all apparent instances of inter-level causation can be understood without loss in terms of the ordinary intra-level causal interactions between the constituents of a system, and the constitution relation that holds between a system and its components.

The version of materialism that I endorse, emergent materialism, says that everything that exists is either: a fundamental microphysical entity, a non-mysteriously emergent system, or a property instantiated in one of these. Chapter 3 presents what is commonly considered to be the biggest challenge for this view: accounting for consciousness. Emergent materialism holds that the relationship between the properties of a system, and the organized interaction of that system's parts in context, should be understandable. Unfortunately, there appear to be principled reasons why our current explanatory methods are incapable of providing an illuminating account of conscious experience. There is, as Levine has put it, a deep "explanatory gap" between the kinds of processes that take place in human brains and conscious experience (1983). In chapter 3 I discuss why accounting for consciousness is so hard and reject the idea that this hardness is a mere artifact of outdated theories of explanations. Following Chalmers, I argue that the explanatory methods of material science are only capable of explaining structures and functions. And, since we do not conceive of conscious experience in structural and functional terms, explaining consciousness presents material science with a uniquely hard problem. Recent developments in the philosophy of science do not undermine this fact: the hard problem of consciousness re-emerges with just as much potency when the more modern, mechanistic account of explanation is employed.

In the final chapter, chapter 4, I turn to discuss a number of metaphysical arguments that claim that the explanatory gap between the facts about neural processes and the facts about conscious experience is not merely an epistemological issue, but an ontological one. Advocates of these arguments believe that the hard problem of consciousness is not merely *hard*, but *intractable*. They think that somewhere along the line consciousness must be taken as fundamental. After presenting these challenges I explore the prospects of resolving them. I argue that there are in fact two problems of consciousness, two explanatory gaps, not merely one. Consider the following two questions that we could ask about the neural processes that followed the stubbing of my toe: 1) Why were these neural processes accompanied by a subjective experience? and; 2)

Why did that subjective experience have the particular 'painful' phenomenal quality that it did? These two questions rely on a distinction between what I call *experience in general*, and *qualitative character*. Rather than presenting a single hard problem, there are two problems associated with phenomenal consciousness that both, currently, resist explanation: there is the problem of accounting for the existence of *experience in general*; and, there is the problem of accounting for the *qualitative character* of experience. Although this distinction not new, it has implications that have been over-looked. I argue that it is only the first of these questions that poses a metaphysical challenge for materialism. If the materialist can account for why certain material processes are accompanied by subjective experience, then materialism is vindicated irrespective of whether or not the specific qualitative character of those experiences can also be accounted for. Furthermore, I argue that there is reason to be optimistic with regard to the prospects of accounting for the existence of *experience in general*. If *experience in general* is not itself something we experience, then there is room for a conceptual renovation that will allow us to understand *experience in general* in terms of the kinds of processes that take place in certain material systems. There is reason to be optimistic with regard to the prospects of material science eventually providing an illuminating explanation of how *experience in general* hangs together with the kinds of processes that take place in functioning human brains.

Chapter 1: Materialism

Despite its current position as the dominant world view among the scientifically informed community today, materialism (or physicalism) has had a long and troubled history. The basic idea is that everything that exists (from atoms to cells to persons to social systems) is constructed from a relatively small set of *fundamental* micro-physical building blocks, usually taken to be those postulated by physics. Another way in which the thesis has been presented, is that if materialism is true, then were you to duplicate all and only the *fundamental* micro-physical entities (whatever they turn out to be) and their relations to and interactions with one another, you would, by necessity, also duplicate all the chemical, biological, social, economic, and phenomenal features of the world.[2] You would have, in other words, a *complete* duplicate.

Variations of the thesis date back at least as far as the Ancient Greek and Indian philosophers, and have persisted in some form or other since then.[3] Despite its longevity, materialism has had very few adherents among eminent philosophers prior to recent times. Historically, materialism met with considerable disfavor, not merely because it flies in the face of theological assumptions about the existence of God (although this certainly played a role), but also because the most compelling argument for materialism—the causal argument—was not available until relatively recently. The basic idea behind the causal argument is that we have good reason to believe both: 1) that every event that occurs is caused by some prior material event, and 2) that no event has more than one sufficient cause. If these two premises are true, it follows that anything that does not fit within the materialist picture must be incapable of effecting change in the world. Returning to the example from the introduction, suppose we want to explain why, after stubbing my toe on the stray chair I proceeded to inspect it for damage. We have good reason to believe that a sufficiently complete causal story can be told as to why I inspected my toe in terms of material processes: nerve cells firing, patterns of brain activity, muscles contracting, etc. The problem that this poses for any view that holds that experiences are not processes of some sort taking place in material systems, is that they—the experiences—are thereby left out of the causal story. They are

[2] Phenomenal features are those associated with conscious experiences: pains, emotions, visual experiences, etc.
[3] Examples include the atomism of Democritus and Lucretius, as well as the atomism evident in the Nyaya-Veiseskia and Jainist schools of thought. Whether Indian thinking influenced the Greeks, or visa versa, or whether both developed independently is a matter of debate (Berryman, 2008; Teresi, 2003).

'causal danglers' to co-opt a term of Feigl's (Feigl, 1967). If the pain that I felt on that occasion was not a feature of material processes taking place in my brain, then the fact that my toe 'hurt' had no causal role to play in the events that lead to me inspecting it. The causal argument seems to force the anti-materialist about consciousness to the extremely counter-intuitive position known as epiphenomenalism—the view that conscious experiences are real, but causally inert. I will return to the causal argument in section 1.3. Before doing so however, it is worth laying some more robust foundations.

This chapter will proceed as follows. In section 1.1 I contrast the general idea of materialism with rival theories before briefly addressing some initial concerns about how to formulate materialism in section 1.2. In section 1.3 I provide some motivation for endorsing materialism by briefly considering three arguments in its favor: the argument from methodological naturalism, Occam's Razor, and the causal exclusion argument. In section 1.4 I address a terminological issue. Although the terms 'physicalism' and 'materialism' are typically taken to be synonyms, they are sometimes used to refer to distinct theses. I intend to use the term materialism to refer to the familiar picture of the world as stratified into distinct levels (very roughly: fundamental particles, atoms, molecules, cells, organisms, etc.), together with the assertion that only the entities of the lowest level are instantiated in a basic way, everything else (and all the properties instantiated in those things) arises in a non-mysterious fashion by virtue of the organized interaction of those fundamentals. The chapter will conclude with a brief discussion of two failed attempts to account for the relationship between the entities at different levels—crude reduction, and supervenience—before providing an account of this relation in terms of emergence in Chapter 2.

1.1 What Materialism is Not

An important initial step in introducing and clarifying the concept of materialism is to specify what it is not: what views it rules out or is incompatible with. First and foremost, that means dualism. Historically, dualism has been the dominant metaphysical view and it continues to dominate outside of the philosophical and scientific communities (although its grip appears to be waning with each generation). The distinction between body and soul common among religions is a distinctively dualist notion. On these views, the body and the soul are taken to be two fundamentally different things (or substances), the latter being able to live on in some ghostly

realm after the demise of the former.

The notions of dualism that I will focus on here are those with their roots in the dualism of Rene Descartes. Descartes' was a dualism between mind (which he saw as essentially subjective and rational, consisting of thoughts and experiences), and body (which he saw as essentially material, spatially located, and objectively observable). To get a more intuitive grip of this distinction consider again the example from the introduction. As I kicked the stray chair leg, a host of material events took place: nociceptors under my skin and in the joints of my toe sent bursts of signals to my brain where a whir of information processing lead to a number of objectively observable behaviors. All of this, according to Descartes' view, was just permutations of the 'material substance'. By contrast, the 'pain' that followed, was constituted by an entirely different substance, that of the mind.[4]

The traditional philosophical way of putting this is in terms of concrete particulars. Typically, when philosophers talk of particulars they have in mind a contrast with properties. Particulars (or individuals) are the sorts of things that can instantiate properties. They are things like books, chairs, The University of Adelaide Philosophy Department, the number 7, and so on. Properties on the other hand are the sorts of things that are instantiated by particulars. A book, for example, may instantiate the property of being well written; a chair, being comfortable; the University of Adelaide Philosophy Department, being underfunded; the number 7, being prime; and so on.

The class of particulars can be divided into those that are concrete and those that are abstract. Very roughly, concrete particulars are those that occupy some space at some time. Things like

[4] It has been pointed out to me by a helpful reviewer that there is a dispute among Descartes scholars as to whether or not he considered *sensations*, such as pain, to be part of the realm of mind. The debate seems to revolve around a particularly awkward passage in his second meditation.

> And finally it is the same I that perceives by means of the senses, or who is aware of corporeal things as if by means of the senses: for example, I am seeing a light, hearing a noise, feeling heat.—But these things are false, since I am asleep!—But certainly I seem to be seeing, hearing, getting hot. This cannot be false. This is what is properly meant by speaking of myself as having sensations; and, understood in this precise sense, it is nothing other than thinking. (Descartes, 2008, p. 74)

As I read him, he is suggesting that when he sees a light, hears a noise, and feels heat, while he can doubt that there is in fact any light, noise, or heat, he cannot doubt that he has an experience as if there were light, noise or heat, and as a result these experiences are a part of his mind. I won't explore this further since nothing of importance to this thesis turns on it. Certainly, the modern conception of substance dualisms sees experiences such as pains as among 'the mental' and wholly distinct from 'the physical'.

books, trees, and people are all concrete particulars. In contrast abstract particulars do not seem to occupy space and time in the same way. For example, the number 7, if it exists, is not the sort of thing that has a particular spatial location.

Cartesian dualism holds that there are two essentially different kinds of concrete particulars, those made of material substance and those made of mental substance (whatever that might be). Cartesian dualism has very few (if any) adherents among scientists and philosophers today. Nonetheless, dualistic intuitions persist. More than one quarter of the 3226 active philosophers that participated in a recent survey conducted by David Chalmers and David Bourget either endorse or lean-towards non-physicalist (non-materialist) views with respect to the mind (2013).[5] What is abundantly clear however, is that modern dualism is not a dualism of substances. Todays' dualists do not believe in the existence of an immaterial mind or mental realm. Rather, modern dualism is a dualism of properties.

Property dualists hold that although there is only one kind of substance (material substance) there is a duality of kinds of properties that it can instantiate: material properties, and mental properties. The property dualist is happy to accept that mental properties (like the pain I felt when I stubbed my toe) are properties of material systems (namely brains, or brain-body-environment systems), however, she will deny that this suffices to make them material properties. They are, she will insist, essentially different.

For those not steeped in the philosophy of mind (or alternatively, for those deeply entrenched materialists) this might seem hard to get an intuitive grip on. What sense can be made of the idea that something can be a property of a material system but not *be* a material property? A little reflection however reveals why someone might want to say this.

Recall again when I stubbed my toe. Stubbing my toe was followed by two, prima facie distinct, phenomena: certain patterns of activity in my brain, and a particularly uncomfortable feeling in my left foot. The materialist is committed to saying that these two prima facie distinct phenomena are really one and the same thing. Although we may use different terms to describe them,

[5] It is important to note that neutral monist theories, which hold that the fundamental building blocks of nature are neither physical nor mental, but rather are somehow 'neutral' between the two, are also included in this figure.

materialism, as I will present it, is committed to holding that the pattern of activity in my head, and the feeling of pain in my foot, are token identical. [6] That is, they stand in the same kind of relation to one another that the *particular global pattern of interactions* that happens to be instantiated in the organized collection of H_2O molecules that now fills my cup, and the *liquidity* of the water than now fills my cup stand in: they are one and the same thing being referred to in different ways. The property dualist wants to insist that this simply cannot be. To illuminate why she might want to say this, consider the following thought experiment.

Imagine you are in the market for a new car and are currently perusing a Honda Civic at the local Honda showroom. Two salespersons, Jerry, and Elaine, notice your interest and bustle over to attempt to make a sale. The Civic on display is in white, and being the savvy shopper that you are you know that white cars have a tendency to look dirtier than darker colored cars. You want to know whether or not the Civic comes in a nice chocolate brown. Who should you ask? Elaine, or Jerry? Assuming both are equally capable salespersons, it does not matter who you ask. Whether or not there are any chocolate brown Civics for sale is an objective fact about the world. It is a fact that both Jerry, and Elaine have access to.

Suppose however, that the question that really plagues you is not whether or not the dealership has any chocolate brown Civics, but whether or not Elaine's left elbow itches. (Perhaps you are an advocate of a rather peculiar philosophical view that 'the good life' is one filled with the purchasing of chocolate brown Civics from dealers with itchy elbows). Who should you ask on this occasion: Elaine or Jerry? Elaine of course. Whether or not Elaine's left elbow itches is not an objectively observable feature of the world in the same way that the availability of chocolate brown Civics is.[7]

[6] Philosophers use the term 'token' to refer to a particular instance of a thing, whereas the term 'type' is used to refer to particular kind of thing. For example, the string of words 'dog, dog, dog', contains three tokens of the word dog, but it only contains one type of word. Some people may want to deny that materialism is committed to either form of identity theory. For example, Stoljar argues that "supervenience physicalism neither implies, nor is implied by token physicalism" (2016). As far as I can see, Stoljar seems to think that token identity can only apply to particulars and not to properties. I am not convinced that this is so, however, I won't explore it at this early stage. Additionally, some versions of functionalism reject token identity, but still claim to be advancing versions of materialism. These views, as far as I can see, are versions of property dualism, and hence not materialism at all.

[7] There are certain objectively observable cues that could help us guess whether or not Elaine's elbow itches: she may be frequently occupied with scratching it, and she may have some visible red bites. But Elaine's behaviour is not

Another way of pressing the same point is to suppose that just before I stubbed my toe we had wired me up to a machine that was able to track, in minute detail, the activity taking place in my brain. You would then have been able to observe, objectively from the third person perspective, the processes taking place in my brain as I stubbed my toe. You would not however (so the property dualist will insist) have been able to objectively observe my experience of pain. What my pain feels like from my subjective point of view, will remain hidden to you. The reason for this, according to the property dualist, is that the properties that characterize my conscious experience of pain are fundamentally different from the properties that characterize the processes that took place in my brain and must be taken as fundamental features of the world. Some may think that this formulation of property dualism is too strong, but I agree with Howard Robinson that "genuine property dualism occurs when the ontology of physics is not sufficient to constitute when there is" even at the token level (2016).

Another view that needs to be distinguished from materialism is panpsychism. Although often considered a version of materialism (Chalmers, 2010; Stoljar, 2010; Strawson, 2006) its central intuitions are distinctly dualistic. Although panpsychism agrees with materialism that the world is indeed constructed wholly from the set of physical fundamentals, where it differs is that it sees conscious experience as among those fundamentals. Panpsychism argues that a complete catalogue of the fundamental building blocks will include consciousness in some form. The main motivation for panpsychism is that there appear to be good reasons to think that conscious experience cannot be accounted for in terms of the organized interactions of things that are not themselves conscious experiences. I will return to discuss these reasons as well as the thesis of panpsychism in more detail in chapters 3 and 4.

What distinguishes materialism from each of these alternatives is that while they each take conscious experience to be a fundamental feature of the world in, materialism does not. Rather, materialism says that the inventory of fundamental features of the world is exhausted by the

her experience. Elaine may have been aware of your preference for making purchases from people with itchy elbows and staged an elaborate rouse. Perhaps she is merely pretending to have an itchy elbow to help her make a sale. Notice then, that even asking Elaine whether her elbow itches will be inconclusive. She can simply lie. All this is in favour of the property dualists' claim that there is something essentially different about mental properties. They are something that only the subject of those mental properties has access to.

basic posits of fundamental physics, and that consciousness is not among them.

Before providing some reasons for preferring materialism to its rivals I will briefly address an initial worry with defining materialism in terms of the fundamental entities postulated by physics.

1.2 An Initial Concern: Hempel's Dilemma

As Carl Hempel (1969) pointed out, there is a bit of a worry with formulating materialism in terms of the 'basic posits of fundamental physics', because physics routinely revises its set of fundamentals. We are faced with a dilemma: either 1) by 'basic posits of fundamental physics' we mean the fundamentals posited by current physical theory, in which case materialism is almost certainly false; or else 2) the phrase 'basic posits of fundamental physics' refers to the fundamentals posited by some final (or complete) theory of physics, in which case we don't know what materialism says. Indeed, as I mentioned in the previous section, there are a number of thinkers around today—the panpsychists—who hold that a final theory of physics will include consciousness among its fundamental posits.

There have been a number of responses to Hempel's Dilemma. David Lewis argued that although Hempel is right to say that we don't know precisely what a final theory will say, we can reasonably think that current physics goes a long way towards a complete and correct inventory, and since "the physical nature of ordinary matter under mild conditions is very well understood" we shouldn't expect any further revisions to radically alter our understanding of ordinary macroscopic phenomena (1994, p. 412). Similarly, Jack Smart argued that the physics of "bulk matter" is essentially complete, and at least when it comes to the mind-body problem we are dealing with macroscopic "bulk matter" (1978, p. 340). It seems reasonable to think that whatever revisions future physics makes will likely occur at scales and energy levels far removed from that of ordinary bulk matter like human brains and the process that take place in them. Andrew Melnyk (2003) takes a different tack and suggests that we should formulate materialism in terms of the properties that current physics postulates. We should, he argues, take the same attitude towards physicalism as scientific realists take towards current physics; namely that it is objectively superior to the alternatives.

Although I am sympathetic with each of these responses, I think the move that Spurrett and

Papineau (1999) make is the right one. They argue that we don't need a precise definition of what the fundamental posits of physics are to address the challenge that consciousness presents. Rather, we can simply say that whatever the fundamentals are, they aren't mental.[8] This strategy can be applied wholesale. In other words, we can arrive at a useful, working definition of materialism by defining it, not in terms of which fundamental posits it is committed to, but rather, in terms of which features of the world it takes to be instantiated in a non-basic way. Doing so gives us an unwieldy but working account of materialism as the view that all the entities and properties associated with macro-physical, chemical, biological, sociological, and psychological phenomena are instantiated in a non-basic way.

A number of thinkers have expressed concerns about this approach. For example, Daniel Stoljar has suggested that this approach would rule out, as if by fiat, the possibility that the fundamentals are both mental and physical (2016). This however is a mistake. Such a formulation does not rule out panpsychism as a possibility. It simply says that panpsychism is not a form of materialism. While there may be intuitive reasons to resist this, there are also good reasons to embrace it. Certainly, panpsychism and materialism, as cashed out in this thesis, are two very different kinds of views.

There is a real threat here of the debate over the truth or falsity of materialism turning into a terminological issue. As far as I am concerned, the interesting, and plausible thesis, is the one that says that conscious experiences (somehow) arise with the organized interaction of non-experiencing entities. If you don't think such a view amounts to materialism about the mind, so be it, call it whatever you will. Ultimately what we call the position is unimportant so long as we are clear on what it says.

The thesis I shall be defending, which I will call materialism, is the thesis that *all macroscopic entities and their properties are non-basic, and everything microscopic is non-mental, non-social, non-biological, and non-chemical.*

1.3 Why Believe Materialism Is True?

Perhaps the most obvious reason for believing that materialism is true is what Stoljar has called

[8] See also Joseph Levine (2001), and Jessica Wilson (2006).

"the argument from methodological naturalism" (Stoljar, 2016). The scientific method has a long standing and well-furnished record of tracking the truth. As a result, it is rational to let the methods and findings of natural science guide our metaphysical commitments. And, as a matter of fact, science tells us that materialism is most likely true. Perhaps more interesting however are the reasons underlying science's preference for materialism. I take there to be two: Ockham's Razon (or parsimony) and causal closure.

Ockham's Razor tells us that if two theories have the same explanatory power but require postulating different entities, then we should prefer whichever postulates the fewest. For example, until the early twenty-first century, there were two competing theories of how it was that living organisms were capable of doing the many remarkable things that they do. One theory, *mechanism*, held that there existed within living organisms, some exceedingly complex, but ultimately material mechanisms, which accounted for all the capacities of living organisms. The other theory, *vitalism*, doubted that material mechanisms could possibly account for the performance of all the seemingly miraculous things that living organisms are capable of, and so postulated an additional force 'elan vital' with which all reproducing organisms were endowed.

Although evidence had been mounting in favour of mechanism for some time, the uncovering of the DNA double helix structure by Francis Crick and James Watson in 1953, was the final nail in the vitalists' coffin. Although many of the details continue to elude us even today, after Crick and Watson it became extremely hard to deny that material mechanisms were capable of accounting for all the phenomena for which 'elan vital' was postulated. We thus had two competing theories, *mechanism* and *vitalism*, with equal explanatory power, but one which postulated a force that the other showed to be superfluous. Ockham's Razor demands that we prefer the simpler, more elegant, of the two theories, and indeed that is what happened. As soon as it became plausible that material mechanisms were capable of accounting for all the phenomena for which 'elan vital' was postulated, advocates of vital forces conceded.[9]

[9] Interestingly there were some tantalizing hints as early as 1828 when Friedrich Wohler produced the organic chemical 'urea', from the inorganic compound 'ammonium cyanate'. A number of thinkers, Peter Slezak for example (personal communication July 2016), regard this as the turning point in the vitalist mechanist debate. However, Wohler himself was cautious of such claims and much vitalist literature followed. For example, the term 'elan vital'

As we will see in chapter 3 however, Ockham's razor alone is not a sufficiently compelling reason for preferring materialism over the alternatives. While materialism is certainly more parsimonious than the competing dualist and panpsychist theories, it is not clear at this stage that it has sufficient resources to explain all that needs to be explained. Whether or not it can will occupy us for much of chapter 3 and 4 so I will postpone discussion until then.

The most compelling reason to prefer materialism over its alternatives is the one I mentioned at the beginning of this chapter: the causal argument.[10] Roughly, the causal argument says that only material events and processes can be causes, and since we know that mental events can be causes, mental events must be material events.

The causal argument can be fleshed out as a pair of premises, which, if both true, entail the conclusion that mental events are in some sense material events.

The first premise of the argument is:

1. Mental events have material (physical) effects.

This is a very compelling premise. As John Searle is fond of putting it, when I decide to raise my arm the damn thing goes up. The same is true for any number of mental events. My belief that the shops will be closed tomorrow, coupled with my desire to have food to eat for breakfast, played a causal role in my going to the shops this afternoon. Likewise, the pain that I felt when I stubbed my toe on the wayward dining chair caused me to inspect it for damage. This causal interaction also runs the other way. Not only do mental events have material effects, material events have mental effects. If you doubt this there are a number of experiments I can advise you to perform that will, I assure you, be convincing. One involves taking to your *material* hand with a *material* hammer. Another, involves pouring a full *material* bottle of *material* whiskey down your *material* throat. I guarantee that each will have noticeable *mental* effect.

---

wasn't coined until Henri Bergson's 1907 book "Creative Evolution" almost 80 years after Wohler synthesized urea (Bergson, 1911).

[10] As a bonus, the causal argument accounts for why materialism has historically been a minority view. The crucial premise is indeed this second premise, and it was not available until relatively recently. See David Papineau (2001) for discussion.

The second premise is:

2. Every material effect has a complete material cause.

In general, material effects can be fully accounted for by a prior history of material processes that act as causes. Consider again my inspecting my toe for damage after stubbing it on the stray chair. We expect to be able to provide a full physiological account of both why and how that happens by citing, among other things, how external stimuli are transduced into electrochemical signals by particular cells under my skin, how these signals lead to certain patterns of electrochemical activity in my brain, and how the subsequent signals cause the relevant muscle fibers to contract. From a physiological point of view, we expect to be able to model this causal process entirely at the level of functioning material systems (or mechanisms).

These two premises, if they are accepted, force the conclusion that:

3. Mental events are material events.

Of course one does not have to accept either of these premises—and indeed there is ongoing debate as to whether or not we should. The most intuitively natural premise to reject is the second. Perhaps, one might think, there are immaterial causes. On a closer inspection however, this is not a very attractive position. This premise is typically written as the causal closure of the physical, which is often mistakenly read as implying that microphysics is causally closed (see for example Earley, 2008). Many papers have been written refuting the causal closure of the microphysical by citing clear cases of macroscopic causation in chemistry and biology. However, the causal closure premise is not disputing that complex material systems have causal powers that their parts lack. As Jaegwon Kim, the author of the causal argument, has pointed out, that systems can and typically do instantiate macroproperties that have their own causal powers, causal powers that go beyond the those of their microconstituents, is an "obvious but important point to keep in mind" (Kim, 1998, p. 85). To take a simple example due to Carl Craver: a lawnmower can cut grass, a sparkplug cannot (2007). All that is required for causal closure to be true, is that there are no causal powers other than those had by fundamental physical entities, *and* those that systems composed of those entities possess by virtue of the organized interactions of their parts in context.

So unappealing is rejecting premise 2 that a number of prominent anti-materialists such as David Chalmers, and reformed anti-materialists such as Frank Jackson, have argued that premise 1 is the dubious one. They have argued that conscious states are in fact epiphenomenal. They have suggested that although we have extremely strong intuitions against epiphenomenalism, they are, after all, only intuitions. Chalmers (1996) suggests that if we have good arguments for accepting the causal closure of the material world, as he thinks we do, and we have good argument for thinking that experiences cannot be material, as he thinks we do, then we should be willing to give up the intuition that our mental events do in fact have any material effects.

I agree with Chalmers about the causal closure of the material world, but what should we make of these arguments that the mental cannot be material? I will return to these arguments in chapter 3 and 4. My next task is to clear up a terminological confusion that has muddied some discussions of materialism and related issues: Namely, is this view materialism or physicalism?

1.4 A Note on Terminology: Materialism or Physicalism?

Despite for the most part being considered to be synonyms, the terms 'physicalism' and 'materialism' are sometimes used to refer to different theses. Sometimes it is suggested that physicalism represents a somewhat more general thesis than materialism. The root term of materialism, 'matter', has historically referred to the sort of stuff that takes up space. A materialist then would be someone who thinks that everything is constructed from tiny, indivisible bits of extended 'matter'. This was indeed the materialist picture in the 17th century when pre-Newtonian physics ruled the day. Although a dualist about minds, Descartes held that everything else was made solely of tiny bits of unchanging matter interacting with each other via direct physical contact. C.D. Broad refers to this view as ‘pure mechanism’. According to ‘pure mechanism’ there is only one kind of basic thing, ‘bits of extended matter’ and only one kind of basic interaction ‘exchange of momentum via physical contact’ (Broad, 1925).

Of course, developments in physics have shown pure mechanism to be false. In the late 1600’s Sir Isaac Newton postulated the existence of gravity—a force—and expanded the inventory of fundamentals beyond the ‘material’ in the historic sense. After Newton, physics’ inventory of fundamentals included more than merely tiny bits of extended matter bumping into each other; it now also included force fields. Similarly, in the 1860’s when James Clerk Maxwell postulated

electromagnetism as a fundamental force of nature, the set of physical fundamentals grew again.

Modern physics has of course extended on this. Now, not only do we recognize strong and weak nuclear forces as well as gravity and the electromagnetic force, but also point particles—particles that have a location and instantiate various properties, but which lack spatial extension in any traditional sense. In fact, most of the entities and properties postulated by modern physics are not much like 'matter' at all. As a result, a number of theorists have suggested abandoning the term materialism in favor of physicalism. According to those who make this distinction, where materialism is 17th century pure mechanism, physicalism is the view that everything is constructed from the fundamentals postulated by physics.

I do not see the value of this distinction. To quote one of the most influential materialists of recent history:

> [Materialism] was so named when the best physics of the day was the physics of matter alone. Now our best physics acknowledges other bearers of fundamental properties… But it would be pedantry to change the name on that account, and disown our intellectual ancestors. Or worse, it would be a tacky marketing ploy, akin to British Rail's decree that second-class passengers shall now be called 'standard class passengers'. (Lewis, 1994, p. 413)

Not disowning our intellectual ancestors is not the only reason for preferring the term materialism to physicalism however. When the term physicalism was first introduced into the philosophical vernacular in the 1930's by the logical empiricists Otto Neurath and Rudolf Carnap, it represented the thesis that every statement in empirical science is equivalent in meaning to a statement in some common vocabulary, typically understood to be the language of physics. Drawing on this work, in the 1940's-60's Carl Hempel and Ernest Nagel used the term physicalism to describe their particular view about the unification of the sciences and the idea that in the final analysis, all of the special sciences (chemistry, biology, psychology, etc.) will be *reduced* to physics. On their view, the *truest* or *best* explanations are those couched in the terms of physics.

Two things to say about this. First, the idea that everything can be expressed in the terms of physics alone is false. In fact, it is not even true within physics itself: the principles of statistical

mechanics cannot be expressed solely in terms of either classical or quantum mechanics. Suppose we have a glass of water and we add to it a measure of whiskey. If we want to explain why the two fluids mix together, we could go about it by explaining how each individual molecule interacts with each other in exquisite detail and end up with a description of the fluid in terms of the location of each of its constituent molecules. Such an explanation would tell us why this particular glass of whiskey and water mixed together but it would miss the underlying point. The reason measures of whiskey reliably mix with bodies of water is a matter of statistics: there are radically more possible states that the combined fluid can be in, in which the whiskey and water are mixed than there are in which they remain separate.

Second, materialism is first and foremost a metaphysical thesis about the relationship between real features of the world: things that exist independent of our theorizing about them. By contrast, physicalism as introduced by the logical empiricists, represents a linguistic thesis about the relationship between our statements about the world, and as employed by Hempel and Nagel, represents an epistemic thesis about the relationship between our theories about the world. I think Galen Strawson is correct when he points out that, "real physicalism (or materialism) can have nothing to do with physicSalism, the view—the faith—that the nature or essence of all concrete reality can in principle be fully captured in the terms of physics" (Strawson, 2006, p. 4).[11]

Further complications arise when we note that within the philosophy of science literature, the term physicalism is sometimes construed as synonymous with micro-physicalism, the view that ultimately only the fundamental building blocks postulated by physics really exist. For example, according to Mario Bunge, physicalists hold that "although there may be different levels of analysis or description [beyond the fundamental level], these have no counterparts in reality" (Bunge 2003:146). However, many of those advocating views they call physicalism are certainly not committed to micro-physicalism in this sense. David Papeneau, has devoted an entire article to arguing the point (Papineau, 2010). Similarly, Jaegwon Kim, often accused of being a micro-physicalist, has argued that "[p]hysicalism need not be, and should not be, identified with micro-

[11] By "real materialism" Strawson means materialism that takes consciousness seriously, as a real phenomenon, in need of explanation. While he thinks this implies panpsychism, I do not.

physicalism" (Kim, 1998, p. 117).

To quote Kim at length:

> [The picture that we have today] is the familiar multilayered model that views the world as stratified into different "levels," … It is part of this layered picture that at each level there are properties, activities, and functions that make their first appearance, or "emerge," at that level (we may call them the characteristic properties of that level). Thus among the characteristic properties of the molecular level are electrical conductivity, inflammability, density, viscosity, and the like; activities and functions like metabolism and reproduction are among the characteristic properties of the cellular and higher biological levels; and consciousness and other mental properties make their first appearance at the level of higher organisms. For much of this century, a layered picture of the world like this has formed a constant – tacitly assumed if not explicitly stated – backdrop for debates on a variety of issues in metaphysics and philosophy of science. (Kim, 1998, pp. 15-16)

Note that this view is *metaphysical* not *epistemological*. It is a claim about the actual features of the world and the way they relate to each other, not about our conception of those things or our theories about them. As such, the levels in question are not mere "levels of description" as Bunge suggests. Rather, this layered picture forms the metaphysical "backdrop" for debates about those epistemological issues.

There are two interrelated and rather difficult issues in front of us. The first is to say something about how we are to understand this notion of levels, and the second is to specify how entities at different levels are related. In what remains of this chapter I will introduce the mechanistic view of levels before briefly discussing two failed attempts at accounting for the relationship between entities at various levels, namely: *crude reductionism* and *supervenience*. The task of providing a positive account of the relationship between entities at adjacent levels will be addressed in Chapter 2.

1.5 Levels

A thorough analysis of the levels concept is far beyond the scope of this thesis, however, a few quick clarifications are needed. In "The Unity of Science as a Working Hypothesis", Oppenheim

and Putnam carved the world up into six distinct levels based broadly on the kinds of entities studied by each of the major sciences. At the lowest level were the elementary particles, and then came atoms, molecules, cells, organisms, and finally social groups (Oppenheim & Putnam, 1958). Levels, so understood, are not intended to merely carve the world up in terms of size. Rather, they are intended to carve the world up mereologically: in terms of parts and wholes. For example, although elementary particles certainly are smaller than the atoms that they compose, they are lower-level than atoms, not just because they are smaller, but because they are constituents of atoms. Although discussions about the levels concept have become considerably more nuanced in recent years, this loose characterization of levels as 'levels of science' still gets deployed today; as is evident in the quote from Kim above.

Recently, this simplistic mereological account has been called into question. Carl Craver and William Bechtel have each argued that in many situations it is a mistake to say, for example, that an atom is lower-level than a molecule (Bechtel, 2008; Craver, 2007; Craver & Bechtel, 2007). One clear example of this is within functioning neurons. Part of the mechanism via which neurons transmit electrochemical signals involves the rapid influx of Na+ (sodium ions) across the cell membrane via specialised voltage gated sodium channels. Although it is certainly the case that these ion channels are composed of atoms, the ions that cross the cell membrane via these channels (which are themselves atoms) are not its constituents. Rather it is more appropriate to say that both are constituents of the functioning cell. It seems intuitive then to say that although the ion channel is a molecule, and the sodium ion an atom, both occupy the same level within the functioning neuron. As a result, Craver and Bechtel have argued that levels need to be defined locally. They suggest that the best way to understand the idea of levels is in terms of individual mechanisms.

A mechanism is an organized set of parts and activities which collectively exhibit a particular phenomenon of interest. According to the mechanistic account of levels, what belong at each level are the individual active entities which, when appropriately organized (both spatially and temporally) constitute the higher-level mechanism which exhibits the phenomenon in question. Consider the phenomenon of circulation as it is instantiated in the human body. The mechanism that exhibits this phenomenon is the functioning circulatory system: the various parts of the

circulatory system (the heart, lungs, veins, blood, etc.) performing their orchestrated activities as an organized whole. The functioning circulatory system as a whole belongs to one level in the mechanistic hierarchy. One level below the functioning circulatory system are the individual active entities, such as the 'heart pumping' and the 'lungs oxygenating' for example. One level down again are the individual active entities that constitute the components. In the case of the pumping heart, the contracting and relaxing ventricles and atria, as well as the flow directing valves belong to the next level down in the mechanistic hierarchy.

Carl Craver has provided a helpful visual representation of the idea of levels of mechanism in the diagram below (fig. 1). In this diagram, entity X_3 performing activity φ_3 ($X_3\varphi_3$-ing) belongs to a lower mechanistic level that $S\psi$-ing because $X_3\varphi_3$-ing is one of the four active entities which, when appropriately organized, constitute $S\psi$-ing. But $X_3\varphi_3$-ing belongs to a higher mechanistic level than $P_1\rho_1$-ing, since $X_3\varphi_3$-ing is constituted by the organized collection of P's ρ-ing. On the mechanistic account of levels, $S\psi$-ing, $X_3\varphi_3$-ing, and $P_1\rho_1$-ing belong to three different levels of mechanism.

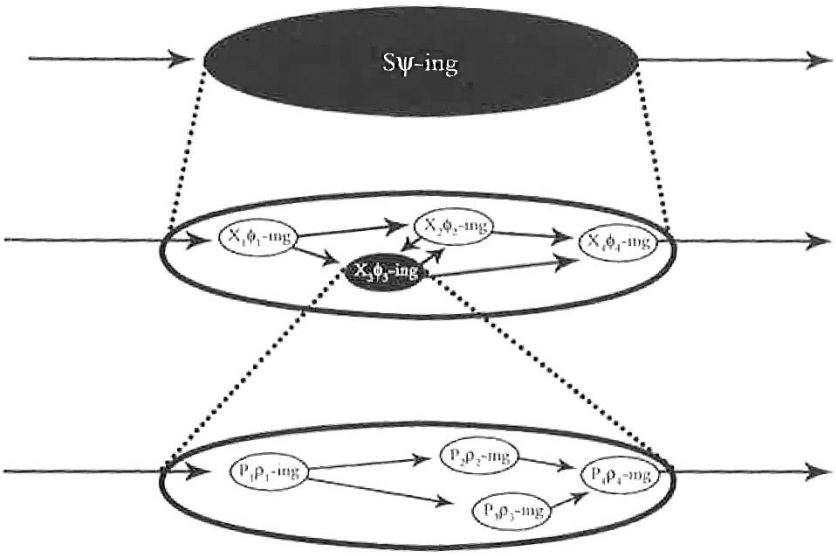

Figure 1 Craver's levels or mechanism: "At the top is a mechanism S engaged in behavior ψ. Below it are the φ-ings of Xs that are organized in S's Below that are the ρ-ings (pronounced ''rho-ings'') of Ps (English pronunciation) that are organized in the φ-ing of Xs" (Craver, 2007, p. 189).

The mechanistic account of levels has several implications. One is that it shows the world to be considerably more stratified than suggested by the six levels of Oppenheim and Putnam. Another

is that it implies that questions relating to whether disconnected entities occupy different levels are meaningless. For example, while my pumping heart belongs to a higher mechanistic level than the contracting and relaxing of my left ventricle, and it belongs to a lower mechanistic level than my functioning circularity system taken as a whole, as Craver points out it, "it makes no sense to ask if my heart is at a different level of mechanism than my car's water pump because there is no mechanism containing the two" (Craver, 2007, p. 191).

Before discussing how entities at different mechanistic levels are related, it is important to distinguish this notion of levels of mechanisms from a related notion prevalent in metaphysical discussion, which Craver (2007) refers to as 'levels of realization' and Kim (1998) calls 'orders' (I will use Kim's terminology). While 'levels of mechanism' refers to the relationship between a mechanism and its constituents, 'orders' refers to the relationship between a mechanism and the various properties it possesses by virtue of being the particular 'components-plus-organization' that it is. Consider my pumping heart again. Levels of mechanism refers to the relationship between my pumping heart as a whole, and the individual active entities that constitute it. The orders relation on the other hand, refers to the relationship between being 'two ventricles and two atria, connected by flow directing valves, contracting and relaxing in a particular orchestrated manner' (call this property *H*), and being a pump. The mechanism in my chest (my pumping heart) instantiates *H*, and by virtue of instantiating *H,* it also instantiates the property of being a pump. The converse, however, is not true. My heart does not instantiate *H* by virtue of being a pump. The reason for this is that there are many ways of being a pump that do not require being *H*. For example, a car's water pump is a pump, but it certainly does not instantiate *H*. The orders relation refers to the kind of relationship that obtains between being a pump, and being *H.* Being a pump is higher-order than being *H.* It is crucially important to note that the orders relation does not cross mechanistic levels. Higher-order properties do not belong to a higher-level of mechanism than their lower-order realizers, and lower-order properties do not belong to a lower-level of mechanism than the properties they realize. Being a pump and being *H* are both properties of the same mechanism (the thing in my chest) and hence are both at the same mechanistic level.

I will return to the distinction between levels of mechanism and orders in the next chapter. For now, note that when I use the term 'level' in this thesis, unless I explicitly state otherwise, it is

this notion of levels of mechanisms that I have in mind.

In the remainder of this chapter I will discuss two familiar accounts of the relationship between levels broadly construed, that fail to satisfactorily relate entities at various levels of mechanism: crude reduction, and supervenience. In the chapter that follows I explicate an ontological notion of emergence—emergence as systemic novelty—to account for the relation between entities at various levels of mechanism.

1.6 Two Failed Attempts at Relating Levels of Mechanism

1.6.1 Crude Reductionism and Nothing-buttery

The simplest attempt to account for the existence of higher-level phenomena without having to postulate new fundamental features of the world is 'crude reductionism' and the infamous 'nothing buttery' of J.J.C Smart and others. Crude reductionism is the view that only a system's parts actually exist, or alternatively, that only a system's parts are relevant in explaining the behavior of the system. Both crude reductionism and nothing-buttery are often presented, as Mario Bunge does, as the view that "composition is everything, whereas structure is nothing" (2003, p. 86). This then, can be easily refuted by noting that structure is relevant even to systems as basic as a simple lever constructed of a stick and a rock; without being organized in the appropriate way, the stick and rock will not function as a lever and will not endow the user with any mechanical advantage.

If crude reductionism was indeed the thesis that structure (or organization) is irrelevant, then I doubt that it is a position that anyone ever actually held. For example, although Smart never retracted his 'nothing-but' claims, he certainly did not advocate the view that 'composition was everything and structure was nothing'. He makes this explicitly clear in a 1981 paper on materialism and emergence where he had this to say.

> … 'nothing buttery' is often said to be a heinous metaphysical crime, but I see nothing wrong with it: in saying that a complex is nothing but an arrangement of its parts, I do not deny that it can do things that a mere heap or jumble of the parts could not. (Who on earth would want to deny this? If 'nothing-buttery' had such an absurd consequence it would be a view that no one has ever held.) (Smart, 1981, p. 110)

Nonetheless, there are deep problems with both nothing-buttery and crude reduction. While Smart accepts that organization (or structure) must be taken seriously (who on Earth would want to deny this?) he overlooks the role that the environment plays in determining the way in which a set of parts will be organized. As I argue in the next chapter, while it is true that the behavior of a system is determined by the organized interaction of its constituents, the environment plays an ineliminable role in determining how they will interact.

Despite the scorn that 'nothing-buttery' receives, there is something obviously right about it. If you subtract from a thing all its parts, nothing remains. If there are no longer any parts to be organized, organization is impotent. However, to move from this to the idea that only the parts are real or the idea that only the parts are explanatorily relevant, is to ignore the vast majority of causally potent features of the world.

1.6.2 Supervenience

One approach to characterizing the relationship between the more traditional discussions of levels of science and levels of mereology that has received a lot of attention over the past 40 years, is to employ the notion of supervenience. But, as a number of people have noted recently, supervenience is too weak to serve as the backbone of a leveled materialist ontology (Kim 1998, Bunge 2003, Stoljar 2010).

Although he didn't use the term supervenience, one of the first to employ the basic idea was G.E. Moore. In trying to understand the relationship between morality and the natural world, Moore noted that no two events that are exactly alike in terms of their non-moral features can differ in terms of moral features. Or as he put it, "[t]wo things can differ in intrinsic *value*, only when they have different intrinsic *natures*" (Moore, 1922, p. 263 my emphasis).[12]

The notion of supervenience is typically cashed out as a relation between sets of properties as follows.

> A set of properties *A* supervenes upon another set *B* just in case no two things can differ with respect to [the instantiation and distribution of] *A*-properties without also differing

[12] Donald Davidson was the first to apply the notion to the mind-body problem. As he put it, "there cannot be two events alike in all physical respects but differing in some mental respect" (1970).

> with respect to [the instantiation and distribution of] their *B*-properties. In slogan form, "there cannot be an *A*-difference without a *B*-difference". (Bennett & McLaughlin, 2005)

Making the notion of supervenience fit with 'levels of mechanisms' is a bit of a fudge since levels of mechanisms are not merely sets of properties, rather they are sets of active entities. Recall Craver's representation of levels of mechanism in figure 1. What belongs at the level below *S* ψ-ing is not the organized collection of X's φ-ing, but the unorganized set of individual active entities {X_1 φ_1-ing, X_2 φ_2-ing, ...}. Mechanistic supervenience then would be the thesis that the properties of a mechanisms supervene on the properties that characterize each of its constituents taken individually.

A criticism that is often leveled against the notion of supervenience is that it ignores the fact that the organization of the mechanism's parts plays an ineliminable role in determining the properties that it instantiates. For example, were the constituents of my heart scattered throughout the galaxy—say my left ventricle was contracting and relaxing as per usual while orbiting Alpha Centuri, while my aortic valve was doing its thing in the proximity of Betelgeuse—they would certainly not constitute a working heart (and I would be in serious need of medical attention). However, I think this criticism is misplaced. While it is certainly true that organization plays an ineliminable role in determining the properties of a whole, it is a mistake to think that supervenience ignores this.[13] To say that the properties the characterize *S* ψ-ing supervene on those that characterize the X's φ-ing is not to say that all there is to *S* ψ-ing is the unorganized set {X_1 φ_1-ing, X_2 φ_2-ing, ...}, nor is it to say that the properties of *S* ψ-ing are fixed once the properties of the individual X's φ-ing are fixed. Rather, it is to say that any two phenomena, S_1 ψ_1-ing and S_2 ψ_2-ing, that are exactly alike in terms of the properties that characterize the organized interactions of their X's φ-ing, are exactly alike in terms of the properties that characterize their *S* ψ-ing.[14]

[13] Although it is perhaps fair to say that a number of those who advocate supervenience do.

[14] It is important to note that nor all of the properties that a mechanism instantiates necessarily supervene on its constituents. Some properties are 'non-local' in the sense that they are properties that a thing possesses by virtue of its relation to features of the wider environment. To use an example due to Teller (1992), being the longest pencil in a box of pencils does not supervene on the properties of the pencil alone, the other pencils need to be taken into account also. Similarly, two physically identical fish may differ in terms of the biological properties they instantiate; one may be 'fitter' than the other on account of being in an ocean rich in the resources that it needs, while the other

When someone working in the philosophy of mind says that an individual's mental properties supervene on the properties that characterize the constituents of the relevant neural mechanism in their brain, they do not mean to say that an individual's mental state is 'fixed' by the unorganized collection of *individual* neurons firing—that would be to deny the importance of organization, and as Smart notes in the quote above, "who on earth would want to deny this?" What they mean is that there cannot be a difference in two individuals' mental states, without there being a corresponding difference in the organized interactions of the neurons firing that constitute the neural mechanisms that collectively exhibit those mental properties. Similarly, when someone working in the philosophy of mind says that consciousness supervenes on the physical, they do not mean that an individual's conscious state is 'fixed' by the unorganized set of *individual* fundamental microphysical active entities that are the ultimate constituents of her brain—that would again be to deny the importance of organization, and as Smart notes in the quote above, "who on earth would want to deny this?" What they mean is that there cannot be a difference in two individuals' conscious states, without there being a corresponding difference in the organized interactions of the microphysical active entities that constitutes the atoms, that constitute the molecules, that constitute the cells, that ultimately constitute the neural mechanisms that exhibit those mental properties. In other words, the claim that the mental supervenes on the physical is the claim that two functioning brains (whole brains) that are exact microphysical duplicates are exact mental duplicates also.[15]

The thesis that the local properties of a mechanism supervene on the properties that characterize its constituents is extremely plausible. Unfortunately, the supervenience relation is insufficient for formulating materialism since it is consistent with, and accepted by, a number of metaphysical positions that are distinctly not materialism. Because supervenience refrains from specifying in detail the nature of the covariation between two families of properties (it just specifies that covariation occurs), a property dualist, who holds that mental properties (such as pain, desire, or

is in a live tank at a sushi restaurant. It is only the local (or non-relational) properties of a mechanism that supervene on the properties of its constituents.

[15] This is in fact the claim that the mental supervenes locally on the physical. Some thinkers maintain that mental properties are more analogous to properties like 'biological fitness' or 'being the longest pencil in a box'. But I won't explore this issue here.

love) are fundamentally non-material properties, can still embrace supervenience. She can maintain that although mental properties are fundamentally different from the properties that characterize the processes that take place in normally functioning human brains, nonetheless, the former supervenes on the latter.[16]

For this reason, materialists should seek a more robust relation to account for the relationship between entities at the various levels of mechanism. A number of relations have been discussed in the literature: realization, grounding, and emergence. While I think that there is considerable philosophical merit to both realization and grounding, and even though much of what I say below could be cashed out in those terms, the notion of emergence has some advantages over the alternatives. For one, emergence has more robust ties with empirical science and, as a result, can facilitate interdisciplinary collaboration between philosophers and scientists. Such collaboration can lead to a more scientifically informed metaphysics and a more philosophically sound science. Furthermore, it has the potential to do more than merely provide an account of the philosophical nature of the relationship between the levels of reality, but in addition, provide us with an illuminating, empirically guided account as to why that relation holds and how it works.

In the next chapter I will give a brief survey of some of the many ways the term 'emergence' has been used in philosophy and science, and examine in detail the sense that is relevant to the formulation of materialism and to the mind-body problem, namely: emergence as systemic novelty.

[16] It is not clear whether she can consistently endorse a mechanistic view of levels though, since implicit in the mechanistic account of levels is the idea that higher-level phenomena can be accounted for in terms of the organized interactions of their lower-level constituents and the property dualist is likely to reject this.

Chapter 2: Emergence

In chapter 1, I sketched the basic outlines of a version of materialism and provided some motivation for preferring it to the alternatives. As we saw, the materialism that I advocate endorses a leveled picture of the world and argues that all macroscopic entities and their properties are non-basic, and everything microscopic is non-mental, non-social, non-biological, and non-chemical. It is only when these fundamentals become organized in particular ways that they can collectively constitute whole systems (or mechanisms) that possess higher-level properties. Towards the end of last chapter I rejected two failed attempts at accounting for how the lower-level entities are related to the higher-level systems that they compose. I rejected 'crude reductionism' on the grounds that it denies that there are any genuine higher-level (systemic) entities at all, and I rejected 'supervenience' on the grounds that it is compatible with some forms of dualism.

In this chapter, I explore the concept of emergence and offer an account of how entities at different levels of mechanism relate to one another. The chapter will proceed as follows. In section 2.1 I lay some foundations and briefly mention a host of uses of the term emergence that I will not be endorsing. In section 2.2 I reject an intuitive but ultimately naïve understanding of emergence that sees higher-level phenomena as somehow 'squirted out by' or 'secreted by' the organized interactions of lower-level entities. In section 2.3, I explicate the version of emergence that I endorse—emergence as systemic novelty. In section 2.4 I address the issue of emergent causation. While emergence as systemic novelty does endorse the emergence of higher level causal powers, it does not license downward causation, nor is it susceptible to Kim's causal exclusion argument. I argue that emergence as systemic novelty allows us to understand how causally potent entities can emerge without jeopardizing our materialist commitments.

2.1 Some Foundations

The concept of emergence is used in a wide range of disciplines: including quantum mechanics, chemistry, biology, sociology, economics, mathematics, and philosophy. While there is certainly some overlap in the way these various fields use the concept, there is also considerable discrepancy. My goal here is not to try to provide a unified account of emergence that applies to all its disparate uses. In fact, I suspect that no such unified account is possible. Rather, I will seek

to clarify a conception of emergence that is prevalent in systems thinking and the biological sciences; one that can account for the relationship between entities at different levels of mechanism. The sense of emergence that I am interested in can be captured by combining the Aristotelian slogan "the whole is more than the sum of the parts" with the idea that, as Murray Gell-Mann puts it, "you don't need something more to get something more" (Gell-mann, 2007). Antti Revonsuo expresses the basic idea in slightly more detail as follows:

> When entities of a certain type become organized in complex ways, engaging in sophisticated causal interactions and forming complex structural and functional wholes, entirely *new* types of phenomena or *new* kinds of properties, unlike those had by any of the parts of the system, may appear in the phenomenon as a whole. (Revonsuo, 2010, p. 26)

It is common for philosophical discussions of emergence to focus on emergent properties, however, in what follows I will also speak of emergent *things* or *entities* (or *emergents* – with a 't'). The reason for this is that I have a fondness for the commonsense metaphysics advocated by Mario Bunge. Bunge, after rejecting the Platonic idea that properties can exist in the absence of some *thing* instantiating them (as if there is some realm in which possible properties wait in earnest and spring forth whenever they are needed), argues that if there are emergent properties, then there must also be emergent *things* which instantiate those properties (1977, 2003, 2010). That is, when a purely material system reaches a certain degree of complexity and exhibits a genuinely new property, one not possessed by any of its parts, it is not merely the property which must be considered emergent, but the system also. In other words, the system which instantiates the emergent property is considered to be a novel *thing* in its own right, and not merely a collection of parts.

To see why we might want to say this consider the familiar example of emergence briefly mentioned in the introduction: the emergence of 'liquidity'. In the common sense of the term, something is *liquid* if it flows in a particular way and has a relatively constant volume (i.e., it does

not dissipate like a gas, but nor does it have a fixed shape like a solid).[17] Now consider a single H_2O molecule. Although no individual H_2O molecule is itself liquid—the property of 'liquidity' simply does not apply to individual H_2O molecules—if you happen to find a large number of H_2O molecules in close proximity, together they can constitute a liquid body of water. Notice that it is not the individual H_2O molecules that constitute the body of water that instantiate 'liquidity', rather, it is the collection taken as a whole. As a result, in order to avoid free floating properties, we must treat the collection as a *thing* in its own right.

"But hold on", one might worry, "aren't we thereby unnecessarily, and problematically, multiplying entities?" If I have a bottle of H_2O molecules at room temperature, I also have a bottle of liquid water. But I certainly don't have *two* things. I have one bottle of liquid water, which just so happens to be ('be' in the sense of 'be token identical with') a bottle filled with H_2O molecules with a particular organization. This, however, is nothing to get into a mereological tizz over.

Mereology is the philosophical study of the relation between parts and wholes. Unfortunately, traditional mereology has a knack of understating (or perhaps overlooking) the importance of organization in relating parts to wholes.[18] I have no intentions of discussing mereology in detail here. I will simply stress that there is more to emergence than merely the part-whole relation: organization can and typically does play an ineliminable role in determining the properties of the whole (and as we will see, the environment can and typically does play an ineliminable role in determining that organization).

To get a clear picture of the importance of organization, imagine yourself diving into a swimming pool filled (as they typically are) predominantly with H_2O molecules. The way in which those molecules are organized will have considerable impact on whether or not you find your swim enjoyable. If the H_2O molecules are loosely bound via the continuous stretching, breaking, and reforming of weak hydrogen bonds, then you will splash, pleasantly, into *liquid* water. By contrast,

[17] A standard chemistry textbook account is that "like gases, liquids are fluid, and most flow easily from place to place. Unlike gases, however, liquids are compact, so they cannot expand or contract much." (Blackman, Bottle, Schmid, Mocerino, & Wille, 2016, p. 260)

[18] Although if I understand him correctly, Roberto Loss has made considerable progress on this front. In a recent paper, Loss has pointed out that the two competing views on the relationship between a whole and its parts—"composition as identity", and "grounding"—are in fact compatible. On his view, the "scattered plurality" of parts ground the whole, which is identical to the organized set of parts (Loss, 2015). This seems right to me.

if they are rigidly bound into a lattice structure, you will splat, not so pleasantly, onto *solid* ice.[19]

As I mentioned, the term emergence is used to refer to a wide range of ideas. It is worth briefly mentioning some of those here to mitigate confusion before explicating in more detail the sense in which I will be using the term. Sometimes emergence is used to describe 'the coming into existence of something new', such as the evolution of a new species, or the coagulation of matter as the universe gradually began to cool following the big bang. It is this sense that Harrold Morowitz has in mind when he outlines twenty-eight steps that have led to the world as we currently know it (2000). This *diachronic* sense of emergence is not what I have in mind. Rather, the sense of emergence that I am interested in here is the *synchronic* sense which tracks the relationship between a thing, and its constituents at a particular time (or over a particular time period).[20] Although an account of how the various phenomena of the world emerged in the diachronic sense would certainly be interesting, it presupposes what I will be arguing, namely that the various phenomena of the world can in fact synchronically emerge from simpler entities.

Another way that the term emergence is used is in reference to features of the world that elude our current attempts at explanation. Elanor Taylor advocates such a use (2015). In her view, the reason consciousness is an emergent phenomenon is not that conscious experience is ontologically distinct from the underlying neural constituents, but merely because we currently cannot explain it scientifically. On her view, emergence is a feature of our attempts to understand the world and not necessarily a feature of the world itself. I shall not be exploring the various versions of epistemological emergence in this thesis since it is the ontological sense of emergence that is relevant to relating the systems with emergent properties to their constituent parts and activities.[21]

Somewhat confusingly, the distinctly epistemic notions of *deducibility* and *explicability* have also been used to characterize versions of ontological emergence. On these accounts, a thing or property is emergent just in case it cannot, as a matter of principle, be *deduced from*, or *explained*

[19] Credit due to Liz Scheir for this vivid example (Schier, 2010).
[20] Morowitz also discusses emergence in this synchronic sense but the general outline of his book provides a nice example (or host of examples) of emergence in the diachronic sense.
[21] For detailed discussion on this epistemic sense of emergence see (Silberstein, 2002, 2012; Van Gulick, 2001)

in terms of, the organized interactions of its parts. This *in principle* inexplicability or indeducibility is not tied to human limitations. Rather, emergence in this strong sense, if it exists, entails that the world itself is fundamentally mysterious. Such strongly emergent phenomena would be inexplicable or indeducible even to a god.

These strong notions of emergence stem from the work of the British Emergentists, most notably C.D. Broad. According to Broad, certain features of wholes cannot be deduced even given complete knowledge of their parts, together with the way in which they are organized in the whole (1925, p. 61). Rather, these features must "simply be swallowed whole with that philosophical jam which Professor Alexander calls 'natural piety'" (1925, p. 55). Precisely what Broad meant by "complete knowledge" is an interesting question. Although he explicitly states that it applies only to the parts as they are when not incorporated into the whole in question, it remains unclear exactly what "complete knowledge" is supposed to mean (1925, p. 61). If "complete knowledge" of a thing is to be understood as an exhaustive list of its macroscopic properties, then I think Broad is right. As we will see, the environment often plays an ineliminable role in determining the way in which a thing's parts will interact, and, as a result, in determining what properties a thing instantiates. Being included in a more complex whole changes its environment, potentially changing its internal organization, and, potentially changing the properties it instantiates as a result. However, if "complete knowledge" is understood as an exhaustive account of the mechanisms responsible for each of its macroscopic properties, then I am inclined to reject Borad's claim. More recently similar strong versions of emergence have been put forward by Timothy O'Connor (2005), David Chalmers (2006), and Silberstein and McGeever (1999): O'connor has argued that "some basic [fundamental] properties are had by composite individuals" (2005); Chalmers has argued that phenomenal consciousness is an example of this stronger sense of emergence (2006); and, Silberstein and McGeever have argued that this sort of emergence occurs in quantum mechanics (1999).

Emergence has also been tied to the notion of *unpredictability*. Although in principle both explainable and deducible, the behavior of complex chaotic systems is in principle *unpredictable*: there are no computational short cuts that allow you to know in advance what the future states

of a complex system will be.[22] Unpredictability is typically associated with notions of diachronic emergence since it is the future states of a system that cannot be predicted. However, in an interesting recent paper Boogerd et al. have argued that there is a synchronic variant to this notion of unpredictability as well (2005). Drawing on C.D. Broad, they point out that even if you have complete knowledge of a systems parts as they behave when not incorporated into the relevant system, and complete knowledge of how they are organized in the system of interest, there is no way to predict what influence the parts will have on each other when combined into the system. What you get, in essence, is a synchronic version of the many-body problem. The simplest way to figure out what the behavior of each of the parts will be, even for a god, is to arrange them into the system and see the system in action. As a result, the simplest way to figure out what the features of the system as a whole are, even for a god, is to build it (either conceptually, or physically).

Although Boogerd et al. claim to be offering a version of strong emergence[23], what these examples of unpredictability really demonstrate is that science, is by pragmatic necessity, both bottom-up and top-down. Since many emergent phenomena are in principle unpredictable, they are in practice indeducible. The only practical way that limited beings such as ourselves can come to know about them is to observe them in systems already operative in the world. What often gets overlooked in discussions of emergence within philosophy of science (or perhaps it is merely taken for granted), is that *although science is by necessity both top-down and bottom-up, metaphysics is not*: there are no spooky, sui-generis, fundamental emergent properties, entities, or laws that cannot in principle, be accounted for in terms of the organized interactions of

[22] The claim that the behaviour of chaotic systems is in principle deducible is likely to raise some eyebrows, so it is important to elaborate what I take 'in principle deducibility' to mean. Firstly, in principle deducibility is not constrained by the computational and measurement limitations of actual scientific research. Secondly, it doesn't rule out the use of models. As far as I am concerned, to say that a phenomenon is 'in principle deducible', is to say that it follows from a perfectly precise model. Obtaining a perfectly precise model may, for pragmatic reasons be impossible, but this does not undermine in principle deducibility. Quantum indeterminacy complicates matters somewhat, but I won't explore those complications here.

[23] Boogerd et al. distinguish between what they call vertical and horizontal emergence. This distinction is roughly analogous to the distinction between emergence as 'inexplicability' and emergence as indeducibility. They claim to be arguing for the latter. However, the example from cell biology that they examine, and the conclusion that they come to, is not that certain system features are synchronically indeducible, but that they are synchronically unpredictable. This is an interesting finding to be sure, but not a version of strong emergence.

fundamental entities in context.

It is common within the literature to think that at least one of these notions—*unpredictability, indeducibility, or inexplicability*—is an essential feature of a theory of emergence. I, however, do not share this intuition. The key insight of emergence is not that certain systemic features are difficult to understand, but rather, that certain systemic features are genuinely novel (and causally potent) features of the world. The central idea of emergence is that when things become organized in particular ways they can constitute wholes with entirely new *kinds* of properties and capacities. Whether these are inexplicable, indeducible, or unpredictable is secondary.[24] Mario Bunge makes this point in relation to the idea that emergent phenomena must be inexplicable: he argues that "explained emergence is still emergence" (2003, p. 21). I am inclined to extend this. On my view, not only is explicable emergence still emergence, deducible emergence is still emergence, and predictable emergence is still emergence as well.

At this stage, it would not be unfair to complain that the conception of emergence that I am advocating is so far removed from its historical roots in the British Emergentism as to be undeserving of the name. However, it does adhere to the deeper historical roots in the Ancient Greeks.[25] Moreover, there are several contemporary thinkers employing similar conceptions of emergence, see for example, Mario Bunge (2003) and William Wimsatt (2000). Emergent materialism, as I shall defend it, is the idea that all of the various macroscopic phenomena in the world can be accounted for in much the same way as a body of water can be accounted for in terms of the organized interactions of a set of lower-level active entities. Of course, with more interesting and complicated phenomena, the kinds of interactions that occur between the parts will be considerably more sophisticated (involving for example, amplification effects due to positive feedback loops, and stability resulting from negative feedback loops) but, there is nothing fundamentally different. Naturally, more needs to be said about how to understand this relation. Before doing so however, it is worth taking the time to reject a naive conception of emergence

[24] Notice that I specify new *kinds* of properties. I do not consider what van Gulick calls 'specific value emergence' to count as emergence (2001). An example of specific value emergence is as follows. A 1kg lump of lead instantiates the property of 'being 1kg', which is not a property instantiated by any of its parts. This is not an emergent property on my view for it is not a new *kind* of property: each of the lump of lead's parts instantiates some value of mass.
[25] For example see Galen's "On the Elements According to Hippocrates", 1.3, 70.15-74.23.

that is sometimes conflated with the one I will present here.

2.2 Naïve Secretion Emergence

I mentioned in the introduction that the concept of emergence has a somewhat dubious reputation among philosophers. While much of the responsibility for this lies with the spooky notions of emergence involving *in principle indeducible* or *in principle inexplicable* phenomena, at least part of the blame needs to be reserved for the thought that emergence is the naïve idea that when entities become organized into a system they somehow *cause*, *emit*, or *secrete* a new, causally potent, feature of the world. I call this notion of emergence *secretion emergence*.

Patricia Churchland provides a particularly memorable example of this sort of thinking in her critique of a well-known cookbook author.

> In her microwave oven cookbook, Betty Crocker offers to explain how a microwave oven works. She says that when you turn the oven on, the microwaves excite the water molecules in the food, causing them to move faster and faster. Does she, as any high school science teacher knows she should, end the explanation here, perhaps noting, "increased temperature just *is* increased kinetic energy of the constituent molecules"? She does not. She goes on to explain that because the molecules move faster, they bump into each other more often, which increases the friction between molecules, and, as we all know, friction causes heat. *Betty Crocker still thinks heat is something other than molecular KE; something caused by but actually independent of molecular motion.* (Churchland, 1994, p. 30)

In other words, Betty Crocker thinks that the molecular motion, or the friction generated by molecules bumping into each other, emits, or secretes 'heat': a distinct *thing*. And that it is this, 'heat', and not the increase in molecular kinetic energy, that is responsible for the cooking of our food. The problem with thinking of heat in this way is obvious; it postulates something that has no causal role to play. As science has confirmed, it is the increase in molecular kinetic energy that cooks our food. If heat, as Betty thinks, is something secreted by but actually independent from molecular motion, then there is no role left for it to play in the cooking process.

However, when we turn to the relationship between conscious experience and the processes that

take place in functioning brains, the temptation towards secretion emergence is hard to resist. It fits very neatly with our pre-theoretical conceptions about consciousness as being fundamentally different from neuronal processes and, so long as you don't look too closely it gives the impression of paying homage to the advances of science by acknowledging the important role that brain processes play. Furthermore, graphical depictions of the mind-brain relationship lend themselves to being interpreted as involving secretion emergence. Consider the image below, taken from Annti Revonsuo's 2010 book "Consciousness: The Science of Subjectivity". Although it is clear from the caption and surrounding context that Revonsuo is not endorsing anything like secretion emergence, it is natural to interpret the image as suggesting that consciousness is something emitted by activity in the brain, but, despite this, somehow floats free of its neural underpinning.[26]

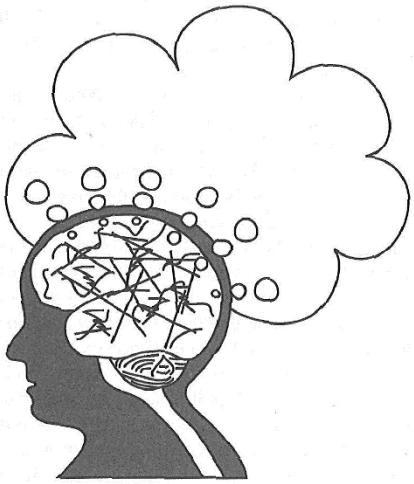

Figure 2 An image from Revonsuo (2010) intended to represent the 'emergent materialist' picture of consciousness. He captions the image as follow. "When brain activities reach a high degree of complexity, a higher level of physical reality—consciousness—emerges. The higher level cannot be reduced to traditional neurophysiology, because it has higher level features (such as qualia) not present in any lower level neurophysiological systems. Still, even the higher level of consciousness is a purely physical phenomenon and a part of the material world. It is unclear whether the emergence of the higher level of consciousness can be explained by studying the brain. According to weak emergent materialism, explanation is possible. However, according to strong emergent materialism, we will never understand how the higher level of reality comes about from the brain." (Revonsuo 2010, p. 27).

[26] It is also quite easy to make the same mistake when considering Carl Craver's representation of the levels of mechanism *(fig 1.)*. Given the way it is depicted, it is quite natural to interpret Sψ-ing as belonging to a higher level than the organized collection of X's φ-ing. This however is a mistake. The organized collection of X's φ-ing is just a higher-resolution depiction of Sψ-ing. They both belong to the same level of mechanism. As Opie and O'Brien (2015) emphasize, what belong at each level are the *individual* active entities whose organized interaction collectively exhibit the phenomenon at the higher level. In this case, Sψ-ing, $X_1\varphi_1$-ing, and $P_1\rho_1$-ing belong to three different levels.

The problems with thinking of the relationship between conscious experience and the neural processes that underpin it in terms of secretion emergence are obvious. First and foremost it is clearly a form of dualism. As Searle puts it, if you think consciousness is somehow "squirted out by the behavior of the neurons in the brain, but once it has been squirted out, it has a life of its own" then you are back in bed with Descartes (Searle, 1992, p. 112). Furthermore, it is subject to the same worries about causal exclusion as those addressed section 1.3. Unless you are willing to give up the idea that the material world is causally closed—which, as I argued there is not easily given up—then secretion emergence commits you to epiphenomenalism; it leaves nothing for conscious experiences to do. In the words of noted British Emergentist, Samuel Alexander, thinking of consciousness this way "supposes something to exist in nature which has nothing to do, no purpose to serve, a species of *noblesse* which depends on the work of its inferiors, and is kept for show and might as well, and undoubtedly would in time be abolished" (Alexander, 1920, p. 8 vol. 2).

Antti Revonsuo has argued that much of the intuitive drive towards thinking of the relationship between conscious experience and the underlying neural processes in terms of secretion emergence stems from the misleading way we have asked the question. It is common to ask. "How does a nonconscious phenomenon, such as a neuron firing, produce or become a conscious phenomenon?" But this question is misguided. Just as individual molecules of H_2O do not mysteriously *become* liquid or unaccountably *emit* liquidity, so too, nonconscious entities do not mysteriously *become* conscious, or unaccountably *emit* consciousness. Rather, at one level of mechanism there are nonconscious entities that, when appropriately interrelated, form a higher-level entity capable of having conscious experiences. As Revonsuo puts it:

> … nonconscious phenomena do not mysteriously emit consciousness, they collectively constitute it. To ask, "How does a nonconscious phenomenon produce or become a conscious phenomenon?" is like asking, "How does a subatomic particle become an atom?" or "How does a water molecule emit liquidity?" or "How does a DNA molecule become alive?". The question ascribes a higher-level feature to a lower-level entity, which in none of the cases makes any sense." (Revonsuo, 2006, p. 358)

Secretion emergence inevitably leads to a form of dualism (or a radical pluralism). As such, it is obviously not suitable to serve as the relation between levels of mechanism in a materialist picture of the world. Any emergence concept that will be useful for formulating materialism must maintain what Carl Craver has called, a sense of "ontological intimacy" between the emergent phenomenon and the organized collection of parts and activities which constitute it (Craver, 2015b). Craver is cautions about how to unpack this notion of "ontological intimacy". He says, that ontological intimacy "is meant to denote an exhaustive ontological grounding of the behavior of a mechanism as a whole in the organized interactions of its components in a given causal context, however that is properly to be unpacked" (Craver, 2015b). I am inclined to be a little more ambitious. The reason an emergent phenomenon is ontologically intimate with the organized interaction of parts in which it is instantiated, is that they are one and the same thing: they are token identical.

2.3 Emergence as Systemic Novelty

The sense of emergence that I am interested in is the one that pervades discussions in systems theory, the biological sciences, and mechanistic philosophy. It sees emergence in any whole (or system) that has at least one property that is of a kind not had by any of its constituents.

I have already briefly discussed the example of liquidity: liquidity is a property that can be instantiated in collections of molecules, but no single molecule can instantiate liquidity. Consider a few others. A snowflake has a six-fold symmetry, while none of its constituent molecules do (Deacon, 2006, 2013). A candle flame extracts resources from its environment in order to maintain its own existence in a way that none of its parts do (Campbell, 2015). A bacterium can alter its behavior to increase its chances of being in an environment rich in the raw materials that it needs to maintain itself; none of its parts can (Bickhard, 2000). Organized networks of neurons can represent features of the world in ways that no individual neuron can (Bunge, 2003, 2010).

Emergence as systemic novelty can be characterized as a commitment to the following three theses:

1. Emergent systems (wholes) have at least one property that is of a *kind* not possessed by any of its constituents. Emergent properties then, are properties that only exist at the

level of the system as a whole. They do not apply to the systems parts taken individually.

2. The organized interactions of the systems constituents over a given period of time *non-causally determine* the local (or non-relational) properties of the whole over that period of time.[27]

3. The environment (or the wider systemic context) often plays an important and ineliminable role in determining how the parts will be organized and how they will interact. How much of an impact the environment has varies from system to system.

The best way to get a feel for this is to explore some examples. However, before doing so it is worth saying something about *non-causal determination*. As I use the term, to say that the local properties of a whole are *non-causally determined* by the organized interactions of the systems constituents, is to say to say they are metaphysically necessitated by them. It is to say that once the organized interaction of the systems constituents is fixed, so too are the local properties of the whole system. It is important to stress that non-causal determination is not intended to imply a static view of systems. Many of the interesting emergent features of systems are necessarily temporally extended. One of the instances of systemic novelty that I will look at shortly is the *self-maintenance* of a candle flame. Self-maintenance is necessarily a temporally extended phenomenon. It makes no sense to ask whether a state (or temporal instant) of a system is self-maintenant; the idea of 'continued existence over time' is implicit in the notion. Rather, non-causal determination applies to the relationship between the properties of the system as a whole (over a given time period) and the properties of its parts and their organized interactions over that time period. The most natural way to interpret this is in terms of identity. On my view, a system's local properties are determined by the organized interactions of its parts because the system and the organized set of interacting parts are one and the same thing being described in

[27] In the literature, this idea is often presented in terms of 'mereological supervenience' which holds that no two systems with the same constituents, interacting in the same organized fashion, can have different systemic properties. Since I have rejected supervenience as inadequate for the job of relating entities at various levels of mechanism it wouldn't do to introduce a version of it here. In any case, it is the stronger thesis of determination that people typically have in mind when they speak of mereological supervenience. For example, Kim defines mereological supervenience as "the doctrine that properties of wholes are *fixed* by the properties and relations that characterize their parts" (Kim, 1998, p. 18 my emphasis).

two different ways.

It is also important to note that not all properties of the whole are fixed by its constituents and their organized interaction. Certain properties are non-local (or relational) meaning that they depend on how the system is related to various features in the environment. As I mentioned in the previous chapter, two fish that are identical in terms of the organized interaction of their constituents can nonetheless differ in terms of evolutionary *fitness*. If one is happily swimming around a coral reef rich, both in the resources it needs to survive and in potential mates, and well equipped to evade the local threats (sharks say), it will be considerably *fitter* than an identical fish in a live tank at a sushi restaurant. The capacities that the fish are endowed with by virtue of the organized interaction of their constituents may allow the former to evade sharks and win a mate, but they will not do much to protect the latter from the knife and the keen eye of a well-trained sushi chef. For an even simpler example consider the one provided by Teller (1992). The property of being 'the longest pencil in the box' depends on more than just the organized interaction of the pencils constituents. Being the longest pencil in the box depends also on the other pencils. For this reason, it is only the local (or non-relational) properties of a system that are non-causally determined by the organized interaction of its constituents.

Consider how these three criteria apply to a body of liquid water—say, the contents of the glass of water on my desk. The first criterion is satisfied since the body of water taken as an organized whole instantiates at least one property, liquidity, that is of a kind not had by any of its constituents taken individually. As we saw earlier, liquidity is not the kind of property that an individual molecule can possess. The second criterion is satisfied since the collection of molecules, together with the way they are organized and how they interact, non-causally determines that the contents of the glass is liquid; it is metaphysically impossible for something with the same constituents, interacting in the same organized fashion, to fail to instantiate liquidity. The third criterion is satisfied since the temperature of the surrounding environment effects how the molecules are organized and how they interact. When the glass of water is at room temperature, the kinetic energy of the individual molecules is sufficient to stretch and even break the hydrogen bonds that hold the molecules together. However, it is not sufficient to allow the molecules to escape entirely, and new hydrogen bonds are quickly reestablished. As a result,

molecules that constitute the body of water are loosely bound via the continual stretching breaking and reforming of weak hydrogen bonds, and this determines that the body of water will be liquid. Were you to put the glass of liquid water into the freezer however, as the individual molecules became less energetic, they would eventually be unable to overcome the strength of the hydrogen bonds and the individual molecules would become bound together into a ridged lattice structure. As a result, the contents of the glass would no longer be a liquid water; rather, it would be solid ice. To rehearse this succinctly, systems composed of interacting water molecules instantiate at least one emergent property (either liquidity, solidity, or being a gas). Which of these properties they possess is determined by the way in which their constituent molecules interact and how they are organized. But this, in turn, is influenced by the surrounding environment.

To take another example consider the phenomenon of self-maintenance as it is instantiated in a candle flame. A candle flame consists of a number of parts—oxygen, paraffin wax, a wick, carbon dioxide—and activities—vaporizing, combusting—with a particular organization. What makes the candle flame an interesting case from the perspective of emergence is that the system as a whole alters the environment so as to encourage its continued existence; it is a self-maintenant system. The self-maintenance of a candle flame is achieved via the following mechanism.

The combustion of oxygen and vaporized paraffin wax generates heat which both: 1) melts, and then vaporizes more of the paraffin wax of which the candle is made, and 2) generates a convection current, drawing in fresh oxygen and expelling the spent carbon dioxide produced by its own combustion.

However, there is no mysterious downward causation going on here. This is merely a more sophisticated (although still relatively simple) example of emergence as systemic novelty. The candle flame system as a whole exhibits at least one novel property that is of a kind not possessed by any of its constituents (in this case the one that I am interested in is the property of 'being self-maintenant'). This property is non-causally determined by the organized interactions of the oxygen, paraffin, wick, and carbon dioxide that constitute the candle flame system. Thus criteria 1 and 2 from above are met. In addition, the environment plays an ineliminable role in

determining how these parts are organized and how they interact. For example, the concentration of oxygen in the surrounding environment impacts the intensity with which a candle flame burns, and even whether it will burn at all. Due to the lower concentration of oxygen in the atmosphere at high altitudes, a candle flame at the top of Mount Everest will burn considerably dimmer (and longer assuming it isn't blown out by the wind) than one at sea level. More interesting for present purposes however, once the concentration of oxygen in the surrounding environment drops below a certain threshold, the flame will no longer be able to acquire sufficient fuel in order to continue burning; which is why, a candle flame burning in a small enclosed environment (under a glass jar say), will soon extinguish itself. The temperature of the surrounding environment is also crucial to the flames ability to sustain its own existence. If the temperature in the environment surrounding the candle is 1000°C (roughly the temperature of the candle flame), the combustion of oxygen and paraffin will no longer generate the convection current that draws in fresh oxygen and expels the waste carbon dioxide. Again, the flame will no longer be able to sustain its existence and will cease being a self-maintenant system. The self-maintenant candle flame provides another example of a system which possesses a novel property by virtue of the organized interaction of its constituents, but which is in turn influenced by the environment.

2.4 Emergent Causation

There has been a lot of discussion over the years about the possibility of emergent causation. While it is impossible to do justice to all that has been said, a lot of progress can be made by simply being clear about which notion of levels is in play. There are two issues to address: one is whether or not higher-level systems can be causally potent at all, and the other is whether they exhibit any *downward* causal influence on their constituents. I will address these in turn.

The causal argument that we looked at in the previous chapter has often been thought to threaten the possibility of causally efficacious entities existing above the level that characterizes the world's most basic parts. For example, employing something like the Oppenheim/Putnam notion of 'levels of science', Ned Block worries that the causal exclusion argument shows that all macro-level causation might "drain away" to the level of fundamental microphysics (Block, 2003). However, it is important to note that these causal drainage worries do not arise for entities at

higher levels of mechanism. Jaegwon Kim, the author of the causal exclusion argument, is quite explicit that his argument does not rule out the possibility of wholes possessing novel causal powers by virtue of the organized interactions of their parts. That they do, Kim takes to be an obvious but important point to keep in mind when thinking about causal exclusion. As he puts it:

> H_2O molecules have causal powers that no oxygen and hydrogen atoms have. A neural assembly consisting of many thousands of neurons will have properties whose causal powers go beyond the causal powers of the properties of its constituent neurons, or subassemblies, and human beings have causal powers that none of our individual organs have. Clearly then *macroproperties can, and in general do, have their own causal powers, powers that go beyond the causal powers of their microconstituents*. (Kim 1998, p. 85)

Craver considers the of existence higher-level entities with higher-level causal capacities to be "so obvious, so prosaic, and so banal as to be hardly worth mentioning". As he rightly points out, even extremely simple systems can do things that their components alone cannot: "two toothpicks stacked perpendicular to one another have the mechanistically emergent capacity to act as a lever or catapult; neither toothpick can do so on its own" (Craver, 2015a, p. 20). Obvious? Yes. But since so much ink has been spilt arguing against it, it is, regrettably, important to say something about it.

Recall from chapter 1 the distinction between 'levels of mechanism' and 'orders' (or 'levels of realization'). Where 'levels of mechanism' refers to the relationship between a mechanism as a whole and the individual active entities that constitute it, 'orders' refers to the relationship between properties; higher-order properties are properties that a thing possesses by virtue of possessing some other property. Recall the example of my pumping heart. Levels of mechanism refers to the relationship between my pumping heart as a whole, and the individual active entities that constitute it. The orders relation on the other hand, refers to the relationship between being *H* (recall that this is the property of being 'two ventricles and two atria, connected by flow directing valves, contracting and relaxing in a particular orchestrated manner') and being a pump. The mechanism in my chest (my pumping heart) instantiates *H*, and by virtue of instantiating *H,* it also instantiates the property of being a pump. Being a pump is higher-order than being *H.*

What is crucially important to remember is that the orders relation does not cross mechanistic levels. Being a pump and being *H* are both properties of the same mechanism (the thing in my chest) and hence are both at the same mechanistic level.

What Kim's causal exclusion argument shows is that higher-order properties do not imbue the bearer with any causal powers over and above those that it already possessed by virtue of its lower-order properties. For example, the higher-order property of 'being a pump' does not imbue my heart with any causal capacities that it doesn't already possess by virtue of being 'two ventricles and two atria, connected by flow directing valves, contracting and relaxing in a particular orchestrated manner': that is, by virtue of being *H*. Kim's argument shows that in any given instance the causal powers of higher-order properties drain down to those of their lower-order realizer.[28] The important thing to note here is that causal drainage worries do not arise for entities belonging to higher levels of mechanism because the orders relation does not cross mechanistic levels. What belongs at the lower mechanistic level is the (unorganized) set of individual active entities that collectively constitute the higher-level mechanism when appropriately organized. Causal drainage worries do not arise for entities at higher-level of mechanism because organization has an ineliminable role to play in determining the properties and causal capacities of systems, and lower levels of mechanism lack that organization. Any attempt to account for the causal capacities of a mechanism solely in terms of the causal capacities of its constituent parts, individually or as an unorganized group, is bound to fail.[29]

[28] It is important to note that this does not prohibit the use of higher-order properties in causal explanations. Being 'below 0°C' is a higher-order property since there are many ways of being below 0°C (e.g., being -3°C, being -13°C and -113°C). Nonetheless, citing that the ambient temperature is below 0°C is a perfectly satisfactory causal explanation as to why a glass of water froze rather than remaining liquid; however, it would not be a satisfactory causal explanation for why the water froze at the particular rate that it did (see Craver 2007 chapter 6 for a detailed discussion). The reason this argument has had such an impact on the literature is that it showed that multiply realizable property types are not causally potent over and above their token realizations. This made things awkward for those who held that the properties picked out by the special sciences are radically multiply realizable (i.e., could be realized in completely different mechanisms), and worse still for those who denied that higher-order properties were token identical to their lower-order realizers. If token identity does not hold between realized and realizing properties, then it is hard to see what causal role is left for the realized property.

[29] Kim is again quite explicit about this, although his terminology is different. In his words: "[Causal drainage worries] do not arise for micro-based properties in relation to their constituent properties because the former do not supervene on the latter taken individually or as a group. Rather, they supervene on specific mereological configurations involving these microproperties—for the obvious and uninteresting reason: they are identical with these micro-configurations" (Kim 1998:118).

More controversial than the existence of causally potent wholes are the debates about whether wholes have any downward causal influence on the behaviour of their parts, or alternatively, whether parts have any upward causal influence on the whole they compose. Inter-level causation, both bottom-up and top-down, is problematic on traditional conceptions of causation for a number of reasons.[30] First, causation is typically construed as occurring between two (or more) distinct objects (e.g., a rock and a window, a cue ball and the 8, a depolarizing neuron and the neurons it synapses on). Second, according to traditional conceptions of causation, a cause must precede its effect. Neither of these criteria are met in the case of inter-level causation. A system and its constituents are neither wholly distinct objects, nor are they temporally separate.

On the account of emergence that I have provided here, while there are certainly causally potent phenomena at higher-levels of mechanism, there is neither top-down, nor bottom-up causation. As Craver and Bechtel have argued (2007, 2013), all apparent inter-level causal influences can be understood as the combination of two non-mysterious metaphysical relations: the ordinary intra-level causal interactions among components in a mechanism, and the constitution relation that holds between the organized collection of active components and the behavior of the mechanism (system) as a whole.

Consider the relationship between my functioning circulatory system as a whole and my pumping heart. My pumping heart is one of the constituents of the mechanism that circulates blood throughout the body. In order to perform its activity (pumping), the heart requires a steady stream of oxygen rich blood. What provides the heart with this steady stream of oxygen rich blood? Well the functioning circulatory system: the very mechanism of which it is a part. But to think of this as an instance of downward causation would be a mistake. The steady supply of oxygen that the heart receives can be understood, without loss, in terms of the ordinary intra-level causal interactions between the various components of the functioning circulatory system: the heart, veins, lungs, and blood. Very loosely, oxygen depleted blood is pumped from the heart to the lungs where it is re-oxygenated. The blood then flows back the heart where most of it is

[30] Inter-level causation is less problematic for more recent conceptions of causation such as Woodward's causal relevance account (Woodward 2003, Craver & Bechtel 2013).

pumped back out to the rest of the body, but a small amount is sequestered in order to sustain the pumping of the heart (see *figure 3* in the next chapter and the discussion there for a slightly more detailed account). Of course, any change to a higher-level mechanism of which the heart is a constituent can have an impact on the behaviour of the heart, but again this should not be thought of as downward causation. Suppose I were to begin a strenuous hike up a mountain. Very quickly, sensory receptors monitoring the position of my limbs and muscles (proprioceptors) would send signals to the cardiovascular center in my brain. The cardiovascular center would in turn signal for my heart to pump faster. Why did my heart being pumping faster? Because I started hiking up a mountain. But there is no downward causation here. As Craver and Bechtel put it, this is just a case in which "a change in the activity of the mechanism as a whole *just is* a change in one or more components of the mechanism which then, through ordinary intra-level causation, cause changes in the other components of the mechanism" (Craver & Bechtel, 2007, p.559 my emphasis). In this case, the higher-level mechanism in question is not just the functioning circulatory system, but the functioning circulatory system together with the relevant aspects of the autonomic nervous system that are responsible for regulating the flow of blood throughout my body. My beginning to hike was in part constituted by increased muscle activity, which via straightforwardly intra-level causal interactions, increases the rate at which my heart pumps.[31]

The take home message is this: while the account of emergence that I have advanced in this chapter—emergence as systemic novelty—certainly endorses the existence of causally potent systems and causal interactions taking place at higher-levels of mechanism, it does not license claims of downward causation.

2.5 Summary

The view of the world that we get according to emergent materialism as outlined in the last two

[31] The details of the autonomic regulation of heart rate are obviously considerably more complicated than this. A full account will require considering causal interactions at many levels of mechanism lower than the circulatory system and autonomic nervous system as a whole. For example, a full account will require considering how impulses in the sympathetic cardiac accelerator nerves that synapse onto the sinoatrial node in my heart trigger the release of norepinephrine, which binds with the *beta-1* receptors, accelerating the rate at which the heart's pacemaker cells spontaneously depolarize (Tortora & Derrickson, 2012). Nonetheless, in each case we have nothing more than ordinary intra-level causation and the constitution relation.

chapters is that although only the fundamental posits of physics are instantiated in a basic way, when these become organized they can form complex wholes with novel properties and novel causal capacities which, in turn, can constitute higher-level wholes with further novel capacities. Emergent materialism is the view that all of nature can be accounted for in this way. The major challenge for this view is consciousness. Emergent materialism is committed to the view that the mind is related to the brain (or the brain-body-environment system) in the same kind of way that liquidity is related to a body of weakly bound H_2O molecules, or self-maintenance is related to a candle-flame. As we will see in the next chapter, when it comes to consciousness, this view faces a serious challenge.

Chapter 3: The Problem of Consciousness

The picture that I have presented so far sees all non-microscopic features of the world as explicably emergent features of material systems. On this view, conscious experience is an explicably emergent feature of specialized systems within normally functioning brains (or perhaps brain-body-environment systems), not radically dissimilar from other explicably emergent phenomena.[32] This view has a lot in its favor. It allows us to take consciousness seriously, as a real, causally efficacious feature of the world, without compromising the view that the world is ultimately constructed from a relatively small set of fundamentals building blocks. The major challenge facing this view comes from what Levine calls the "explanatory gap" (Levine, 1983, 2001). All of the emergent phenomena we have discussed so far (e.g., liquid bodies of water and self-maintaining candle flames) can ultimately be accounted for in terms of the organized interactions of a set of parts plus the relevant environmental conditions. The emergence of consciousness however, continues to stubbornly resist such explanation. There appears to be a deep epistemic gap between the objectively observable processes that take place in material systems on the one hand, and subjective conscious experience on the other. Moreover, a number of arguments have been developed that purport to show that this explanatory gap is not merely an *epistemic gap* that will one day be closed by advances in science, but represents a deep *ontological gap* between consciousness and the material world and ultimately demonstrates the inadequacy of materialism.

In the next two chapters I will address the challenge of bridging this gap. My goal here is not to provide a definitive solution. Following Stoljar (2006), I think definitive solutions are beyond what can be achieved with our current conceptual and empirical tools. Rather, I will seek to bolster the materialist picture by arguing that the problem is not as dire as it first appears. I will do this by emphasizing the distinction between *experience in general* (and the challenge of explaining why certain neural processes are accompanied by subjective experience), and *qualitative character* (and the challenge of explaining why any particular experience *feels* the way that it does) on the other. This distinction is not new—I provide three examples of it from the literature on

[32] Whether non-neural systems can also exhibit these features is an open question however the majority view is that they can.

consciousness in section 4.3. However, it has implications that have been overlooked. I argue that it is only the existence of *experience in general* that provides a metaphysical challenge for materialism. Even if materialism is incapable of accounting for the qualitative character of experience—of accounting for what experience *feel* like from a subjective perspective—if the existence of experience in general can be accounted for, materialism is vindicated. Furthermore, I argue that there is reason to be optimistic with regard to the prospects of accounting for the existence of *experience in general*. If *experience in general* is not itself something we experience, then there is room for a conceptual renovation that will allow us to understand *experience in general* in terms of the kinds of processes that take place in certain material systems.

Before elaborating on this however, it is important to be clear about what the problem is and why it is so hard. That is the task of this chapter. In section 3.1 I introduce the problem of consciousness in more detail, and clarify why it is indeed a "hard problem". In section 3.2 I take a short detour into the philosophy of science and briefly address some concerns that the "hard problem" is merely an artifact of outdated theories of how explanation works in the sciences. I show that the underlying philosophical issues surrounding consciousness re-emerge with just as much potency when more modern views of explanation are deployed.

3.1 The Hard Problem of Consciousness

When discussing the ideas in this chapter with friends and colleagues who are unfamiliar with traditional philosophical issues regarding the mind, I have found that presenting them with questions such as "Why is it the case that when my amygdala is active in a particular way, I feel fear?" or "Why do certain patterns of neural activity produce in me experiences of pain?" or even "Why is it the case that when these things happen I experience anything at all?" I typically receive a response along the following lines: "Conscious experiences provide an evolutionary benefit. Therefore, it is perfectly understandable why such traits evolved. For example," they continue, "the pain you feel when you hold your hand too close to the fire provides a very strong incentive to not do that again, thus preventing further damage to your hand. Similarly, the persistent pain you feel after rolling your ankle ensures that the now-weakened joint is not over-strained, and is given sufficient time to repair."

The fact that pain is valuable from an evolutionary perspective is made palpably clear by

examining cases in which it is absent. Consider the case of Miss C, a university student at McGill University who suffered from congenital analgesia (a disorder which renders sufferers incapable of feeling pain, or insensitive to pain). When psychologist Gordon McMurray examined Miss C in the laboratory, he found that not even subjecting her to stimuli which hare normally considered to be forms of torture could cause her to feel any pain (1950). For example, pinching her Achilles tendon, injecting her with doses of histamine, and even inserting a stick up her nostril, all failed to produce any feelings of pain.

Miss C's inability to feel pain had a severely detrimental impact on her physical body, and ultimately played a central role in the events that lead to her death at the age of twenty-nine.

> As a child, she had bitten off the tip of her tongue while chewing food, and has suffered third-degree burns after kneeling on a hot radiator to look out of the window. … She exhibited pathological changes in her knees, hip and spine, and underwent several orthopedic operations. Her surgeon attributed these changes to the lack of protection to joints usually given by pain sensation. She apparently failed to shift her weight when standing, to turn over in her sleep, or to avoid certain postures, which normally prevent the inflammation of joints … All of us quite frequently stumble, fall or wrench a muscle during ordinary activity. After these trivial injuries, we limp a little or we protect the joint so that it remains unstressed during the recovery process. This resting of the damaged area is an essential part of its recovery. But those who feel no pain go on using the joint, adding insult to injury. (Melzack & Wall, 1988, pp. 4-5)[33]

In Miss C's case there is little doubt that her inability to feel pain was responsible for the "extensive skin and bone trauma that contributed" to the massive infection that led to her death (Baxter & Olszewski, 1960).

On the face of it, there is something right about this response to my question "Why do certain patterns of neural activity produce in me experiences of pain?" Pain most certainly does provide an evolutionary benefit, and citing this fact does provide some form of explanation as to why we

[33] For more on the evolutionary benefit of pain see (Grahek & Dennett, 2011).

experience pain. However, there are considerable problems with this response.

The first thing to point out in response to these considerations is that it is not the *experiences* of pain as such that conveys the evolutionary benefit, it is the associated tendency to avoid situations which cause physical harm that they engender. As certain philosophers may want to point out, it is perfectly conceivable that we may have evolved each of these tendencies without the associated experiences of pain or fear. More problematically however, this response misses the deeper issue. When we ask "Why do certain patterns of neural activity produce in me experiences of pain?" we are not asking for a causal, or historical story as to why we have experiences of pain. Rather we are asking, "How could it possibly be the case that patterns of neural activity in my brain is all there is to my experience of fear?" The hard problem is not to explain why organisms with complex neural circuitry capable of having subjective experience evolved, but to explain how it could possibly be the case that subjective experience is an emergent feature of neuronal systems in the first place. To introduce some philosophical technicalia, what we need is a *reductive* or *constitutive* explanation (I use the terms synonymously), rather than an *etiological* explanation.[34] Where an etiological explanation looks backwards and tells a causal, or historical story as to why the phenomenon in question occurred, reductive explanation looks inwards (or downwards) and describes the mechanism (the organized set of active parts) via which it works. The problem that consciousness poses is that it is not clear that a reductive explanation is possible.

A typically unstated but plausible assumption of the claim that consciousness poses a hard problem, is the thought that explanations must be *illuminating*. A good explanation of a phenomenon, be it reductive or etiological, should leave no deep mystery as to whether or not the phenomenon occurred. In the case of reductive explanation, a detailed description of the internal workings of the system in which the phenomenon occurs, should leave no deep mystery as to whether or not the phenomenon was exhibited in that system. And indeed, this is what we

[34] When philosophers of mind use the term 'reductive explanation' I take it that it is typically 'constitutive explanation' that they have in mind (Chalmers, 1996; Kim, 2008). Some philosophers of science however, are inclined to think of reductive explanation as closely tied to the kind of *explanation reduction* that takes place in Ernest Nagel's account of inter-theoretic reduction. I won't provide any argument for my usage here (see Kim 2008 for a detailed discussion).

find with the myriad examples of successful reductive explanations that science has thus far provided. For example, science has explained how a collection of H_2O molecules (when loosely bound together via the continual stretching, breaking, and reforming of weak hydrogen bonds) can exhibit the central tendencies of liquidity. The details need not concern us here, but it is possible to create models in which the most salient features of the inter-molecular interactions are captured, and literally *see* how, as a whole, they exhibit the central tendencies of a liquid.[35] If, after being presented with such a model, you still felt compelled to ask, "But where's the liquidity? I see how collections of molecules can constitute a whole that flows and has a relatively constant volume, but I don't see how they can constitute a liquid." Then you simply don't understand what the word 'liquid' means.

Illuminating explanations is also what we find, or at least what we expect to find, in cases where much explanatory work remains to be done. For instance, while there are still some deep unanswered questions surrounding the phenomenon of *life*—What are the precise mechanisms of epigenetics? (Carey 2012). What is the role of "junk DNA" that does not code for proteins? (Carey 2015)—we do know, or at least have exceeding good reason to believe, that explanations of these phenomena are within reach for current methods of explanation (i.e., explanations in terms of the organized interaction of a set of active parts in context). And, we also know that after such explanations are given, nothing about life will be left unexplained. After explaining how purely material mechanisms are capable of metabolizing, reproducing, self-repairing, and the like, no further explanation is required. If you still feel compelled to say, "Yes, yes, I see how a mechanism could do all that, but I still don't see how it could be alive" then you are making a conceptual mistake. All it means for something to be alive (in this sense of the word at least) is for it to be capable of doing all these things.

What a number of thinkers have pointed out is that this does not seem to be the case with conscious experience. Joseph Levine makes this point by emphasizing the challenge of explaining his subjective experience when he looks at a red diskette case that lives on his writing desk. He

[35] Note that we do not need precise, or even complete details of these interaction for the purpose of explaining how liquidity can be a property of systems constituted by non-liquid parts. However, more precision and completeness is needed if one wants to explain the particular characteristics (such as viscosity, boiling point, etc.) of a particular liquid.

says:

> As I now look at my red diskette case, I'm having a visual experience that is reddish in character. Light of a particular composition is bouncing off the diskette case and stimulating my retina in a particular way. That retinal stimulation now causes further impulses down the optic nerve, eventually causing various neural events in the visual cortex. Where in all of this can we see the events that explain my having a reddish experience? (Levine, 2001, pp. 76-77)

Setting aside objections about the extreme oversimplification of the neural events in question, and the omission of further processes that may prove necessary for consciousness[36], Levine's point is well made. Intuitively it is hard to see how an account of the neural processes, however detailed they become, could ever amount to an illuminating explanation of something like a reddish experience. Even after we have explained how Levine's retina is able to transduce light stimulus into neuronal signals, and how his visual system is capable of using those signals to produce a complex representation of the external world via which he can navigate, even after we have explained all the capacities that are associated with visual experience, it still seems reasonable to say, "yes, yes, I see how a neural system could do all that, but I still don't see how it could be an experience of a reddish diskette". There is currently a deep explanatory gap between the neural processes on the one hand and the facts about conscious experience on the other. [37]

David Chalmers has clarified this issue nicely by articulating the difference between the 'easy problems' and the 'hard problem' of consciousness (1995, 1996). Chalmers points out that there are two steps to successful reductive (constitutive) explanation: the first involves (re)conceiving the phenomenon to be explained in terms of a set of capacities or functions, the second step involves explaining how those capacities or functions can be exhibited by the organized interaction of lower-level parts and activities. When we already conceive of the higher-level phenomenon as a set of capacities or functions, the first step is quite straightforward. This is

[36] For example, neural synchrony (Tononi & Koch, 2008), stability (Opie & O'Brien, 1999), or recurrent neural activity (Lamme, 2004).

[37] Much of this presentation of the problem is drawn from Chalmers (1995).

indeed the case with the example of liquidity above. To instantiate the property of liquidity *just is* to behave in particular ways: to flow, and have a relatively constant volume. Sometimes however, this first step is not so easy. Indeed it may take considerable scientific investigation before such a conception of the phenomenon is arrived at.[38] Once a conception of the phenomenon in terms of a set of capacities and functions is arrived at, scientists can begin to uncover the sorts of mechanisms capable of performing them.

This two-step process is also clearly presented in Craver and Darden's (2013) book length treatment of the search for explanations in biology and the neurosciences. They explicitly state that the first stage in discovering mechanisms in the world is to adequately characterize the phenomenon you seek to explain. This requires "developing a more or less precise description of the *behavior* or *product*" in need of explanation (2013, p. 8 emphasis added). It is only once this is complete that we can begin to construct hypotheses about the kind of mechanisms that could be responsible for it.[39]

There is however, an important difference between Chalmers' characterization of this process and the one presented by Craver and Darden. While Chalmers presents these steps as occurring in a linear, stepwise fashion—first comes conceptual analysis, then science—Craver and Darden emphasize that in practice this is rarely the case. Rather, these two steps proceed in parallel and with mutual interaction. As we experiment on, and learn more about the underlying mechanisms, our conception of the phenomenon evolves, which in turn leads to more fine-tuned experimentation. Thus, the search for mechanisms is a virtuous cycle in which experimentation

[38] 'Life' might provide an interesting case study here. For example, Bruce Webber has suggested that both Hogben, and Haldane, two influential figures in the vitalist/mechanist debate in the 1930's, considered consciousness to be an "integral part of the problem of life" (Weber, 2015). Although, according to Webber, neither Hogben nor Haldane were vitalists, if the conception of life being employed by the vitalists included consciousness under its rubric, then it is no wonder that they considered mechanisms incapable of accounting for it. Peter Godfrey-Smith makes a related point in a forthcoming paper. He notes that "there has been a partial deflation of the concept of life". Our current conception of life is "of a sort that removes any appearance of a large-scale problem that might motivate vitalism" (Godfrey-Smith, forthcoming).

[39] Craver and Darden consider this second step to itself be a two-step process: first we develop a host of mechanism schemas or "how-possibly" explanations, and then we evaluate these for against the real-world phenomenon. For simplicity I have considered these as a single step.

and conceptual analysis mutually reinforce one another.[40]

For Chalmers, the easy problems of consciousness, such as explaining how neural systems are capable of learning, or how neural systems are capable of reporting on inner states, are easy precisely because we already conceive of them as sets of capacities. Obviously neither of these problems are actually easy. There is still a huge amount about the mechanisms responsible for these capacities that we do not understand. But, Chalmers argues, since we already conceive of these phenomena in terms of sets of capacities, all that is required to explain them is to describe the mechanisms that exhibit those capacities. For example, learning is just the ability to adjust one's responses to input in light of environmental feedback. As such, in order to explain learning we just need to explain how a set of parts could be organized so as to constitute a mechanism capable of adjusting its responses to inputs in light of environmental feedback (Chalmers, 1995).[41] Similarly with 'report on inner states'. All that is required of such an explanation is to describe how a system could have access to, and report on, its own internal operations.

One way of reading Chalmers, is as suggesting that the problem of explaining consciousness experience is 'hard' precisely because our folk conception of what consciousness is, does not consist in a set of structural or functional capacities. Rather, our folk conception of conscious experience refers to something like the presence of a subjective perspective, or of it being like something to be you to use Nagel's famous phrase (Nagel, 1974). Since our conception of what conscious experience *is*, does not consist in a set of objectively observable capacities, even after one has completely described all of the capacities that fully functional human brains possess, there remains a further question to ask: why is conscious experience associated with those capacities?

To return to Levine and his red diskette, even after we have explained how Levine's visual system is capable of distinguishing various features of the world via the quality of light hitting his retina and the subsequent neural processing, there remains a further question: Why is all this

[40] For those familiar with the literature, I am here denying Chalmers and Jacksons' claim that conceptual analysis is a priori endeavour (Chalmers & Jackson, 2001).

[41] Explaining more sophisticated instances of learning, such as language acquisition in human children will undoubtable be more complicated but would still count as an easy problem is Chalmers terms since all that is required is to uncover the relevant mechanisms via which human children acquire the capacity to use language.

accompanied by Levine's having a reddish experience?

3.2 Do Developments in Philosophy of Science Help?

These philosophical worries about 'explanatory gaps' and 'hard problems' are, at times, received with a groan of impatience from those working in the philosophy of science. As Peter Godfrey-Smith colourfully puts it:

> Late at night in the bar at the Philosophy of Science Association meetings, one might hear grumbling: ''People in metaphysics and philosophy of mind have such an antiquated view of philosophy of science!'' (Godfrey-Smith, 2008, p. 62)

The antiquated view of philosophy of science being referred to is the package of ideas that surround the Deductive-Nomological model of explanation (hereafter, the D-N model), and the model of inter-theoretic reduction that it spawned (Nagelian reduction), both of which were dominant mid-century.[42] According to this package of ideas, to explain a phenomenon is to logically deduce it from a description of initial conditions together with the relevant universal laws which govern the behavior of the things involved. Furthermore, it was typically held that these laws could be deduced from the laws that govern fundamental microphysics, together with the requisite 'bridge laws' linking the vocabularies of the various sciences.

In the 1970's, both the D-N model of explanation and Nagelian reduction were called into serious question. Several thinkers, most memorably Sylvain Bromberger (1966), pointed out that certain explanations that satisfy the D-N model fail to be explanatory on the grounds that they fail to cite the requisite causal relations. For example, according to the D-N model, it is just as legitimate to explain the height of a flagpole by citing the position of the sun, the length of the shadow it casts, and the law that states that light propagates in straight lines, as it is to explain the length of the shadow by citing the height of the flagpole. The logical deduction works both ways. Despite this, few people are willing to accept that the height of a flagpole can be explained by the length of the shadow. Thus, the D-N model fails to provide sufficient conditions for explanation. Others, such as Michael Scriven (1962) and more recently those exploring the nature of explanation in

[42] The D-N model of explanation was introduced by Hempel and Oppenheim (1948) and later refined by Hempel (1966). The model of inter-theoretic reduction that it spawned was most clearly presented in Nagel (1961).

the biological and cognitive sciences such as Carl Craver (2007) and William Bechtel (2008), have argued that many perfectly adequate explanations don't have the form of logical deductions involving laws. So, the D-N model also fails to provide necessary conditions for explanation.

Around the same time, Nagel's model of inter-theoretic reduction was seriously called into question by the multiple realizability arguments of Hilary Putnam (1967), and Jerry Fodor (1974). They argued that many of the features of the world that are picked out by the predicates of higher-level sciences can be realized by a radically heterogeneous set of lower-level bases. As a result, the 'bridge laws' that Nagel's account of inter-theoretic reduction required were simply not available for many of the higher-level sciences.

That some philosophers of mind are still working with this outdated model of explanation and inter-theoretic reduction is quite clear. Consider the following quote from Levine:

> … if materialism is true we have reason to expect that any phenomenon can be explained by reference to the physical laws and principles that govern nature as a whole. Adding the basic deductivist claim about explanation, it follows that we should be able to show how a description of the phenomenon to be explained can be *deduced* from an ideal explanatory text that includes all the laws of physics, together with *whatever constitutive principles are necessary to bridge the vocabulary of physics with the vocabulary within which the description of the phenomenon to be explained is couched*. (Levine, 2001, p. 76 my emphasis)

What is equally clear is that explanations in the sciences of the mind (biology, neuroscience, and cognitive science) are typically not of this form. Rather they take the form of mechanistic explanations. Mechanistic explanations seek to explain a phenomenon by describing how a set of active entities, when appropriately organized, can constitute a system which *exhibits* that phenomenon.

Clearly the grumblings of the philosophers of science in Godfrey-Smith's tale are not unwarranted. However, as he continues:

> … the people in metaphysics and philosophy of mind are well within their rights to march

> into the bar and reply: ‘‘What *difference* does it make, to the truly foundational issues? If I fussily re-express everything in the language of the philosophy of science *du jour*, will the issues be much altered, or will they reappear more or less as before?’’ (Godfrey-Smith, 2008, p. 62)

In the rest of this section, I will briefly discuss the burgeoning mechanistic model of explanation, and note that, at least with respect to the problem of consciousness, the philosophers of mind and metaphysicians are right: the hard problem of consciousness re-emerges more or less unchanged.

3.2.1 Mechanistic Explanation

The model of explanation that is most relevant to the problem of explaining how conscious experiences are related to brains (or brain-body-environment systems) is the mechanistic model of explanation as expounded by Bechtel, Craver, Machemar, Darden, and others. These thinkers have pointed out that, contrary to what the D-N model says, scientists working in biology and the neurosciences typically do not seek to subsume phenomena under laws. Rather, they seek to uncover and describe a set of parts and activities which collectively exhibit the phenomenon. That is, they seek to uncover the underlying mechanisms via which the phenomenon works.

In abstract terms, mechanisms have been defined as follows:

> A mechanism is a structure performing a function in virtue of its component parts, component operations, and their organization. The orchestrated functioning of the mechanism is responsible for one of more phenomena. (Bechtel & Abrahamsen, 2005, p. 423)

> [A mechanism is] a set of entities and activities organized such that they exhibit the phenomenon to be explained. (Craver, 2007, p. 5)

An important point to note is the distinction between ‘mechanisms’, and ‘mechanistic explanations’. While ‘mechanisms’ are real causal systems operative in the world, ‘mechanistic explanations’ are models of mechanisms, specifically designed to illuminate their salient features and render intelligible the relationship between the mechanism, conceived of as an organized set of parts and activities, and the phenomenon they collectively exhibit.

For example, in order to explain the phenomenon of circulation as it occurs within the human body, scientists describe a mechanism operative in the human body, the functioning circulatory system, which consists of a set of parts (a heart, lungs, blood, veins) and their activities (pumping, oxygenating, carrying oxygen, directing flow), and demonstrate how the spatial and temporal organization of these parts and activities constitutes a system which exhibits the phenomenon of circulation.[43]

One of the central features of the mechanistic approach is the diversity of representational resources available for describing mechanisms. Under the D-N model, theorists are, for the most part, restricted to logical operations over linguistically or mathematically represented statements (Bechtel & Abrahamsen, 2005, p. 426). Not so under the mechanistic model. Especially within biology, scientists often deploy cartoon-like sketches, box and arrow diagrams, animated videos, graphical models, mathematical simulations, and even build scale models in order to represent the salient features of mechanisms (Craver & Darden, 2013, p. 30). For example, the mechanism of circulation within humans is often represented in cartoon-like sketches such as the one below. The arrows depict how the deoxygenated blood (represented in blue) travels from the various oxygen-using features of the body to the heart, from where it is pumped to the lungs. The change in color, from blue to red, represents that the lungs resupply the blood with the oxygen that the body needs. From the lungs, the now oxygen-rich blood flows back to the heart where it is pumped back out to be utilized by the body. Naturally, there are many details missing in this sketch of the circulatory system. For example, it provides no account of how the heart functions as a pump or how the blood can function as an oxygen carrier. Nonetheless, it presents an intelligible account of how a set of parts and activities can be organized so as to constitute a mechanism that exhibits the phenomenon of circulation.

[43] Often in the literature on mechanistic explanation, mechanisms are said to "produce", or "be productive of" the phenomena they explain (Machamer, Darden, & Craver, 2000), or else they "generate" the phenomenon (Bechtel, 2008, p. 146). I think this is an unfortunate choice of words. While it is certainly true that mechanisms can be productive, they do not produce the phenomenon they explain. For example, the mechanism of protein synthesis produces a protein, but it doesn't produce *protein synthesis*. The reason that it is important to be cautious about our talk of mechanisms producing or generating their phenomenon is that is extremely easy to misinterpret the view as a kind of secretion emergence (discussed in section 2.1).

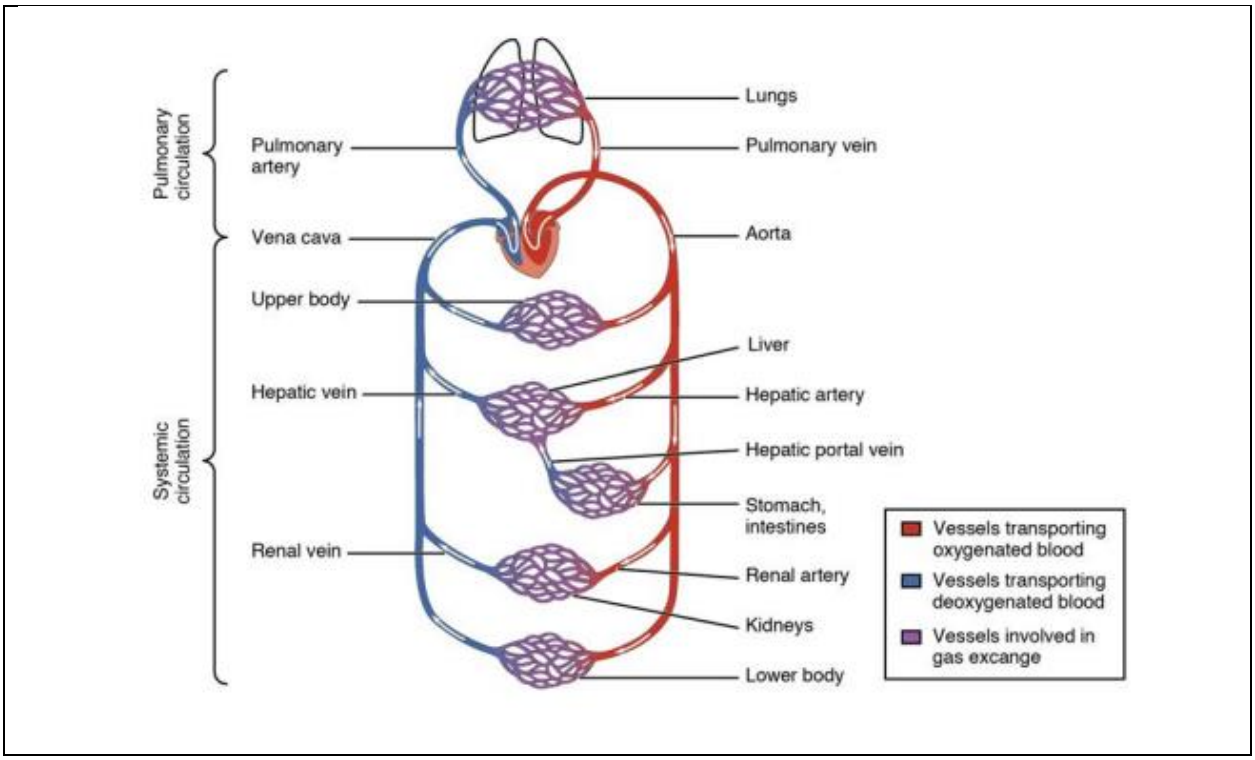

Figure 3. A mechanism sketch of the phenomenon of circulation in the human body. Image sourced from OpenStax College, 2013 p. 839.

Following Wesley Salmon (1984), mechanistic explanations are those that seek to locate the explanandum phenomenon within the causal structure of the world by describing a mechanism.[44] However, it is important to recognize that there is more to explanation than just identifying a mechanism. As Bechtel and Abrahamsen note, explanation is an epistemic activity (2005). The mechanism in nature does not by itself perform any explanatory work at all. In order to explain how collections of non-liquid molecules can constitute a liquid, it is insufficient to merely point at a glass of water. Rather the relevant parts and activities must be emphasized so as to make the link between the mechanism and the phenomenon intelligible. Likewise, the mechanist/vitalist

[44] Recently, it has been argued that dynamical modeling offers distinctively non-mechanistic (mathematical) explanations (Silberstein & Chemero, 2012). However, others consider dynamical modelling to be nothing more than a useful tool for providing mechanistic explanations (Kaplan & Craver, 2011; Craver, 2014; Kaplan, 2015). There is a clear dispute about what lies at the core of the mechanistic approach to explanation. To use a helpful distinction due to Eric Hochstein, the dispute is between those who think that mechanistic explanation is committed to providing explanations presented as *mechanistic models*—which explicitly depict a set of organized active entities (as is the case in the crude depiction of the circulatory system above)—and those who think mechanistic explanation is merely committed to providing *models of mechanisms*—models which describe the behavior of mechanisms operative in the world (Hochstein, 2016). Although I am inclined to agree with Kaplan and Craver, for the purposes of this thesis this dispute is inconsequential, since the hard problem of consciousness re-emerges with just as much potency for dynamical modelling as it does for models that explicitly depict a set of organized active entities.

debate was not brought to an end by pointing at a living organism and saying "Look! Here are the mechanisms." Rather it was resolved by *making intelligible* how a set of manifestly non-living entities can collectively constitute a living system.[45]

Precisely how to cash out this notion of 'make intelligible' is an interesting question, however, it is not one that needs to be explored here. The point that Chalmers and other gap aficionados are pressing is that whatever these conditions of intelligibility are, there are principled reasons why they cannot be met in the case of consciousness.

In what I consider to be the most potent of all of Chalmers' attacks on materialism—a short paragraph from his 1995 paper "Facing up to the hard problem of consciousness"—he presents this problem with extreme clarity. It can be paraphrased to specifically address the mechanistic model of explanation as follows:

> Models of mechanisms, however they are represented, are great for explaining structure and function. They are perfectly suited for explaining macroscopic structures in terms of micro-structures, and for explaining the dynamical features of macro-structures in terms of the organized interactions of micro-structures. As a result, they provide very satisfying explanations of the performance of functions. But the structure and dynamics of material systems yield only more structure and dynamics, so structures and functions are all we can expect models of these systems to explain. But, and here is the crux of the matter, consciousness simply does not seem to be a matter of structure and function. (Chalmers, 1995)

Recall from the previous section that there are two steps to successful reductive explanation. The first involves (re)conceiving the phenomenon to be explained in terms of a set of structural or

[45] In discussions of this particular issue in the literature, Bechtel and Craver are typically construed as holding opposed and incompatible positions. Craver holds an *ontic* view of explanation, which sees real-world mechanisms as an essential feature of explanation, where Bechtel holds an *epistemic* conception of explanation, emphasizing that there is more to explanation than merely citing a mechanism. I'm inclined towards a middle ground that agrees that explanation is necessarily an epistemic activity, but places a strong ontic restraint on the kinds of things that count as explanations. For example, contrary to Bechtel and Abrahamsen, I do not consider "incorrect mechanistic explanations" (explanations which appeal to a mechanism, but not one operative in nature) to be explanations at all, since they fail to locate the phenomenon within the causal structure of the world (2005, p. 425). Nonetheless, I agree with them there is more to providing a mechanistic explanation than merely citing the mechanism. I suspect that this is in fact the view that Craver holds.

functional capacities, while the second involves uncovering the underlying mechanism responsible for those capacities. The mechanistic model of explanation, even bolstered by dynamical modelling techniques, is only capable of providing explanations of structure and function. As a result, in order to mechanistically explain a phenomenon we need to be able to conceive of it as a set of structural or functional capacities. The problem with consciousness is that there are reasons to think that phenomenal qualities, such as the *reddishness* of Levine's experience, simply cannot be (re)conceived of in this way.

There are a number of complex arguments as to why we cannot (re)conceive of phenomenal qualities in terms of structure and function (some of which I will discuss in the final chapter), however, in the end they all fall back on the fact that our conception of phenomenal qualities is arrived at from a subjective perspective, whereas the structural and functional concepts deployed in mechanistic explanation are distinctly objective. It is worth fleshing this out in a bit more detail.

The framework of emergent materialism that I presented in the previous two chapters is committed to the view that conscious experiences and their phenomenal qualities are non-mysterious structural and functional features of neural systems in human brains (or brain-body-environment complexes). According to this view, that is what phenomenal qualities *are*. But that is certainly not how we *conceive* of them. We conceive of phenomenal qualities in terms of how they appear *from the inside* as it were. We conceive of them in terms of how they appear to us from our subjective perspective. Structural and functional concepts (such as oscillations in V4 say) can only describe how they are *from the outside*: from an objective point of view.

John Searle has very clearly demonstrated why this presents a problem. Searle argues that reconception of a phenomenon can only take place when the appearance-reality distinction can be made. That is, we can only reconceive of a phenomenon when the following type of scenario is possible: Initially the phenomenon *appeared* to be like so, but after poking it for a while and examining it more closely, we discovered that in *reality* the phenomenon was actually like so. There are myriad examples of such reconceptions of phenomena in the history of science. For example, from our vantage point here on Earth, the Sun (and other celestial bodies) *appear* to be orbiting the Earth, and until the 1500's our conception of the solar system was colored by this

appearance. Close observation by Nicolaus Copernicus and Galileo Galilei revealed that in *reality* it is the Earth which is orbiting the Sun; it merely appears as though the Sun orbits the Earth because the Earth is rotating. The problem with consciousness, as Searle points out, is that the appearance-reality distinction cannot be made when the phenomenon of interest is the phenomenal qualities of experience because phenomenal qualities *just are* appearances. In the case of phenomenal qualities, "the appearance is the reality" (Searle, 1997, p. 76). Any (re)conception of phenomenal qualities in terms of the structure and function leaves out any reference to the way that they appear to us from a subjective perspective. And, since in this case the phenomenon of interest, *reddishness* say, just is the way a particular phenomenal quality appears to us in experience, any (re)conception of *reddishness* in terms of structure and function (oscillations in V4 say), leaves out the *reddishness* and misses its explanatory mark.

So, while it is certainly true that at least some philosophers of mind and metaphysicians are operating with outdated views about how explanation works in the higher-level sciences (the sciences most appropriate for offering explanations of consciousness), the underlying philosophical point that they are making remains as potent as ever. As it stands, we completely lack an intelligible link between the kinds of processes that take place in material systems, and the phenomena associated with conscious experience. The 'hard problem' is just as hard as ever.

3.3 Summary

I began this chapter by pointing out that the kind of explanation that we really need for consciousness is a reductive (constitutive) explanation, one that allows us to understand consciousness as an emergent feature of certain material systems. I argued that there are two steps to providing such an explanation. The first involves (re)conceiving of the phenomenon to be explained in terms of a set of structural or functional capacitates. The second involves uncovering the mechanisms capable of having those capacities. The reason consciousness poses such a hard problem is that we do not conceive of consciousness in terms of structural or functional capacities. And, as we saw, developments in the philosophy of science have done little to resolve this issue.

In the next chapter I explore some of the implications of the 'hard problem'. Intuitions, about the intractability of the hard problem have been used to fuel a number of metaphysical arguments

that purport to show that materialism must be false. I briefly present a number of these arguments before offering the beginnings of a novel solution.

Chapter 4: Arguments Against Materialism and the Outlines of a Novel Response

In the first two chapters of this thesis, I sketched the outlines of a version of emergent materialism which allows us to maintain that macroscopic features of the world are both real and causally potent without jeopardizing the thesis that only the basic microphysical entities are truly fundamental. In the previous chapter, we saw that conscious experience poses a serious challenge for this view. There appear to be good reasons to think that current methods of explanation are incapable of accounting for consciousness in material terms. Some thinkers have suggested that this does not merely present a challenge for materialism, it demonstrates it to be irreparable flawed. They have devised a number of argument to illustrate their point. Two such arguments are: Frank Jackson's Knowledge Argument, and David Chalmers' Conceivability Argument. Both purport to demonstrate that consciousness poses a metaphysical challenge as well as an epistemological challenge to materialism. In section 4.1 I briefly discuss these arguments and sketch the common materialist responses. In section 4.2 I introduce a distinction between *experience in general* (and the challenge of explaining why certain neural processes are accompanied by subjective experience), and the specific *qualitative character* of experience (and the challenge of explaining why any particular experience *feels* the way that it does) on the other. As the examples in section 4.3 demonstrate, this distinction is not new. However, it has implications for anti-materialist arguments that have been overlooked. I argue that there are in fact two problems of consciousness that often get conflated: the problem of accounting for the existence of *experience in general*; and, the problem of accounting for its *qualitative character*. And, moreover, the conflation of these two problems has made the materialists position appear considerably more dire than it in fact is. I argue that so long as materialists can account for the existence of *experience in general*, then materialism is vindicated. In other words, materialism does not need to be able to explain the relationship between neural systems and the specific *qualitative character* of experience, but it does owe us an explanation of how subjective conscious experiences could be emergent features of material systems. This, of course, is no *easy problem*. However, there is reason to think that it is not *intractably hard*. In the final section, section 4.4, I argue that if *experience in general* is not itself something we experience, then there

is room for a conceptual renovation that will allow us to understand *experience in general* in terms of the kinds of processes that take place in certain material systems. There is, I argue, reason for optimism with regard to the prospects of material science eventually providing an illuminating explanation of the emergence of *experience in general.*

4.1 Arguments against materialism

Over the years, a number of intuitively powerful thought experiments have been developed that pump the intuition that the explanatory gap is not just an artifact of our current incomplete understanding of the relevant neural processes. Rather, they give us reason to think that the explanatory gap is *in principle* unbridgeable. These thought experiments have been used to fuel metaphysical arguments that purport to show that materialism about conscious experience (including the version of emergent materialism advocated in this thesis) must be false. These arguments share the same basic structure. They employ thought experiments to show that an epistemic gap exists between facts about material structure and function on the one hand and facts about conscious experience on the other (an *epistemic* gap occurs between X and Y when Y is such that it cannot be explained in terms of the organized activity of X).[46] They then make the move from an epistemic gap to an ontological gap (an *ontological* gap occurs between X and Y when Y is such that it cannot emerge from the organized activity of X). And finally, they draw the conclusion that materialism must be false. Following Chalmers, the basic structure of these arguments can be presented as follows (2010, p. 110).

1. There is an epistemic gap between material truths and phenomenal truths.
2. If there is an epistemic gap between material truths and phenomenal truths, then there is an ontological gap, and materialism is false.

Therefore

3. Materialism is false.

[46] Advocates of these arguments typically present them in terms of 'facts' and 'truths', and I will follow their lead here. This is nothing more than a short-hand for talking about the instantiation of properties: 'material facts' and 'material truths', is just shorthand for 'the instantiation and distribution of material properties'; and 'phenomenal facts' and 'phenomenal truths', is shorthand for 'the instantiation and distribution of phenomenal properties' (Chalmers 1996, p. 33, 361).

The two most widely discussed such arguments are Frank Jackson's Knowledge Argument and David Chalmers' version of the Conceivability Argument. Before moving on to discuss the implications of distinguishing between *experience in general* and the *qualitative character* of experience, I will briefly present these two arguments and the common materialist responses.

4.1.1 The Knowledge Argument:

In his 1982 paper "Epiphenomenal Qualia", Frank Jackson posed the problem of consciousness in the form of a thought experiment involving Mary, a super-scientist who has spent her life confined to a black and white room. Mary, we are told, knows everything that a completed science of human vision could possible tell her. She has acquired "all the physical [material] information there is to obtain about what goes on when we see ripe tomatoes, or the sky, and use terms like 'red, 'blue', and so on" (Jackson, 1982, p. 130). She knows what neural states correlate with red experiences, what sorts of objects and events bring about those neural states, and what neural processes are responsible for typical behavioral responses to red stimuli (for example, vocalizations such as "What a nice red tomato"). And yet, despite knowing all the physical information that correlates with red experiences, all she's ever seen is black, white, and the various shades of gray. The question Jackson poses is this: If we let Mary out of her room and she sees a red rose for the first time, has she learned something new about color? The seemingly inescapable conclusion we are to draw is that she has, now she knows not only what the objective physical correlates (or mechanisms) of red experiences are, now she knows what red *looks like*.

Although he has since recanted[47], Jackson used this thought experiment to fuel a metaphysical argument against materialism which can be summarized as follows.

1. There are truths about consciousness that are not deducible from material truths.
2. If there are truths about consciousness that are not deducible from material truths, then materialism is false.

Therefore

[47] In his 2003 article he confesses that "for some time [he has] thought that the case for physicalism is sufficiently strong that we can be confident that the arguments from the intuitions go wrong somewhere" (Jackson 2003, p. 251).

3. Materialism is false.

The facts that are not deducible from the material facts according to Jackson's argument are the facts about *the particular qualitative character of experience*. Despite knowing all the material facts about the brain's structure and function, Mary does not know what red *looks like*. Here 'deducible' is being used in the strong sense briefly discussed in chapter 2. This indeducibility is not merely an artifact of human cognitive limitations, nor is it something that will go away with further scientific research—Mary already knows all that scientific research can reveal. Rather, it implies either: that there is more to the word than the fundamental physical entities and the systems and properties that can explicably emerge from them; or else, that the relationship between conscious experience and neural processes is a brute empirical fact that we must simply accept with "natural piety" (Alexander, 1920 p. 47).

This can be put more generally as:

1. There is an epistemic gap between material facts and facts about the qualitative character of conscious experiences.
2. If there is an epistemic gap between material facts and facts about the qualitative character of conscious experiences, then there is an ontological gap, and materialism is false.

Therefore

3. Materialism is false

4.1.2 The Conceivability Argument

David Chalmers has pressed the problem in a slightly stronger fashion. According to Chalmers, not only is materialism is incapable of accounting for what particular conscious experiences are *like,* it is incapable of accounting for why there are any conscious experiences in the first place. He makes this point by way of a thought experiment involving zombies. A zombie is a complete physical and functional duplicate of you or I who nonetheless lacks experience altogether. Despite being objectively indistinguishable from normal conscious people, both in their physiological functioning and in terms of their behavioral responses to stimuli, zombies are supposed to have

no subjective experience whatsoever. Although Chalmers does not claim that zombies are a physical possibility, he does think, and has fervently argued, that they are both logically and metaphysically possible. His intuition is that the possibility of zombies is a conceptual truth, and that no amount of physical or functional detail will make the possibility of zombies inconceivable (1996, 2010).[48]

Chalmers argument can be summarized as follows:

1. It is conceivable that there are zombies
2. If it is conceivable that there are zombies, then it is metaphysically possible that there are zombies.
3. If it is metaphysically possible that there are zombies, then materialism is false.

Therefore

4. Materialism is false.

This can be put more generally as:

1. There is an epistemic gap between material facts and facts about the existence of conscious experiences.
2. If there is an epistemic gap between material facts and facts about the existence of conscious experiences, then there is an ontological gap, and materialism is false.

Therefore

3. Materialism is false

4.1.3 Common Replies to these Arguments.

Following Chalmers, we can carve the common materialist responses to these arguments into three camps: Type A, Type B, and Type C (2010, p. 110-123).

[48] It is important to note here that in order to faithfully track the notion of conceivability that is relevant to the conceivability argument—ideal primary conceivability—this epistemic gap will have to be quite strong. The epistemic gap cannot be a mere feature of our current ignorance, it would have to remain even for a perfect (or ideal) reasoner, with perfect (or ideal) concepts. See Chalmers (2002) for details.

Type A materialists, such as Dan Dennett, Patricia Churchland, and Paul Churchland, are those who simply deny that there is any such epistemic gap between the material facts and the facts about conscious experience. Typically, they take the form of either eliminativists, analytic behaviorists, or analytic functionalists. Eliminativists deny that there is an epistemic gap by simply denying that there is anything at all that needs explaining. That is, they make the unbelievable claim that conscious experience is merely an illusion, that it doesn't really exist. Despite a number of thinkers being charged with eliminativism, whether or not anyone has ever genuinely been an eliminativist about 'the phenomenon' of experience is not entirely clear to me. Rather, it seems that what they want to eliminate is a particular conception of the qualitative character of experience that sees the qualitative features of experience—the qualia—as essentially private, ineffable, and incorrigible.[49] This is entirely consistent with experiences still existing. Of course this sort of reading leaves them open to the zombie argument. If they do not in fact eliminate the phenomenon of experience, then zombies still present a problem. On this front, the eliminativists will most likely join forces with the analytic behaviorists or the analytic functionalists.

Analytic behaviourists and analytic functionalists, rather than deny the existence of experiences or their qualitative character, simply redefine them in behavioural or functional terms. For analytic behaviourists, all it means to see red is to respond in a particular way. According to the analytic behaviourist, Levine's *reddish* experience is nothing more than his behaving in a prototypically *red* fashion (whatever that might be). Since it is typically not disputed that material facts can explain behaviour, according to analytic behaviourism there is no deep issue associated with explaining Levine's *reddish* experience in material terms. Analytic functionalism makes a similar move, but rather than (re)defining experiences in terms of behaviour, analytic functionalists (re)define them in terms of their relations to environmental input, behavioural output, and other mental states. Chalmers argues, and I agree, that ultimately, both views are deeply unsatisfying (especially with respect to the qualitative character of experience).

[49] What Dennett has to say in the opening remarks of his paper "Quining Qualia" are telling in this regard.

> Which idea of qualia am I trying to extirpate? Everything real has properties, and since I do not deny the reality of conscious experience, I grant that conscious experience has properties. … My claim … is that conscious experience has no properties that are special in any of the ways qualia have been supposed to be special. (Dennett 1988, pp. 43)

Behavioural concepts and functional concepts are inherently objective, they refer to features of the world from an objective, third-person perspective. As I argued towards the end of last chapter, there is a subjective aspect to conscious experiences that simply cannot be captured in objective terms. Redefining *reddishness* in terms of behavior or function, leaves out how red *appears* to us from a subjective perspective. And, since the phenomenon of interest *just is* how something appears from a subjective perspective, accounts of behaviour and function miss their explanatory mark.

Type B materialists are those who accept that there is indeed a deep epistemic gap but deny that this entails an ontological gap. They endorse what have been called brute necessities. When Levine looks at his red diskette, the cascade of neuronal activity in his brain that follows is indeed one and the same thing as his *reddish* experience, however, according to Type B materialism, there is simply no illuminating explanation as to how or why this is so—it is a brute fact of the universe that we must simply take in our stride.

Type C materialism tries to walk a path between these two positions. Type C materialists say "Yes experiences are real" and "No they cannot be (re)defined in behavioral or functional terms", but they don't think that this implies that they are brute inexplicable facts about the world. They accept that there is currently a deep epistemic gap between the material facts and the facts about conscious experience, but they hope that one day this gap will be closed. Type C materialism is essentially an 'I owe you' for future work. Despite acknowledging its intuitive appeal, Chalmers has argued that ultimately there is no room for the Type C materialist. According to Chalmers, Type C materialism inevitably collapses into one of the other options: either future science will define conscious experiences in terms of material structures and functions—giving us a version of Type A materialism; or it won't—resulting in Type B materialism (or else some distinctly non-materialist alternative).

As we will see however, there is an avenue that Chalmers has missed that stems from the distinction between *experience in general* and *qualitative character* that I shall introduce in the next section. It is possible to embrace both Type A and Type B materialism at the same time. It may be that future science will allow us to define *experience in general* in terms of structure and

function, while accepting the brute inexplicability of the *qualitative characteristics of conscious experiences*.

4.2 The Two Problems of Consciousness

The common assumption is that the two arguments above—the Knowledge Argument and the Zombie Conceivability Argument—represent a single 'explanatory gap', and a single 'hard problem' for materialism. I want to call this into question. In this section I will argue that there are two distinct problems of consciousness: the problem of accounting for the existence of experience in general, and the problem of accounting for the qualitative character of experience. As I have mentioned, this distinction is not new. It, or some very close analogue, is evident in a number of areas within the literature on consciousness. After introducing this distinction in more detail I will provide evidence of it in three places: the contrast between state-based and creature-based investigations into the neural correlates of consciousness (NCC); the contrast between two kinds of philosophical theories of consciousness (representational theories and higher-order theories); and the contrast between the various philosophical arguments against materialism—such as the two briefly discussed in the previous section. Despite being implicit in a wide range of discussions, the distinction between these two problems is seldom explicitly stated. More importantly however, its implications for materialism have not been thoroughly explored. The task of this section is to present and motivate this distinction. In the next section I discuss one of its implications for the arguments against materialism.

To get a clearer picture of the distinction that I have in mind, consider again Levine's presentation of the explanatory gap.

> As I now look at my red diskette case, I'm having a visual experience that is reddish in character. Light of a particular composition is bouncing off the diskette case and stimulating my retina in a particular way. That retinal stimulation now causes further impulses down the optic nerve, eventually causing various neural events in the visual cortex. Where in all of this can we see the events that explain my having a reddish experience? (Levine, 2001, p. 76-77)

What exactly is Levine asking here? Is he asking, "Where in all of this can we see the events that

explain my having a *reddish* experience?", rather than, say, a *bluish*, or a *greenish* experience? Or, is he asking, "Where in all of this can we see the events that explain my having a reddish *experience*?", as opposed to no experience at all? When Levine looks at his red diskette, there are two questions that we want to be able to answer that we currently cannot:

1. Why does he experience?
2. Why does his experience have a *reddish* character?

These two questions represent what I take to be the two problems of consciousness: the problem of accounting for the existence of *experience in general*, and the problem of accounting for the *qualitative character* of experience. The problem of accounting for the existence of *experiences in general* is the problem of determining what is required for an organism or system to have experiences of any kind. It is the problem of (not) being able to conclusively determine whether or not any particular entity is the subject of experiences. By contrast, the problem of accounting for the *qualitative character* of experience is the problem of determining what a particular conscious agent's experiences are like for it. It is the problem of (not) being able to answer Thomas Nagel's famous question: what is it like to be a bat? These two problems represent two distinct explanatory projects. Answering the first question, and solving the problem of the existence of experience requires uncovering the basis (material or otherwise) of *experience in general*. Answering the second question, and solving the problem of the *qualitative character* of experience requires uncovering not only the basis of experience in general, but, in addition, explaining why particular experiences *feel* the way they do.

Before providing evidence of this distinction already at work in the various literatures on consciousness, a clarificatory comment is in order. It is important to stress that the distinction between these two explanatory projects is not intended to imply that *experience in general* and the *qualitative characteristics* of experiences can exist independently. They can't. There are no instances of experience in general that are devoid of qualitative character, and there are no instances of qualitative character in the absence of experience in general.[50] Rather, following Tim

[50] Somewhat surprisingly, some people do want to dispute this. For example, Rosenthal thinks that qualitative character, or what he calls "mental qualities", can exist in the absence of experience; there can be unconscious pains

Bayne (2007) I take *experience in general* and *qualitative character* to be related in much the same was that *being coloured* is related to *being red*, or in much the same way that *precipitating* is related to *snowing*: via the determinate/determinable relation. To borrow an example from McClelland (2014), when it is snowing, there are not two wholly distinct phenomena taking place—*snowing* and *precipitating*—rather, snowing just is one way of precipitating. Likewise, when Levine has his reddish experience there are not two distinct phenomena taking place—*reddishness* and *experience*—rather, there is just Levine's *reddish experience.* Nonetheless, just as accounting for why it is precipitating is a different explanatory project than accounting for why it is snowing, so too accounting for why Levine is the subject of experience is a different explanatory project than accounting for why his experience is *reddish* in character.

One of the places this distinction (or a very close analogue) is already at work in the consciousness literature is the in the contrast between state-based and creature-based research into the neural correlates of consciousness (NCC). Both Tim Bayne (2007) and Jakob Hohwy (2007) have noted that empirical investigations into the neural underpinnings of consciousness can be grouped in terms of whether they focus on "the contrast between one state of consciousness and another", or whether they focus on "the contrast between consciousness and its absence" (Bayne, 2007, p. 2). Bayne suggests that these two distinct approaches in NCC research indicates that there are two distinct phenomena, each in need of explanation: "creature consciousness", and "state

for example (2011 .435). If pain is to be understood in terms of how it feels, in terms of its qualitative character, then this strikes me as incoherent. I can understand how representational content can exist in the absence of experience, but I cannot understand how *it's being like something in particular* can exist in the absence of *it being like anything at all*. Of course, Rosenthal can insist that qualitative character *just is* representational content, but then in explaining representational content he has not given us an explanation of the phenomenon we are after—namely the qualitative character of experience. Conversely, others hold that experience in general can occur in the absence of qualitative character. The notion of "witness-consciousness" from within meditative traditions, such as Buddhism, is supposed to refer to a deep meditative state in which all phenomenal content has been stripped away. Again, if phenomenal content is understood as qualitative character, then I find this to be incoherent. Intuitively, someone is experiencing just in case it is like something to be them. If all qualitative character is stripped from their experience, it is no longer like anything to be them, and, intuitively, they are no longer experiencing. In defence of witness consciousness, Miri Albahari has argued that even once all the phenomenal content is stripped from consciousness, it still feels like something to be conscious (personal communication August 12, 2016). Again, this seems straightforwardly contradictory to me. I can understand how all 'meaning' could be stripped from conscious experience and it still be 'like something'. But I cannot see how it is possible to have all the 'what it's likeness' stripped away, and it still remain 'like something'. Given my lack of meditative training, and lack of first person experience with the kind of state described by witness consciousness, perhaps I should withhold judgement, but my philosophical intuition calls balderdash.

consciousness" (Bayne 2007). It is important to note that by 'creature consciousness' Bayne does not have in mind mere sentience or wakefulness as the term is sometimes used.[51] Rather, for Bayne, 'creature consciousness' is the property of being phenomenally conscious—it is the property of there being something it is like to be you. As such, Bayne's use of 'creature consciousness' is in line with my notion of *experience in general*. Similarly, "state consciousness", is analogous to my use of *qualitative character*. Conscious states are individuated in terms of their phenomenal character—in terms of what it is like to be in them. It is natural to see creature-based research as trying to solve the problem of the existence of experience in general, while state-based research as trying to solve to problem of qualitative character.

Another place this distinction (or a very close analogue) is evident is in the contrast between representational theories of consciousness and higher-order theories of consciousness. After distinguishing between a phenomenal state's "qualitative character"—which is what makes it the phenomenally conscious state that it is—and its "subjective character"—which is what makes it a phenomenally conscious state at all—Uriah Kreigel notes that theories of consciousness tend to target one or the other of these aspects of consciousness depending on where they think the true mystery of consciousness lies (2009, p. 1).[52] For example, representational theories of consciousness seem to be targeting the qualitative character of experience (they try to account for qualitative character in terms of representational content), whereas higher-order theories seem to be targeting its subjective character (they try to account for subjective character in terms of self-representing systems—systems that represent themselves as representing the world).[53]

At first glance it appears as though Kriegel's distinction between "qualitative character" and "subjective character" is the same as the one I am making. However, closer inspection leads me to question this. The issue is to do with precisely how to understand *subjective character.* Sometimes, Kriegel uses this term to refer to something very much like my notion of *experience*

[51] Rosenthal (2005) uses creature consciousness this way.

[52] Levine (2001) makes a similar point. See pages 104, and 175 especially.

[53] A similar point can be made for theories of consciousness that attempt to tackle both of these problems. For example, O'Brien and Opie (1999) seek to account for the qualitative character of experience in terms of structural similarities in the vehicles of representation, and they seek to account for the existence of experience in general in terms of stable activation patterns in suitable connectionist networks.

in general. This is evident when Kriegel says that subjective character "is invariant across all phenomenal characters, and captures the existence (rather than identity) conditions of phenomenality" (2009, p.58). Elsewhere however, subjective character seems to refer more to the existence of the 'self' or the existence of a particular 'point of view', and not necessarily to the existence of conscious experience. This appears to be the case when Kriegel says that "to say that my experience has subjective character is to point to a certain *awareness* I have of my experience" (2009, p.8). Here it seems as though experience is already present, and subjective character is merely a feature of that experience.[54] Additionally, although I don't think Kriegel makes this conflation, it is important to be clear that qualitative character, as I use the term, is not synonymous with representational content.

I bring up this possible discrepancy between our views to be explicitly clear that it is not *representational content* and *subjectivity* that I am distinguishing between. What I am interested in is the distinction between *experience in general*, understood as a determinable property shared by all instances of phenomenal consciousness, and *qualitative character*, understood as the vast array of possible determinates of experience in general. One may hold that qualitative character can be accounted for in terms of representational content. And one may hold that experience in general can be accounted for in terms of the existence of a self or the existence of a subjective point of view. But these are substantive claims that need to be defended. When it comes to explaining consciousness, the phenomena of interest are not representational content, and subjectivity. Rather, when it comes to explaining consciousness, the phenomena of interest are the existence of *experience in general*, and the *qualities* that characterize what those experiences are like from the perspective of the experiencer.

A third place where this distinction is evident is in the contrast between the various metaphysical arguments against materialism. That consciousness poses two distinct problems is implicit in the distinction between the Knowledge Argument, and the zombie version of the Conceivability Argument that I briefly discussed earlier (I will elaborate on this shortly). It is also implicit in the distinction between the *invert* and *zombie* versions of Chalmers' conceivability argument. Where

[54] Levine (2001) has a similarly unclear usage of subjectivity.

zombies are physical duplicates of you or I that lack conscious experience altogether, inverts are physical duplicates that have experiences, but whose experiences are qualitatively different from our own.[55] To see the difference between the two, consider how they apply to Levine and his reddish experience. When *real* Levine looks at his red diskette case, he has a reddish experience and says "How can we explain my reddish experience?" When *invert* Levine looks at his red diskette case however, he has, what would be for real Levine be a greenish experience but still says, "How can we explain my reddish experience?" Whereas, when *zombie* Levine looks at his red diskette case, he has no experience whatsoever but still says "How can we explain my reddish experience?"

The invert version of Chalmers' conceivability argument says that if inverts are conceivable then materialism is false, whereas the zombie version says that if zombies are conceivable then materialism is false. That there are two problems of consciousness as mentioned above is implicit in the distinction between these two versions of the argument: the conceivability of inverts (if the conceivability arguments are to be believed) shows that the problem of accounting for the qualitative character of experience is insoluble, whereas the conceivability of zombies shows that the problem of accounting for the existence of experience in general is insoluble.

Despite the fact that both the problem of accounting for the existence of experience and the problem of accounting for its qualitative characteristics are widely discussed in both the philosophical and scientific literature on consciousness, anti-materialists about consciousness typically conflate these two problems into a single challenge for materialism. That is, it is typically assumed that if either of the two problems cannot be given a materialist solution, then materialism is in trouble. To see that this is the case consider again the two arguments against materialism presented in the previous section.

In Jackson's thought experiment, Mary learns something new when she leaves her black and white room. What she learns is a fact about *the qualitative character of experience*: she learns

[55] The roots of the invert/zombie distinction can be traced back to Block and Fodor's (1972) paper "What Psychological States are Not" in which they distinguish between absent vs inverted qualia cases: in a case of inverted qualia we have two functionally identical systems who's experiences differ in terms of qualitative character; in the case of absent qualia we have two functionally identical systems, only one of which has any experiences at all.

what red *looks like*. Presumably however, she doesn't learn anything new about *experience in general*: about what it is for something to *be* a conscious experience. Jackson's thought experiment clearly targets the problem of accounting for the qualitative character of experience. It purports to show that the particular qualitative character of any particular experience cannot be reductively explained in terms of the underlying material (neural) processes. If it could, then presumably Mary would not learn anything new when she sees color for the first time. His thought experiment does not address the prospects for reductively explaining the existence of *experience in general*. Jackson argues that when Mary leaves her room she learns something new about the *qualitative character* of color experiences; he says nothing about whether she learns anything new about the relationship between the kinds of material processes that take place in the human brain and the existence of *experience in general*. Jackson's thought experiment presents the problem of consciousness as the problem of explaining the qualitative character of experience. The Knowledge Argument says that since the *qualitative character* of conscious experience cannot be reductively explained in material terms, there must be an ontological gap between material processes and the qualitative character of experience, and, as a result, materialism must be false.[56]

Unlike Jackson's thought experiment, Chalmers' zombie argument is not intended to show merely that we will never be able to attain an illuminating reductive explanation of the *qualitative character* of experience, it makes the stronger claim that we will never even be able to attain an illuminating reductive explanation of why there are experiences at all. The zombie intuition targets the problem of accounting for the very existence of *experience in general*. The zombie version of Chalmers' Conceivability Argument says that since *experience in general* cannot be reductively explained in material terms, there must be an ontological gap between material processes and *experience in general*, and, as a result, materialism must be false.

While there is a wealth of literature debating the validity of these arguments, the overwhelming consensus is that each of these problems presents a metaphysical challenge to materialism. That is, the two problems of consciousness—the problem of accounting for the existence of

[56] Thomas Nagel's famous portrayal also targets this problem. Recall that the title of Nagel's paper was "What is it like to be a bat?" not "Is it like anything to be a bat?" (Nagel, 1974).

experience in general, and the problem of accounting for the qualitative character of particular experiences—are conflated into a single, super-hard problem for materialism.

Although, as I have demonstrated, the distinction between these two problems is not at all new, as far as I am aware, the only person to have begun to explore the metaphysical implications of the possibility that the solutions to these two problems come apart, is Tom McClelland (2014). After explicitly distinguishing between the problem of accounting for the existence of experience in general and the problem of accounting for the qualitative character of experience, McClelland argues that the problem of consciousness is neither 'hard' nor 'easy', rather it is "tricky". The problem of consciousness is "tricky" if one of the two problems can be given a materialist solution but not both.

McClelland's extremely interesting paper does an excellent job of laying out the metaphysical options that this distinction offers. However, McClelland still considers both of these problems to be presenting a metaphysical challenge to materialism, thus reinforcing the general consensus that if either the existence of experience, or the qualitative character of experience cannot be reductively explained, then materialism is in trouble. This is a mistake. In the next section I sketch an argument that, if sound, shows that only the problem of accounting for the existence of experience in general poses a serious threat to materialism. I argue that if the existence of conscious experience can be reductively explained then materialism is vindicated irrespective of whether the qualitative character of experiences can also be reductively explained.

Before proceeding, it is important to say something about why I take the converse not to hold; about why explaining the *qualitative character* of experience does not have the same materialism-vindicating quality.

Some thinkers, for example Antti Revonsuo (2006), are drawn to the idea that we may be able to explain much of the *qualitative character* of experience, without having to explain why there are experiences in the first place, by uncovering structural properties in neural activity that reflects certain structural properties of phenomenal consciousness. For example, we may be able to explain the 'circularness' of someone's experience when they look at the moon by uncovering certain structural properties in the patterns of neural activation that take place in their brain at

that time. Additionally, we may be able to account for the fact that an experience of *red* is more like an experience of *orange* than one of *green* by noting that the neural activity that underlies those experiences is similarly related (i.e., the neural activity that underlies *red* experience may be more similar to the neural activity that underlies *orange* experiences than *green* experiences). Ultimately however, I think that such attempts fail to solve the problem of the qualitative character of experience for three reasons. The first is that they still leave Mary in the dark (or in the grey) with regard to what red looks like. The second is that at best, such approaches could give us knowledge of the conditional 'if X experiences anything, then its experiences are like Y'. The third is that for certain aspects of conscious experience (colour experiences for example), inversions of qualitative character are possible even when the structural relations between the different colour experiences are taken into account (Palmer, 1999). As a result, these approaches to explaining qualitative character without explaining experience in general leave all the metaphysical arguments against materialism unscathed: Mary still doesn't know what red looks like; zombies are still conceivable since at best all we know is what agents' experiences are like if they are like anything at all; and inverts are still conceivable for certain types of experiences.

4.3 Why Materialists Need not Fear Qualia

The argument that follows relies on two plausible theses:

> (T1) Experiences necessarily have qualitative character.
>
> (T2) Reductively explaining the existence of experience does not require reductively explaining the qualitative character of experience.

I take T1 to be a conceptual truth. As I mentioned in the previous section, I hold that experience in general and qualitative character are related as determinable to determinate (i.e., they stand in the same kind of relation that 'being coloured' and 'being red' stand in). A feature of this relation is that any object instantiating a determinable "must also instantiate some determinate under that determinable" (Funkhouser 2006 p. 549). So, just as no object is colored without being some particular colour, so too no-one is experiencing without experiencing in a particular way. There is no such thing, in other words, as a bare experience—an experience devoid of qualitative character. In Nagel's parlance, to be conscious is for it to be like something to be you. If there is

nothing it is like to be you (if you are devoid of subjective qualitative character) then you are not experiencing (Nagel, 1974). To put this concisely: the existence of *experience in general* entails the existence of *qualitative character*.

The second thesis is not so obviously true. However, as the examples from the previous section demonstrate, I am certainly not alone in endorsing this view. It is natural to see higher order theories of consciousness, such as David Rosenthal's (2005) and Robert van Gulick's (2004), as attempting to reductively explain *experience in general* without, as far as I can see, explaining qualitative character. Conversely, representational theories of consciousness, such as Michael Tye's (1995) and Fred Dretske's (1995), can be seen as attempting to reductively explaining the qualitative character of experience, but they do not offer an explanation of experience in general. A similar characterization can be given to the explanatory targets of creature-based and state-based NCC research.[57] It is worth exploring this matter a little further however, because a lot of what follows rests on this claim.

First, it is worth making explicit what T2 entails. T2 is not committed to the existence of experience in general actually being reductively explainable (although I will speculate that it is in the next section). Nor is it committed to the claim that the qualitative characteristics of experience are not reductively explainable (although on this front I remain skeptical).[58] Rather, all I am arguing here is that the following conditional claim is true: if it is possible to reductively explain the existence of experience in general, doing so does not also require reductively explaining the qualitative characteristics of experience. What I need to do here is to show that the solutions to these two problems can come apart.

A common reply at this point is the following: "If, as you say, experience in general and qualitative

[57] One could argue that NCC research is not seeking to provide reductive explanations, rather it has the humbler goal of seeking to uncover the neural processes that correlate with the various features of consciousness. I suspect that this is not what scientists working on the NCC's see themselves as doing. They may (and should) accept that merely uncovering the mechanisms that correlate with the various features of consciousness does not by itself constitute an explanation—as I argued earlier, explanation is still an epistemic activity; we need an illuminating account of why the correlation holds—nonetheless, I suspect that it would be a mistake to think that scientists working on consciousness are not seeking an *explanation* of that phenomenon.

[58] My reason for thinking this is that I hold that there is room for a conception of experience in general in terms of structure and function, but I don't think the same is true for qualitative character. I elaborate on this shortly.

character are related as determinable to determinate, can a constitutive explanation of the two really come apart? How can you constitutively explain a determinable without constitutively explaining at least one of its determinates? How can you explain what colour is without explaining at least one colour? How can you explain what it is to be shaped without explaining what it is to be some particular shape?"

To motivate the idea that this is possible, and is actually not uncommon in scientific practice, I will consider two examples. The first concerns snowflake formation. The second concerns the discovery of DNA and the end of the vitalist/mechanist debate.

I should say up front that I don't regard these examples as direct analogies. In fact, I suspect that there are no direct analogies for the case at hand: consciousness is a singularly weird phenomenon. Nonetheless, I think these examples make a strong case for the claim that it is at least possible to have a reductive explanation of a determinable phenomenon, without necessarily having a reductive explanation of any of the determinates of that determinable. It is also important to stress that I am not claiming that actual science always proceeds by first explaining the determinable without explaining any of its determinates, and then later seeks to explain the particular instances. Maybe there are some cases of scientific progress that have proceeded that way (snowflake formation may be an example) but that is not the claim being made here. Rather, all I am trying to demonstrate with these examples is that it is possible to have a reductive explanation of a determinable that does not explain any of its determinates. I have separate reasons for thinking that in the special case of consciousness an explanation of the determinable is all we are likely to get: namely, that only the determinable *experience in general* can be conceived of in terms of structure and function.

We start with snowflakes. Our current understanding of snowflake formation is sufficiently detailed for us to explain how it differs from other forms of precipitation, such as rain drops and sleet, but not sufficiently detailed for us to explain how the various kinds of snowflakes form. Although it is perhaps not common knowledge here in Australia, there are a wide range of different kinds of snowflakes: from small hexagonal prisms, to needles, to the more iconic star-shaped snowflakes. While there is much that we know about the process of snow crystal growth

that produces these various kinds of snowflakes—for example, we know that their hexagonal symmetry is due to the particular structure of the H_2O molecule, and we know that snow crystal growth is influenced by both the temperature and the humidity of the surrounding environment, and we know that at low humidity, snow crystals tend to form into small, solid, hexagonal prisms, while it is only at higher humidity and within a narrow temperature range around -15°C that star-shaped snow crystals form—the specific mechanisms of snow crystal growth that determine which type of snowflake forms remain a mystery (Libbrecht, 2001; 2005).

We do however, have a sufficiently detailed understanding of the general mechanism of snowflake formation to be able to distinguish it from other forms of precipitation such as rain and sleet. While rain drops form when atmospheric water vapour condenses into liquid water, and sleet forms when atmospheric water vapour first condenses into liquid and then freezes, snowflakes form when water vapour in the atmosphere skips the liquid phase and transitions directly from a gas into a solid via the process known as deposition. This is a clear case of having a constitutive explanation (or perhaps explanation sketch) of a determinable (snowflake formation) in the absence of a constitutive explanation of any of its determinates (star-shaped snow crystal formation, or prism formation, for example).

My second example concerns the role that the discovery of DNA played in the final demise of vitalism. Although theorists embraced vitalism for a number of reasons, one of the most persistent was that it was hard to see how something as complex and differentiated in its organization as a living organism could possibly develop from a much simpler and seemingly undifferentiated fertilized egg, without the aid of some vital impulse (Bechtel & Richardson 1998). Whether or not a purely material mechanism could be responsible for the development of living organisms was debated well into the 20th century: the mechanists argued that it could be, the vitalists argued that it could not be. Although evidence had been mounting in the mechanists' favor for some time, the final nails in the vitalists' coffin were the discovery of the structure of DNA by Francis Crick and James D. Watson in 1953, and the specification of the mechanisms of protein synthesis that shortly followed.

The question that I want to ask is this: What did the mechanists actually have to show in order to

end the vitalist/mechanist debate? Did demonstrating that 'development' could be achieved by purely material mechanisms require reductively explaining how any (or every) specific phenotype developed? Certainly not. There are phenotypic characteristics that depend on epigenetic processes that we still do not have complete explanations for. Rather, as John Searle noted in a talk given at the University of Montreal in 2012, all the mechanists needed was a description, in broad strokes, of the basic molecular mechanism by which development was achieved.[59] They did not have to provide an account of how any particular phenotype—let alone how every particular phenotype—developed. In other words, all that was needed to close the mechanist/vitalist debate was a reductive explanation of the determinable 'development', it did not require reductively explaining how any determinate 'phenotypic characteristic' was developed.

The same, I suggest, is true of experience in general. In order to show that material mechanisms are capable of having experiences, we do not need an explanation of any particular qualitative characteristic—let alone every particular qualitative characteristic—of experience.

With these preliminaries in place, we can now set about defending materialism against arguments that revolve around the apparent inexplicability of the qualitative character of experience.

As we saw earlier, Chalmers (2010) has presented the general form of these arguments as follows:

1. There is an epistemic gap between material truths and phenomenal truths.
2. If there is an epistemic gap between material truths and phenomenal truths, then there is an ontological gap, and materialism is false.

Therefore

3. Materialism is false.

In order to capture the general shape of a host of anti-materialist arguments, Chalmers has opted for the general notion of 'phenomenal truths' which conflates the distinction made in the previous section. The fact that Levine had an experience when he looks at his red diskette, and

[59] A recording of this talk can be viewed online at https://www.youtube.com/watch?v=yCii726A4Jc. Reference from minute 40.

the fact that his experience was reddish in character, both count as phenomenal truths.

What is often overlooked is that premise 2, the crucial move from epistemic gap to ontological gap, can only work when the phenomenal truth in question is a truth about the existence (or non-existence) of experience. The second premise does not go through when the phenomenal truth in question is one about the qualitative character of experience. In other words, arguments against materialism can only be effective when they target truths about the existence or non-existence of experience. Those that target truths about the qualitative character of experience, although intuitively powerful, are metaphysically impotent. The reason for this will be fleshed out in detail below, but first, to pump the intuition that this is the case, consider the following hypothetical.

Suppose at some point in the future we uncover a mechanism that explains, in an illuminating fashion, how purely material systems can constitute subjects of experience: call it the 'experience mechanism'. Suppose further that although we have an illuminating understanding of the material basis of experience in general, we are still completely in the dark as to why experiences have the particular qualitative character that they do. For example, in this hypothetical future we know perfectly well that Levine has an experience of some kind when he looks at his red diskette case since he instantiates the 'experience mechanism'. However, we are still completely clueless as to why he has a *reddish* experience rather than a *bluish* or a *greenish* experience. In this scenario, there is still an epistemic gap between the material facts and the facts about the qualitative character of conscious experience but—you will note—there is no ontological gap here. In this scenario, we know that conscious experiences are emergent features of material systems, and since, given T1, conscious experiences are necessarily like something—they necessarily have some qualitative characteristics or other—we know that the qualitative character of experiences are emergent features of material systems also.

I think this is intuitively quite straightforward, yet it is worth taking the time to flesh out the argument in detail. The argument itself is unfortunately dense. There are three premises to the argument and together they entail the conclusion that an epistemic gap between facts about the processes that take place in human brains (or brain-body-environment systems) and the facts

about the qualitative character of experience does not entail an ontological gap between the material world and conscious experience. I will present the argument in a general, detailed form, provide some support for each of the premises, and then present the argument again in a more intuitively accessible fashion by considering how it applies to Levine and his red diskette.

The argument proceeds as follows:

Let **B** be 'the kinds of processes that take place in normally functioning brains (or brain-body-environment systems)', **Q** be 'the qualitative character of experience', and **E** be 'the existence of experience in general'.

1. The existence of an epistemic gap between the facts about **B** and the facts about **Q** does not entail the existence of an epistemic gap between facts about **B** and the facts about **E**.
2. If there is no epistemic gap between the facts about **B** and the facts about **E**, then there is no ontological gap between **B** and **E**.
3. If there is no ontological gap between **B** and **E**, then there is no ontological gap between the material world and conscious experience.

If all three premises are true, then it follows that:

4. The existence of an epistemic gap between the facts about **B** and the facts about **Q** does not entail an ontological gap between the material world and conscious experience, and thus, does not entail the falsity of materialism.

Premise 1 follows directly from T2. If reductively explaining the existence of experience in general does not require reductively explaining the qualitative character of experience, then even if there are principled reasons why we cannot reductively explain the qualitative character of experience, this does not entail that we cannot reductively explain the existence of experience in general.

Premise 2 is a truism. Let's consider what we mean by an ontological gap. An ontological gap occurs between X and Y when Y is such that it cannot emerge from (in the sense of emergence advocated in this thesis) the organized activity of X. For there to be no epistemic gap between the material facts and the existence of experience in general requires that we understand how

experience in general can be an emergent feature of certain kinds of material systems. For us to understand how experience in general can be an emergent feature of material system, trivially requires that it can in fact be so. And, if experience in general is an emergent feature of material systems, then there is no ontological gap between the two.

Premise 3 follows from T1. If experiences necessarily have qualitative character, and experiences are emergent features of material systems, then it follows that the qualitative characteristics of experience are also emergent features of material systems. If both the existence of experience and the qualitative character of experience are emergent features of material systems, then there is no ontological gap between the material world and conscious experience, and materialism is not false.

As I mentioned, the argument can be made much more digestible by considering how it applies in the case of Levine and his red diskette. In this case the argument runs as follows:

1. Even if we can't reductively explain the *reddishness* of Levine's experience we may still be able to reductively explain why he *experienced.*
2. If we can reductively explain why he *experienced,* then *experience* is an emergent feature of material systems.
3. If *experience* is an emergent feature of material systems, then properties like *reddishness* are also emergent features of material systems, and materialism is not false.

Therefore:

4. Not being able to reductively explain the *reddishness* of Levine's experience does not entail the falsity of materialism.

As I see it, there are good reasons to think that all three premises are true. If they are, then it follows that an epistemic gap between the material facts and the facts about *what it feels like* to be in a particular conscious state is not sufficient to establish an ontological gap between the material world and conscious experience. As a result, anti-materialist arguments that rely on intuitions about the epistemic gap between the material facts and the facts about the qualitative character of experience (such as Jackson's Knowledge argument and the invert version of

Chalmers conceivability argument) do not directly threaten materialism. In other words, Mary (Jackson's colourblind neuroscientist) doesn't need to know what red looks like in order to understand the relationship between conscious experience and its underlying material basis. And she doesn't need to know what red looks like in order to know that reddishness and all other phenomenal properties are emergent features of complex material systems. Despite not knowing what red looks like, Mary can still be a materialist.

4.4 Can we Reductively Explain the Existence of Experience?

If the above is correct, then as materialists we need not fear the apparent inexplicability of the *qualitative character* of experience. We do, however, have to address the apparent inexplicability of the existence of *experience in general*. For the above argument to be of any use to materialists, we need to be able to cast doubt on the idea that the problem of explaining the existence of experience is intractably hard. In other words, materialists need a reply to the zombie intuition.

I am not going to enter into a detailed discussion of Chalmers two-dimensional conceivability argument here, there is a wealth of literature on the topic already and I have little to add. Suffice it to say that I think Stoljar's response is the right one: we are currently ignorant of some crucial "experience relevant non-experiential fact" and as a result, our conceiving does not meet the strict criteria for establishing a link between conceivability and possibility (Stoljar, 2006, p. 72-74).[60] However, unlike Stoljar, I don't want to restrict our ignorance to ignorance of the fundamental physical properties or of their modes of interacting.[61] Rather, I think it is more plausible that ours is a conceptual ignorance. As I see it, we currently lack the proper conception of *experience in general* that would allow us to see how experiences could be constituted by material processes.

[60] See (Chalmers, 2002) for details of the link between conceivability and possibility.

[61] Stoljar notes two kinds of "experience relevant non-experiential facts" of which we may be ignorant. On the one hand, we may be ignorant of some "basic level" fact about the fundamental constituents of the world that accounts for experience—giving us a version of Russellian monism. On the other hand, our ignorance may also be of some higher-level, "intermediate" level fact about how the fundamentals can be combined and what results. Stoljar does not specify how this "intermediate" level ignorance is likely to play out: our ignorance may be remedied with further empirical research, it may not. As a result, intermediate level ignorance is consistent with 'systemic emergence' as advocated in chapter two, and it is also consistent with stronger versions of emergence in which our ignorance is chronic and cannot be overcome. (Stoljar, 2006, pp. 72-74; 2009).

The goal of this final section then, is not to *solve* the problem of the existence of experience, or even to provide a sketch of a solution. Nor will I suggest that once the qualitative character of experience is cordoned off, explaining the existence of experience becomes an easy problem: it doesn't. As it stands we simply have no idea how to understand *experience in general* in terms of structure and function. And, since our model of reductive explanation is limited to providing explanations of structures and functions, the existence of *experience in general* poses a deep challenge. However, the hard problem only becomes intractable when one accepts, as a matter of principle, that *experience in general* cannot be (re)conceived of in terms of structure and function. My goal here is to cast doubt on this thought. I will begin by motivating the idea that conceptual renovation is the way future science is likely to go, before gesturing at how such a conceptual renovation may be possible for *experience in general*.

As I see it, there are two ways in which the future of consciousness science could unfold. It may be, as Chalmers suggests, that we are currently in a situation analogous to the 19th century physicist trying to understand electromagnetism. When pre-Maxwellian physicists passed an electric current though a wire in the vicinity of a compass needle, they found that like magic, the needle began to move. Furthermore, following Farraday's discovery of mutual induction in 1831, it was well known that one could cause an electric current to flow by moving a magnet within a coil of wire. The physics of the day, Newtonian Mechanics, was incapable of accounting for these phenomena. What Newtonian physicists were ignorant of was a fundamental feature of the world—electromagnetism—a theory of which James Clerk Maxwell produced in 1873. In order to account for the phenomena he observed, Maxwell posited a new fundamental feature of reality. In order to account for electromagnetic phenomena, the set of fundamentals had to be expanded. Perhaps, as Chalmers suggests, the same will be true of consciousness. Perhaps in order to account for the existence of conscious experience, the set of fundamentals will have to be expanded—giving us a version of Russellian monism, or a strong (spooky) version of emergentism.[62]

Alternatively, it may be, as Nagel (1974) suggests, that the materialists' present situation is more

[62] Philip Goff (forthcoming) also predicts this future of consciousness science.

analogous to that of a pre-Socratic holding that matter and energy were the same thing: she would be saying something true but would lack the concepts to understand how it could possibly be so. What she would be lacking would be an understanding of the concepts 'matter' and 'energy' that allowed them to co-refer.

Presently, we are not able to say which of these possible futures will eventuate: we are philosophers after all, not clairvoyants.[63] There is however, a disanalogy between the situation that the 19th century physicists found themselves in, and our own, that renders the second possible future more plausible.

19th century physicists were faced with a number of causal phenomena which they were at a loss to explain. When they moved a magnet between a coil of wire, this caused an electric current in the wire. When they passed an electric current though a wire in the vicinity of a compass needle, this caused the needle to move. By contrast, we do not find ourselves in this situation with regard to conscious experience. This is not to say that conscious experiences have no causal powers. No doubt they do; the pain I felt upon stubbing my toe caused me to inspect if for damage. However, as zombie enthusiasts concede, these causal capacities can all, in principle, be accounted for in terms of the behavior of material systems. What cannot be accounted for, according to the zombie enthusiasts, is why these causal capacities are accompanied by experience.

So, while Pre-Maxwellian physicists were dealing with a causally open theory of physics, as there are numerous causal phenomena that cannot be accounted for by Newtonian Mechanics, as I argued in Chapter 1, and as many anti-materialists readily concede, the current edifice of the material science(s) appears to be causally closed. The new fundamental “psychophysical principles” that Chalmers calls for are not analogous to Maxwell’s electromagnetism since they “will not interfere with physical law”. Rather, these fundamental psychophysical principles are intended to “be a supplement to a physical theory”, providing “the extra ingredient that we need to build an explanatory bridge” between material processes and conscious experience (1995, p. 210).

[63] Or whether the bleaker future envisaged by Colin McGinn (1989) where an understanding of experience remains forever beyond us.

This disanalogy gives us considerable reason to doubt that future science will postulate anything like the causally impotent fundamental psychoneural principles that Chalmers calls for. Rather than postulating the existence of some additional fundamental principle, it is far more plausible (and more desirable) that future science will revolutionize our conception of experience in general allowing us to see conscious experience as a perfectly natural, emergent feature of material systems.

Recall the contrast between Chalmers' presentation of the two steps of reductive explanation and that presented by Craver and Darden. Where Chalmers suggests that we must first get clear about our concepts and only then can we begin to uncover the mechanisms via which they work, Craver and Darden point out that in practice these two steps often happen in tandem. As we learn more about the underlying mechanisms our conception of the phenomenon to be explained evolves. What I suggest is that as we continue to probe the underlying neural mechanism responsible for the existence of experience, we will come to think of experience, at least in part, in terms of these underlying mechanisms. As we come to learn more about its underlying mechanisms, we will come to conceive of experience in general, not solely in terms of *the existence of a subjective world for me*, but also, in terms of a particular kind of material structure operative in human brains. This will not eliminate the phenomenon of experience, nor will it eliminate our subjective conception of experience. It will however, carve off from our conception of subjective experience any dualistic notion that consciousness is essentially different from the material processes that take place in our brain. Whether the same will eventually be true of the qualitative character of experience I remain unsure, however as the argument above demonstrates, the truth of materialism does not rest on this. What distinguishes this view from property dualism, is that although it embraces a kind of conceptual dualism, it denies that these concepts have distinct referents. What makes this a version of Type A rather than Type B materialism, at least with respect to the existence of experience in general, is that it holds that eventually this conceptual renovation will allow for an illuminating explanation of the existence of experience.

Of course, there are considerable challenges here also. As I argued earlier, any (re)conception of phenomenal qualities in terms of the structure and function of neuronal systems leaves out any reference to the way that they appear to us in experience. Any (re)conception of *reddishness* in

terms of oscillations in V4 leaves out the *reddishness*. And since, in the case of qualitative character it is precisely this *reddishness* that we seek to explain, explanations in terms of structure and function miss their mark. To echo Searle, in the case of qualitative character the appearance-reality distinction cannot be made. We cannot simply carve off how qualitative characteristics appear to us and explain their underlying material reality because the qualitative characteristics of experience just are 'appearances'. In the case of the qualitative characteristics of experience, "the appearance is the reality" (Searle, 1997, p. 76). For the qualitative character of experience, I take these concerns to be decisive. However, I suspect that this is another place where the distinction between *experience in general* and *qualitative character* can do some good work.

When I introspect, and attend to my own experiences, what *appears* to me are the phenomenal qualities (the qualitative character) of my experiences: the *feeling of hardness* of the chair I am sitting on, the *ache* in my wrist from typing so many words, and of the deep *reddishness* of the wine that fills my glass. However, as far as I can tell, *experience in general* is not something that *appears* to me in the same manner. I suspect that just as the property of *being shaped* is not itself a shape, so too the property of *being an experience* is not itself an experience. And if *experience in general* is not itself an experience, then it seems that for *experience in general* at least, we may be able to make the appearance reality distinction after all.[64]

To put all this another way, recall that the central issue underlying the hard problem is that reductive explanations are only capable of explaining structure and function, and, so Chalmers claims, facts about experience are not facts about structure and function (Chalmers, 1995, p. 208). It seems at least plausible to hold that facts about the qualitative character of experience are not facts about structure and function. I am not sure how to respond to those who insist that the *reddishness* of Levine's experience is a fact about his subjective experience and not in any way a fact about the structure and dynamics instantiated in his brain. But what about Levine's instantiating *experience in general*? What is that a fact about? If *experience in general* is not itself something we experience then it is hard to see how it could be a fact about Levine's experience.

[64] A similar move can be made for those such as Philip Goff who argue that we form our phenomenal concepts *directly* from our experience and thus could not be mistaken about their referent (2011). If *experience in general* is not something we have we experience, then presumably we *can* be mistaken about the essential nature of what our *experience in general* concept refers to.

Presumably it is a fact about Levine. More precisely, presumably it is a fact about the structures and functions instantiated in Levine's brain.

Although much more work is required in order to fully flesh out this view, there is more room for conceptual renovation in the case of conscious experience than first appearances suggest.

4.5 Summary

In this thesis, I have presented and defended a version of emergent materialism. The brand of materialism that I endorse can be characterized as a commitment to two theses. First, only the fundamental entities and properties postulated by physics (whatever they turn out to be) are genuinely fundamental features of the world. Everything else emerges in a non-mysterious fashion through the organized interaction of these fundamentals. And second, nothing chemical, biological, intentional, moral, or social or experiential is among these basic building blocks. In slogan form: *all macroscopic entities and their properties are non-basic, and everything microscopic is non-mental, non-social, non-biological, and non-chemical.*

To account for how the various non-fundamental features of the world arise and how they relate to the more fundamental features, I developed a conception of emergence consistent with discussions of emergence in systems biology and the other sciences of the mind. This account has three central commitments. First, it sees emergence in any system (whole) that has at least one property that is of a *kind* not possessed by any of its constituent parts. Second, it holds that a system's properties are *non-causally determined* by the organized interaction of its constituent parts. And third, it holds that the environment in which a system is embedded plays an ineliminable role in determining how its parts will be organized and how they will interact. This view embraces the emergence of novel, higher-level causal capacities, and the existence of causally potent higher-level systems. But, it does not license downward causation. In other words, emergence as systemic novelty allows us to understand how causally potent entities can emerge without jeopardizing our commitment to materialism. The major challenge for this view is accounting for conscious experience.

In chapter 3 I argued that the reason conscious experience presents such a hard problem is that the methods of material science are only capable of providing explanations of structures and

functions, and, given our uniquely intimate access to the qualitative character of our own experiences, any attempt to (re)conceive of the qualitative characteristics of experience in terms of structure and function leaves out how they appear to us from a subjective perspective. Since, in the case of explaining consciousness, part of what we seek is an explanation of how particular experiences appear from a subjective perspective, any attempt to explain the qualitative character of experience in structural and functional terms ultimately misses its explanatory target.

In the final chapter I addressed a number of arguments that purport to show that materialism's inability to account for consciousness ultimately demonstrates its inadequacy as a metaphysical position. I have offered the beginnings of a novel solution to this challenge. I have argued, contrary to some of the anti-materialist arguments, that the explanatory gap between the facts about the processes that take place in material systems (such as human brains) and the facts about the qualitative character of experience—even if unbridgeable—does not entail the falsity of materialism. Underlying this negative thesis is also a positive one. In order to conclusively demonstrate the truth of materialism (at least with respect to conscious experience), we need only (only!) explain how *experience in general* can be an emergent feature of systems composed entirely of non-experiential material entities. This, of course, is an extremely difficult task, itself the target of anti-materialist arguments. Realizing that the materialist does not need to explain the qualitative character of experience does not in any way make her task easy. It does however, ensure that we don't make her task harder than it already is. Furthermore, there is reason to be optimistic about the prospects of a conceptual revolution allowing us to see experience in general as a non-mysterious feature of complex material systems, allowing us to close the explanatory gap between the material fact and the existence of experience in general, and once and for all close the debate on the mind-body problem.

Reference List

Alexander, S. (1920). *Space, time, and deity: the Gifford lectures at Glasgow 1916-1918.* London: Macmillan & Co Ltd.

Barron, A., & Klein, C. (2016). What insects can tell us about the origins of consciousness. *Proceedings of the National Academy of Sciences, 113*(18), 4900-4908. doi: 10.1073/pnas.1520084113

Baxter, D. W., & Olszewski, J. (1960). Congenital universal insensitivity to pain. *Brain: A Journal of Neurology,* 83, 381.

Bayne, T. (2007). Conscious states and conscious creatures: explanation in the scientific study of consciousness. *Philosophical Perspectives 21*(1), pp.1-22

Bechtel, W. (2008). *Mental mechanisms: philosophical perspectives on cognitive neuroscience*. New York: Lawrence Erlbaum Associates.

Bechtel, W., & Abrahamsen, A. (2005). Explanation: a mechanist alternative. *Studies in History & Philosophy of Biological & Biomedical Sciences, 36*(2), 421-441. doi: *10.1016/j.shpsc.2005.03.010*

Bechtel, W. and Richardson, R. (1998). Vitalism. In Craig, E. (Ed.) *Routledge encyclopedia of philosophy*. New York: Routledge.

Bennett, K., & McLaughlin, B. (2005). Supervenience. *Stanford Encyclopedia of Philosophy. (Spring 2014 Edition),* Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/spr2014/entries/supervenience/>.

Bergson, H. (1911). *Creative evolution.* London: Macmillan.

Berryman, S. (2008). Ancient atomism. *The Stanford Encyclopedia of Philosophy* (Winter 2016 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/win2016/entries/atomism-ancient/>.

Bickhard, M. H. (2000). Emergence. In P. B. Andersen, C. Emmeche, N. O. Finnemann, & P. V. Christiansen (Eds.), *Downward Causation* (pp. 322-348). Aarhus: University of Aarhus Press.

Blackman, A., Bottle, S., Schmid, S., Mocerino, M., & Wille, U. (2016). *Chemistry (3rd edition. ed.)* Brisbane: Wiley.

Block, N. (1997). Anti-reductionism slaps back. *Noûs, 31*(s11), 107-132.

Block, N. (2003). Do causal powers drain away. *Philosophy and Phenomenological Research, 67*(1), 133-150.

Block, N., & Fodor, J. A. (1972). What psychological states are not. *Philosophical Review, 81*(April), 159-181.

Boogerd, F. C., Bruggeman, F. J., Richardson, R. C., Stephan, A., & Westerhoff, H. (2005). Emergence and its place in nature: a case study of biochemical networks. *Synthese, 145*(1), 131 - 164.

Bourget, D., & Chalmers, D. J. (2014). What do philosophers believe? *Philosophical Studies, 170*(3), 465-500.

Broad, C. D. (1925). *The mind and its place in nature*. London: Routledge & Kegan Paul.

Bromberger, S. (1966). Why Questions. In R. Colodny (Ed.), *Mind and Cosmos*. Pittsburgh: Pittsburgh University Press.

Bunge, M. (1977). Emergence and the mind. *Neuroscience, 2*, 501-509.

Bunge, M. A. (2003). *Emergence and convergence: qualitative novelty and the unity of knowledge.* Toronto: University of Toronto Press.

Bunge, M. (2010). *Matter and mind: a philosophical inquiry*. Dordrecht: Springer Verlag.

Campbell, R. (2015). *The metaphysics of emergence*. London: Palgrave Macmillan.

Campbell, R., & Bickhard, M. H. (2011). Physicalism, emergence and downward causation. *Axiomathes, 21*(1), 33-56.

Carey, N. (2012). *The epigenetics revolution: how modern biology is rewriting our understanding of genetics, disease and inheritance*. London: Icon.

Carey, N. (2015). *Junk DNA A Journey Through the Dark Matter of the Genome*. London: Icon.

Chalmers, D. J. (1995). Facing up to the problem of consciousness. *Journal of Consciousness Studies, 2*(3), 200-219.

Chalmers, D. J. (1996). *The conscious mind*. New York: Oxford University Press.

Chalmers, D. J. (2002). Does conceivability entail possibility. In T. Gendler & J. Hawthorne (Eds.), *Conceivability and possibility* New York: Oxford University Press.

Chalmers, D. J. (2006). Strong and weak emergence. In P. Davies & P. Clayton (Eds.), *The re-emergence of emergence.* New York: Oxford University Press.

Chalmers, D. J. (2010). *The character of consciousness*. New York: Oxford University Press.

Chalmers, D. J., & Jackson, F. (2001). Conceptual analysis and reductive explanation. *Philosophical Review, 110*(3), 315-361.

Churchland, P. S. (1994). Can neurobiology teach us anything about consciousness? *Proceedings and Addresses of the American Philosophical Association, 67*(4), 23-40.

Craver, C. F. (2007). *Explaining the brain: mechanisms and the mosaic unity of neuroscience*. New York: Oxford University Press.

Craver, C. F. (2014). The ontic account of scientific explanation. In M. I. Kaiser, O. R. Scholz, D. Plenge, & A. Hüttemann (Eds.), *Explanation in the special sciences: the case of biology and history* (pp. 27-52) Dordrecht: Springer.

Craver, C. F. (2015a). Levels. In T. K. Metzinger & J. M. Windt (Eds.), *Open MIND*. Frankfurt am Main: MIND Group.

Craver, C. F. (2015b). Mechanisms and emergence. In T. K. Metzinger & J. M. Windt (Eds.), *Open MIND*. Frankfurt am Main: MIND Group.

Craver, C. F., & Bechtel, W. (2007). Top-down causation without top-down causes. *Biology and Philosophy, 22*(4), 547-563.

Craver, C.F. and Bechtel, W. (2013). Interlevel Causation. In D. Werner, O. Wolkenhauer, K. H. Cho, H. Yokota (Eds.) *Springer Encyclopedia of Systems Biology*. New York: Springer

Craver, C. F., & Darden, L. (2013). In search of mechanisms: discoveries across the life sciences. Chicago: University of Chicago Press.

Deacon, T. (2006). Emergence: the hole at the wheel's hub. In P. Clayton & P. S. Davies (Eds.), The re-emergence of emergence (pp. 111-150). New York: Oxford University Press.

Deacon, T. (2013). *Incomplete nature: how mind emerged from matter.* New York: W. W. Norton.

Dennett, D. (1988). Quining qualia. In A. Marcell & E. Bisiach, (Eds.), *Consciousness in contemporary science* (pp. 42-47) New York: Oxford University Press.

Descartes, R. (1641/2008). *Meditations on first philosophy with selections from the objections and replies* (M. Moriarty, trans.). Oxford: Oxford University Press.

Dretske, F. (1995). *Naturalizing the mind*. Cambridge, MA: MIT Press.

Earley, J. (2008). How Philosophy of Mind Needs Philosophy of Chemistry. *HYLE*, 14(1), 1-26.

Feigl, H. (1967). *The mental and the physical*. Minneapolis: University Of Minnesota Press

Fodor, J. A. (1974). Special sciences. *Synthese, 28*(2), 97-115.

Funkhouser, E. (2006) The determinable-determinate relation. *Noûs 40*(3), pp. 548-569. doi: 10.1111/j.1468-0068.2006.00623.x

Galen. (1996). *On the elements according to Hippocrates* (P. De Lacy, Trans.). Berlin: Akademie Verlag.

Gell-mann, M. (March 2007). Murray Gell-mann: Beauty truth and physics [Video file]. URL = <https://www.ted.com/talks/murray_gell_mann_on_beauty_and_truth_in_physics>.

Godfrey-Smith, P. (2008). Reduction in real life. In J. Kallestrup & J. Hohwy (Eds.), *Being reduced: new essays on reduction, explanation, and causation* (pp. 52-73). Oxford: Oxford University Press.

Godfrey-Smith, P. (forthcoming). Mind, matter, and metabolism. *Journal of Philosophy*.

Grahek, N., & Dennett, D. C. (2011). *Feeling pain and being in pain* (2nd ed.). Cambridge: MIT Press.

Goff, P. (2011). A posteriori physicalists get our phenomenal concepts wrong. *Australasian Journal of Philosophy, 89*(2), 191-209. doi: 10.1080/00048401003649617

Goff, P. (forthocoming). *Consciousness and Fundamental Reality*. Oxfore: Oxford University Press. (manuscript available at http://www.philipgoffphilosophy.com/publications.html)

Hempel, C. (1969). Reduction: ontological and linguistic facets. In W. S. S. P. M. Morgenbesser (Ed.), *Philosophy, science, and method: essays in honor of Ernest Nagel*. New York: St Martin's Press.

Hempel, C. G., & Oppenheim, P. (1948). Studies in the logic of explanation. *Philosophy of Science, 15*(2), 135-175.

Hochstein, E. (2016). One mechanism, many models: a distributed theory of mechanistic explanation. *An International Journal for Epistemology, Methodology and Philosophy of Science, 193*(5), 1387-1407. doi: 10.1007/s11229-015-0844-8.

Hohwy, J. (2007). The search for neural correlates of consciousness. *Philosophy Compass, 2*(3), 461-474. doi: 10.1111/j.1747-9991.2007.00086.x.

Jackson, F. (1982). Epiphenomenal qualia. *The Philosophical Quarterly, 32*(127), 127-136. doi: 10.2307/2960077

Jackson, F. (2003). Mind and illusion. In Anthony O'Hear (Ed.) *Minds and Persons*. Cambridge: Cambridge University Press, pp. 251–272.

Kaplan, D. (2015). Moving parts: the natural alliance between dynamical and mechanistic modeling approaches. Biology & Philosophy, 30(6), 757-786. doi: 10.1007/s10539-015-9499-6

Kaplan, D. M., & Craver, C. F. (2011). The explanatory force of dynamical and mathematical models in neuroscience: A mechanistic perspective. *Philosophy of Science*, *78*, 601–627.

Kim, J. (1998). Mind in a physical world. Cambridge MA: MIT Press.

Kim, J. (2005). *Physicalism, or something near enough*. Princeton: Princeton University Press.

Kim, J. (2008). Reduction and reductive explanation: is one possible without the other? In J. Hohwy & J. Kallestrup (Eds.), *Being reduced: new essays on reduction, explanation, and causation*. Oxford: Oxford University Press.

Kriegel, U. (2009). Subjective consciousness: a self-representational theory. Oxford: Oxford University Press.

Lamme, V. A. F. (2004). Separate neural definitions of visual consciousness and visual attention a case for phenomenal awareness. *Neural Networks, 17*(5), 861-872.

Levine, J. (1983). Materialism and qualia: the explanatory gap. *Pacific Philosophical Quarterly, 64*(October), 354-361.

Levine, J. (2001). *Purple haze the puzzle of consciousness*. New York NY: Oxford University Press.

Lewis, D. (1994). Reduction of mind. In S. Guttenplan (Ed.), *Companion to the philosophy of mind* (pp. 412-431) Cambridge MA: Blackwell.

Libbrecht, K. G. (2001). Morphogenesis on ice: the physics of snow crystals. *Engineering and Science LXIV,* 1. Available at: http://www.its.caltech.edu/~atomic/publist/engsci2.pdf.

Libbrecht, K. G. (2005). The physics of snow crystals. *Reports on Progress in Physics 68*(4), pp.855-895

Loss, R. (2015). Parts ground the whole and are identical to it. *Australasian Journal of Philosophy, 94*(3), 489-498.

Machamer, P., Darden, L., & Craver, C. (2000). Thinking about mechanisms. *Philosophy of Science, 67*(1), 1-25.

Mahner, M., & Bunge, M. (1997). *Foundations of biophilosophy*. Berlin: Springer.

McClelland, T. (2014). The problem of consciousness: easy, hard or tricky? Topoi. doi: 10.1007/s11245-014-9257-4

McGinn, C. (1989). Can we solve the mind-body problem? *Mind, 98*(391), 349-366.

McMurray, G. A. (1950). Experimental study of a case of insensitivity to pain. *A.M.A. archives of neurology and psychiatry, 64*(5), 650.

Melzack, R., & Wall, P. D. (1988). *The challenge of pain* (Rev. ed., 2nd ed.). New York NY: Penguin Books.

Moore, G. E. (1922). *Philosophical studies*. London: Routledge & Kegan Paul.

Morowitz, H. J. (2000). *The emergence of everything how the world became complex*. New York NY: Oxford University Press.

Nagel, E. (1961). *The structure of science: problems in the logic of scientific explanation.* London: Routledge & Kegan Paul.

Nagel, T. (1974). What is it like to be a bat? *The Philosophical Review, 83*(4), 435-450. doi: 10.2307/2183914

O'Connor, T. (2005). The metaphysics of emergence. *Noûs, 39*(4), 658-678.

OpenStax College. (2013). *Anatomy & Physiology*. OpenStax College.

Opie, J., & O'Brien, G. (1999). A connectionist theory of phenomenal experience. *Behavioral and Brain Sciences, 22*(1), 127-148.

Opie, J., & O'Brien, G. (2015). The structure of phenomenal consciousness. In S. Miller (Ed.), *The constitution of phenomenal consciousness*. Amsterdam: John Benjamins Publishing Company.

Oppenheim, P., & Putnam, H. (1958). Unity of science as a working hypothesis. *Minnesota Studies in the Philosophy of Science, 2*, 3-36.

Palmer, S. (1999). Color, consciousness, and the isomorphism constraint. *Behavioral and Brain Sciences, 22*(6), 923-943.

Papineau, D. (2001). The rise of physicalism. In C. Gillett & B. M. Loewer (Eds.), *Physicalism and its discontents.* London: Cambridge University Press.

Papineau, D. (2010). Must a physicalist be a microphysicalist? In J. Kallestrup & J. Hohwy (Eds.), *Being reduced: New essays on reduction, explanation, and causation.* Oxford: Oxford University Press, UK

Putnam, H. (1967). Psychological predicates. In W. H. Capitan & D. D. Merrill (Eds.), *Art, mind, and religion* (pp. 37-49). London: Pittsburgh University Press.

Revonsuo, A. (2006). *Inner presence: Consciousness as a biological phenomenon.* Cambridge MA: MIT Press.

Revonsuo, A. (2010). *Consciousness: The Science of Subjectivity.* New York NY: Psychology Press.

Robinson, Howard, "Dualism", *The Stanford Encyclopedia of Philosophy* (Winter 2016 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/win2016/entries/dualism/>.

Rosenthal, D. (2005). *Consciousness and mind*. Oxford: Oxford University Press.

Russell, B. (1912). *The problems of philosophy.* London: Oxford University Press

Ryan, A. J. (2007). Emergence is coupled to scope, not level. *Complexity, 13*(2), 67-77.

Salmon, W. C. (1984). *Scientific explanation and the causal structure of the world*. Princeton, N.J: Princeton University Press.

Schier, E. (2010). The explicable emergence of the mind. In W. Christensen, E. Schier, & J. Sutton (Eds.), *ASCS09: Proceedings of the 9th Conference of the Australasian Society for Cognitive Science* (pp. 306-310). Sydney: Macquarie Centre for Cognitive Science.

Scriven, M. (1962). Explanations, predictions, and laws. In H. Feigl & G. Maxwell (Eds.), *Scientific explanation, space, and time.* Minneapolis MN: University of Minnesota Press.

Searle, J. R. (1992). *The rediscovery of the mind.* Cambridge, MA: MIT Press.

Shapiro, L. A. (2004). *The mind incarnate*. Cambridge, MA: MIT Press.

Silberstein, M. (2002). Reduction, emergence and explanation. In P. K. Machamer & M. Silberstein (Eds.), *The blackwell guide to the philosophy of science* (pp. 80-107): Hoboken NJ: Wiley.

Silberstein, M. (2012). Emergence and reduction in context: Philosophy of science and/or analytic metaphysics. Metascience, 21(3), 627-642.

Silberstein, M., & Chemero, T. (2012). Constraints on localization and decomposition as explanatory strategies in the biological sciences. *Philosophy of Science, 80*(5), 958-970. doi: 10.1086/674533

Silberstein, M., & McGeever, J. (1999). The search for ontological emergence. *Philosophical Quarterly, 49*(195), 201-214. doi: 10.1111/1467-9213.00136

Smart, J. J. C. (1978). The content of physicalism. *Philosophical Quarterly, 28*(October), 339-341.

Smart, J. J. C. (1981). Physicalism and emergence. *Neuroscience, 6*, 109-113.

Spurrett, D., & Papineau, D. (1999). A note on the completeness of "physics". *Analysis, 59*(1), 25-29.

Stoljar, D. (2006). *Ignorance and imagination: The epistemic origin of the problem of consciousness.* Oxford: Oxford University Press.

Stoljar, D. (2010). *Physicalism*. London: Routledge.

Stoljar, D. (2016). Physicalism. *The Stanford Encyclopedia of Philosophy* (Spring 2016 Edition), Edward N. Zalta (ed.), URL = http://plato.stanford.edu/archives/spr2016/entries/physicalism/.

Strawson, G. (2006). Realistic monism: Why physicalism entails panpsychism. *Journal of Consciousness Studies, 13*(10-11), 3-31.

Taylor, E. (2015). An explication of emergence. *Philosophical Studies, 172*(3), 653-669. doi: 10.1007/s11098-014-0324-x

Teller, P. (1992). A contemporary look at emergence. In A. Beckermann, H. Flohr, J. Kim (Eds.) *Emergence or reduction? Essays on the prospects of nonreductive physicalism*. (139-153). Berlin: Walter de Gruyter.

Teresi, D. (2003). *Lost discoveries: The ancient roots of modern science - from the Babylonians to the Maya.* New York NY: Simon & Schuster.

Tononi, G., & Koch, C. (2008). The neural correlates of consciousness: An update. *Annals of the New York Academy of Sciences*, 1124, 239-261. doi: 10.1196/annals.1440.004

Tortora, G., & Derrickson, B. (2012). *Principles of anatomy & physiology.* (13th ed.). Hoboken NJ: Wiley

Tye, M. (1995). *Ten problems of consciousness: a representational theory of the phenomenal mind*. Cambridge, MA: MIT Press.

Van Gulick, R. (2001). Reduction, emergence and other recent options on the mind/body problem: A philosophic overview. *Journal of Consciousness Studies, 8*(9-10), 1-34.

Van Gulick, R. (2004). Higher-order global states: an alternative higher-order model of consciousness. In Gennaro, R. (Ed.), *Higher-order theories of consciousness: an anthology*. Philadelphia, PA: John Benjamins Publishing Company.

Wilson, J. M. (2006). On characterizing the physical. *Philosophical Studies, 131*(1), 61-99.

Wimsatt, W. (1997). Aggregativity: Reductive heuristics for finding emergence. *Philosophy of Science, 64*(4), S372-S384.

Wimsatt, W. (2000). Emergence as non-aggregativity and the biases of reductionisms. *The official Journal of the Association for Foundations of Science, Language and Cognition, 5*(3), 269-297. doi: 10.1023/A:1011342202830

Weber, B. (2015). Life. *The Stanford Encyclopedia of Philosophy* (Spring 2015 Edition), Edward N. Zalta (ed.), URL = https://plato.stanford.edu/archives/spr2015/entries/life/.

Zumdahl, S. S., & Zumdahl, S. A. (2014). Chemistry (9th ed.). Boston: Cengage Learning.