DOES THE CAUSAL EXCLUSION ARGUMENT HOLD IN A PROBABILISTIC SETTING?

Assessing the Efficacy of Mental Causation in an Indeterministic World

Lisa Grant

5th March 2021

Submitted in partial fulfilment to the requirements for the PhD

to the

Department of Philosophy
University College London

Supervised by Dr. Luke Fenton-Glynn

DECLARATION

I, **Lisa Grant**, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Signed: Lisa Grant Date: 5th March 2021

ABSTRACT

My thesis examines Kim's Causal Exclusion Argument (CEA) against the existence of mental causation of physical effects, which I ultimately argue is unsound. I generalise the CEA into probabilistic terms as I assume that we live in a probabilistic world. This is because orthodox interpretations of Quantum Mechanics are in principle probabilistic. I argue that the CEA, at least in its probabilistic form, is unsound because the analogue version of the causal closure premise is false. If the world is probabilistic then this opens the door for mental causes to 'top up' (or lower) the probabilities of further physical events occurring. Thus the mental can be causally efficacious. Secondly, I put forward a positive argument in favour of mental causation based on the natural kindhood of mental properties. Each mental state has a corresponding brain state, both of which could be conceptualised as a kind. I will argue that mental kinds are more natural (albeit imperfectly so) than their corresponding brain states and therefore that it is the mental rather than brain states which are the better candidates to feature in scientific laws. So, mental states have causal efficacy. This branch of the argument can apply to worlds whether they're deterministic or probabilistic. Thus, even if the reader does not share my assumption that the world is probabilistic, or doesn't agree with my rejection of causal closure, there are still some reasons to doubt the soundness of even the original deterministic CEA.

IMPACT STATEMENT

This thesis is a defence of the causal efficacy of the mental in particular against Kim's Causal Exclusion Argument (CEA). I put forward a negative argument as to why the CEA is unsound and a positive argument in favour of mental causation. I hope both these strands of thought can be of wider use and interest within academic philosophy.

The negative argument regards the principle of causal closure. Beyond the realm of mental causation, there may be broader areas of philosophical interest where arguing against the principle of causal closure may be fruitful. For example, such arguments are relevant to debates around free will. It's plausible that mental causation is a requirement for free will. This is because not only is the ability to do otherwise is necessary but the choice must come from an individual in order to be considered their will. I discuss the relevance of my arguments to debates in free will in slightly more detail in the introduction to the thesis.

Regarding the positive argument I put forward, this was based on the concept of natural kindhood. This may also apply to wider social science properties as they too can plausibly held to constitute natural kinds. My argument therefore can be used to build a defence for the causal efficacy of special science properties and the autonomy of these sciences. I touch on this topic briefly in the conclusion to the thesis.

Outside of philosophical academia my work has already found some interest from a psychiatrist doing some interdisciplinary work around models of mind within psychiatry. As the mind is integral to psychological and psychiatric fields, ways of conceptualising mental causation may be of use within a physically based medical world view.

ACKNOWLEDGEMENTS

I want to start by saying a huge thank you to my supervisor Dr Luke Fenton-Glynn for his guidance and patience throughout the thesis writing process. His knowledge and expertise made writing my thesis much more enjoyable than it otherwise would have been and I learnt a great deal under his supervision.

Furthermore, I'd like to thank all the members of the UCL Philosophy Department who I've worked with in the past few years. They all contributed to creating a great atmosphere for learning and developing ideas. Contributors to both WIP and thesis prep were invaluable for their feedback and discussion. Catherine Dale, Helena Cicmil and Ane Engelstad deserve particular thanks for being great philosophers and for making the department such a welcoming place to be.

Thanks go also to the A. J. Ayer Scholarship fund and to the UCL Scholarship committee for providing me with financial support while working on the MPhil which preceded this work. Thanks also to the London School of Economics Library for being an accommodating employer during the MPhil, as well as to all the staff there for being excellent colleagues. This thesis could not have been completed without the funding provided by the London Arts and Humanities Partnership. My biggest thanks go to them.

Finally, I must thank those people who, while not directly involved philosophically, were of invaluable support to me while writing this thesis. The staff at Prufrock

have been very kind to me over the years while I nursed a coffee and wrote the vast majority of my work. Huge thanks go to Dr Meadhbh McIvor, Nina Lazic and Lisa Franklin for being there come what may and always understanding. This project would have been impossible without them. Paul Smith and Dr Jianan Bao have been a huge support to me both as flatmates and as colleagues. Many chapters in my thesis have been refined and improved through long dinner debates with them. Thanks go also to Dr Mateusz Bieniek who lent support in many ways. However, my biggest thanks must go to my family without whom none of this would have been possible. They have provided for me financially and emotionally throughout my many years of philosophy. Finally, I dedicate this thesis to my favourite person, my sister, Phyllis.

CONTENTS

1

INTRODUCTION

"`Begin at the beginning', the King said gravely, `and go on till you come to the end: then stop'"

Alice in Wonderland - Lewis Carroll

(1856)

There will be seven sections in my introductory chapter. The first (1.1) will introduce my overall thesis. The second (1.2) will explain why I was interested in this topic and why I think it's a topic worth investigating. Thirdly, in Section 1.3 I will introduce Kim's version of the deterministic CEA and my probabilistic analogue. In Section 1.4 I will set out what I take a mental state to be and how I will be using the concept in my thesis. The fifth Section (1.5) concerns free will, a topic

with close ties to mine but one which I will lack the space to properly discuss. Likewise, consciousness is a topic with close ties to mine but which I will not discuss due to space reasons. I briefly address this in Section 1.6. To end the introduction, Section 1.7 will layout the plan for the rest of the thesis.

1.1 Thesis

The purpose of my thesis is to argue in defence of mental causation. I will do this in two largely independent ways. For my first argument (in Chapter 8) I will examine Jaegwon Kim's Causal Exclusion Argument (CEA) against the existence of mental causation. I will argue that the CEA does not hold and therefore that mental causation may exist in a probabilistic setting. That is, the mental may cause physical effects in a probabilistic setting and it may do so in virtue of its being mental.[1] I will do this by generalising the CEA into probabilistic terms as I will assume that we live in a probabilistic world.[2] I will argue that the CEA (at least in its probabilistic form but possibly in both its forms) is unsound because in each case the relevant version of causal closure is false.

Secondly in Chapter 9 I will put forward a positive argument in favour of mental causation based on the natural kindhood of mental properties. Each mental state has an underlying brain state which is a bundle of neuronal firings both

[1] Whenever I write about the existence or non-existence of mental causation I mean specifically the mental causation of physical effects, whether or not I always specify this.

[2] This is because orthodox interpretations of Quantum Mechanics are in principle probabilistic. Furthermore, many higher-level sciences also invoke probabilities in their laws and explanations. For more on this topic see Chapter 2.

in the brain and the wider central nervous system. Both the mental state and its corresponding brain state could be conceptualised as a kind. However, I will argue that mental kinds are simpler and more (albeit imperfectly) natural than their corresponding brain states. Because mental states make the more perfectly[3] natural kind, I argue that it is them rather than their corresponding brain states which are the better candidates to feature in scientific laws. If I'm right about this then it seems that mental states must be causally efficacious. This positive branch of the argument is independent of the first branch of my argument in that it can apply to worlds whether they're deterministic or probabilistic.[4] Of course, the two arguments do not preclude each other. Indeed, I hope more generally to put forward an overall world view under which both my arguments are sound and mental causation can consistently obtain.

1.2 Motivation

The reason I think this is a question worth attempting to tackle is because it cuts to the heart of our everyday experience and to the heart of philosophical debates which have raged for centuries. It relates to questions which have been asked since philosophy began; although particularly since Descartes. It

[3]To clarify, my argument will be that mental states form imperfectly natural kinds, but that they are less imperfect than their corresponding physical states. Although a stronger argument that mental states can constitute perfectly natural kinds could be made (see section 9.3.2) this is a stronger claim than I need to make for my argument to hold.

[4]The general argument that causal closure is false can be made in deterministic worlds too. However, I believe it is easier to argue that causal closure doesn't hold in a probabilistic world rather than a deterministic one. On the other hand, my natural kind based argument is independent of this consideration and can be made equally well in either case.

would be extremely strange if our most basic experiences of the world should prove to be illusionary or in some way misleading. I believe if mental causation of physical events doesn't exist then our mental lives do systematically mislead us. This is because I take it to be uncontroversial (and indeed integral to our mental lives) that we all feel as though our mental states are causing us to act. For example, when we're hungry we eat, when our heads hurt we reach for the painkillers and when we're tired we reach for the coffee. However, and perhaps surprisingly, it turns out to be hard to defend our intuitive sense of the world philosophically. If the CEA is correct then it is not our mental states which cause us to act, rather our acts are purely the result of underlying brain states. This would be extremely counterintuitive but just because this would be counterintuitive does not mean that it cannot be the case. It does however lead me to think that if there is a way to preserve our everyday fundamental experiences which is also philosophically satisfying, then that picture is to be preferred.

What does it mean for a theory to be philosophically satisfying? It must be consistent with our best current scientific theories (among many other things). Newtonian physics, while still utilised for some practical purposes, is agreed to be fundamentally wrong. The two front runners of our best current partial scientific theories are now quantum mechanics (though there are many interpretations of quantum mechanics each with different physical implications) at the microscopic level, and General Relativity at the macrolevel. There is currently one largely orthodox philosophical way of metaphysically conceptu-

alising the world: physicalism.[5] Orthodox interpretations of quantum mechanics are fundamentally indeterministic theories of the world. Therefore, as a front runner in our best scientific theories, it seems philosophically astute to think of our world as being indeterministic. This is the background against which I want to analyse the CEA. Is the argument still sound in a probabilistic world? Or do one or more of the premises no longer stand up to scrutiny? I will argue that it is not sound in its original deterministic form in the probabilistic setting but that there may be an analogue probabilistic CEA which more plausibly could be. However, I will ultimately conclude that even this analogue CEA is unsound in probabilistic worlds like ours.

1.3 Kim's Causal Exclusion Argument

As so much of my thesis is framed around Kim's CEA I will introduce it very briefly now so that I can lay out the plan for the rest of my thesis more clearly. However, I will hold off going into full detail until Chapter 6. There are three premises to Kim's argument:

Deterministic Causal Exclusion Argument:

(P1) Causal Closure of Physics
Every physical event has a sufficient physical cause.

(P2) No Systematic Overdetermination
It is not systematically the case that there are multiple minimally sufficient causes

[5] I will define exactly what is meant by this in the next chapter.

of any given event which exist simultaneously.

(P3) Non-Identity of the Physical and the Mental
There are physical states and mental states and these are not identical to each other.

(C) There is no Mental Causation
Mental states cannot be causes of physical events.

In a sentence, if physics is closed then we have a sufficient causal picture of the world which means that, assuming the physical and the mental are not identical, and that we reject widespread and philosophically troubling overdetermination, there is no mental causation of physical events. This leaves mental states as epiphenomenal, at best able to cause other mental states, or even as some way illusionary. This is because it will always be the underlying brain state which is actually bringing about effects despite it seeming to us like our mental states are doing the causing.

My goal for this thesis is to show how one argument against the existence of mental causation is unsound in a probabilistic world (as I assume we live in). The probabilistic analogue of the CEA which I generalise from Kim's deterministic version is a fairly straightforward parallel:

Probabilistic Analogue CEA'

(P1') Causal Closure of Physics

Every physical event has a physical cause which is sufficient to fix its probability.

(P2') No Systematic Overdetermination of Probabilities
It is not systematically the case that there are multiple sets of events which exist simultaneously and that are minimally sufficient to fix the probability of a further event.

(P3) Non-Identity of the Physical and the Mental
There are physical states and mental states and these are not identical to each other.

(C) There is no Mental Causation
Mental states cannot be causes of physical events.

This version would be appropriate for a probabilistic setting where physics fixes, or helps to fix, the probabilities of events occurring rather than sufficing for the events themselves. I generalised the CEA into probabilistic terms because I wanted to analyse the best version of the argument possible. As I am assuming a probabilistic world, it follows that a probabilistic rather than deterministic version of the CEA has the greater chance of success. Indeed, it seems clear that the original deterministic CEA would not be sound in a probabilistic world because the deterministic causal closure principle would not hold.[6] This is not to argue that mental causation exists per se, but rather to show that the CEA is not a successful argument against it, at least as far as probabilistic worlds such as ours go. This is of course far from sufficient for demonstrating the existence of

[6]Although I don't think this speaks too much against the deterministic CEA in the sense that it fails here at something it is not 'designed' to handle.

mental causation.

Before continuing I should note there have been other probabilistic versions of the CEA put forward. For example, Papineau presents a similar argument in "Philosophical Naturalism" (1993) albeit in a slightly different form. Papineau, unlike Kim, starts with the assumption that the mental is causally efficacious. This premise together with a probabilistic causal closure premise shows that either there must be rampant overdetermination or that there is identity between the mental and the physical. In this way, like Kim, he argues for a form of physicalism. Tim Crane (1995) structures the problem slightly differently. He presents the problem as follows;

> "(A) Causes have their effects in virtue of some of their properties
> (B) There is mental causation
> (C) The completeness of physics is true
> (D) There is no overdetermination
> (E) Mental and physical causation are 'homogeneous'" ((1995) p.229)

Completeness of physics here can be understood in either deterministic or probabilistic terms. Crane claims that the only way to reconcile these five premises is to be an identity theorist. He sees Kim's position as seeming to "waver between denying (B) and denying (E)" ((1995) footnote 45, p.229). That is, Crane takes Kim to wavering between denying that mental causation exists and the position that mental causation exists but is different in kind from physical

causation.[7]

The plan for the rest of this introductory chapter is to explain what I mean by a mental state (although I will leave explaining what I mean by mental causation until I have discussed some popular theories of causation). Having done this I will quickly note that I will not be heavily focused at this time on debates about free will although what I have to say will bear relevance to them. I will briefly discuss consciousness, another topic I will not have space to directly tackle but which is heavily connected to this work. I will then lay out the plan and structure for the rest of the thesis.

1.4 Mental States

I will take the existence of mental states for granted for the purposes of my thesis and for space reasons I won't attempt an exhaustive and exclusive definition.[8] There are of course some arguments against their existence[9] but I will not discuss these. So what do I mean when I use the term `mental state', as I shall be doing frequently in the course of this thesis? There are many different kinds of mental state; beliefs, desires, emotions, attitudes, and qualia, of which pain is perhaps the most philosophically utilised example, to name a few. At least some are

[7]See Kim (1984). For Kim, epiphenomenal causal processes are "*apparent* causal relations that are *grounded* in some underlying causal processes" ((1984) p.259). Some such causal processes, including mental causal processes, will also be supervenient causal relations in that they supervene on, and are explainable by, microphysical events.

[8]For example, as I will go on to say, I don't want to get into discussion about which things do and do not have mental states or consciousness but I take it that, contra Descartes, at least some non-human animals do have them.

[9]For example, there have been eliminativist arguments put forward by Churchland (1981).

conscious at least some of the time, or are conscious in some aspect.

There are two paradigm types of mental states: qualia and propositional attitudes. Qualia are usually defined as 'what it's like-ness', for example as Nagel does in "What it's Like to be a Bat" (1974). It is the part of mental life which is essentially experiential. Pain and the experience of 'seeing the colour red' are classic examples although other feelings, emotions and perceptual experiences all have a qualia aspect to them. Propositional attitudes are the mental states agents hold which have propositions or states of affairs as their content. Beliefs and desires such as the belief there is ice cream in the freezer and the desire to eat ice cream, are examples of propositional attitudes.

What is the relation between the mental and the physical? What does it mean to say a mental state has an underlying brain state? The most popular view is that the mental supervenes on the physical. An explanation of this is put forward by Lewis (1986*a*) with his dot-matrix example:

> "A dot-matrix picture has global properties - it is symmetrical, it is cluttered, and whatnot - and yet all there is to the picture is dots and non-dots at each point of the matrix. The global properties are nothing but patterns in the dots. They supervene: no two pictures could differ in their global properties without differing, somewhere, in whether there is or isn't a dot" ((1986*a*) p.14).

Furthermore, many thinkers would be willing to endorse the view that the mental is determined, in a stronger than causal manner, by physical brain states (or

perhaps by the wider central nervous system states of which the brain plays a significant role).[10]

1.5 Free Will

Another important note to make from the outset concerns the debate around free will (and moral responsibility) and how it connects with my work. I raise this substantial issue here for two reasons. First, because the philosophical literature about determinism in particular is often couched in talk about free will. It therefore may be expected that I would engage with this literature. However, although what I have to say will be relevant to the topic of free will, I do not have the space to do it full justice here. For now I focus on mental causation.

Second, is to note that mental causation and free will are very closely related and therefore philosophical discussion of one will inevitably be relevant to the other at least to some degree. To clarify, Kim does not reject the phenomenon of mental causation, rather he avoids the conclusion of the CEA by rejecting the non-identity premise (see Kim (2008)). Therefore, as he rejects the conclusion of the CEA, he is not calling freedom of the will into question.

So why is it that mental causation and free will are so closely related? It is because to have free will plausibly requires of a person that they have causally efficacious mental states as well as an ability to do otherwise. That is, the ability

[10]A notable exception would come from content externalism. If you hold a content externalist view, such as Putnam (1975) then a mental state may supervene on both a brain state and the wider environment. McGinn (1989) also discusses this issue in detail.

to do otherwise by itself is not sufficient for free will. Most theories of free will take it to be necessary that our actions are caused by *our* beliefs, desires and decisions.[11] So it seems plausible to me that in order for our decisions to count as our own, mental causation must exist.

So what I have to say in my thesis will bear on the free will debate in that the CEA stands as a possible counter argument to the existence of free will, even if you take a compatiblist stance. This is because if the CEA is sound and valid then it calls into question the existence of mental causation which is taken to be so central to agents making free choices.[12] So, if I successfully argue that the CEA does not hold, this will be to the benefit of the defender of free will, at least in probabilistic worlds.

To recap, I mention free will here only to highlight its relevance but to put it to one side. Doubtless there are many interesting interconnections between mental causation and free will, but this is a topic to discuss more fully at a later time.

[11] Although some thinkers would argue it is not necessary either, for example see Frankfurt cases as a potential counterexample. Frankfurt's original (1969) counterexample involved two agents, one named Black and one named Jones ((1969) p.835). Black wants Jones to perform a certain action and only if Jones decides not to perform this action will Black intervene to ensure that Jones does. Therefore, Jones could not have acted otherwise. But, Frankfurt claims, so long as Jones acted *because* he decided to and not because Black forced him to, then he can still be held morally responsible. So, Jones is morally responsible despite not being able to do otherwise.

[12] Again, Kim himself rejects the non-identity premise (see (2008)) rather than the phenomenon of mental causation.

1.6 Consciousness and Qualia

The last important opening note I want to make relates to consciousness. What it means to be conscious, what things count as having consciousness and exactly how consciousness relates to mental causation are all interesting questions. However they are beyond the scope of the task at hand. I want to make it clear that my main focus is mental causation and therefore I will be making many assumptions about consciousness from the beginning which I hope to highlight now.

Of course, at least some mental states are conscious or can easily be brought to attention. There are some beliefs which are so run of the mill that you know them without thinking about it (take `London is the capital of England' as an example) which can nonetheless be brought to attention without difficulty. It is possible that there are other deeper unconscious mental states which are harder to bring to conscious attention. This is more plausible with some types of mental states than others however, for example it seems unlikely that qualia can be unconscious. That unconscious mental states can be causally efficacious is unclear but is suggested by some psychological treatment. Such treatments are based on the premise that sometimes, to correct unhelpful behaviour patterns, you must examine unconscious thought patterns that may unknowingly be at the root of them. Causation by unconscious mental states is another interesting topic which will be beyond my scope now. It will suffice for my argument to show that at least some mental states can be causally efficacious. So, as conscious

mental states are easier to argue from I will focus on those.

There is evidence from psychiatric medicine regarding conscious mental states causing physical effects. Take cognitive behavioural therapy as an example. The idea of this kind of treatment is to change behaviour (and mood, although this is uncontroversial as far as the CEA goes as this would be mental to mental causation) by changing thinking patterns. Part of changing an individual's thinking pattern is to make that person realise they may have beliefs, often long standing beliefs, about themselves and their abilities which they did not even realise they had.[13] A distinction sometimes made between such conscious and unconscious mental states is termed as the difference between 'standing' or 'dispositional' and 'occurrent' mental states (see Schwitzgebel (2015)).

I'll use an example, which I'll call 'my sister's birthday' example, to illustrate what the difference between a standing and an occurrent mental state is and how it relates to my work. I would be said to believe, as a 'standing' mental state, the date of my sister's birthday even when I am not currently thinking about it. But it's not until it's brought to attention as an 'occurrent' mental state (that I'm currently experiencing) that you take action based on it. I have known

[13]To give a concrete example; a patient may be anxious because they are not performing well at work. When questioned why they are not performing well they may give a list of reasons such as being too tired or being distracted. If that patient was to correct for all these problems they may still find that they are not happy with their performance. If it is suggested to them that it could be because they hold a belief such as 'I can't do it' they may even disagree. However, by bringing to attention such a belief and countering *that*, they may find an improvement in their behaviour and mood. The improvement in outcome would suggest that it was in fact the unconscious belief which was the cause of the problems in the first place. This would suggest causation by an unconscious mental state, although of course my example as it stands is hypothetical and to become convincing would need empirical evidence to back it up.

for years that my sister's birthday is in September. However, unless I bring that fact to mind[14] I will not act on it and will neither buy a present nor wish her happy birthday. When I remember that I've missed it, this will induce guilt and actions such as panic buying and apologising.[15]

What I hope to suggest with this example and through use of such terminology, is that possibly in some cases, mental states may only be causally efficacious *because* they're occurrent. Given that being conscious is essentially a mental quality, then if such mental states are only causally efficacious because they're occurrent, then they are efficacious because of an essentially mental quality.

One way to circumvent the CEAs conclusion is to identify mental states and physical states (indeed Kim takes such a path). In that case there's no violation of causal closure. I will argue however that qualia at the very least, and more likely all mental states, are something over and above purely physical states.[16] Since consciousness necessarily involves qualia, it stands to reason that consciousness is not a purely physical phenomenon and must be fundamentally mental in some aspect. However you view standing beliefs, I agree with Chalmers when he claims that occurrent mental states involve qualia. He says:

> "It is often hard to pin down just what the qualitative feel of an occur-

[14]And assuming I also have the desire to make my sister happy, the belief that remembering her birthday will make her happy and so on.

[15]Someone could question which belief exactly is it that is causing the action. Is it the belief that her birthday is on x, or the belief that that date has passed, and so on? I'm not sure the exact belief is important to my example as long as it causes my later actions.

[16]See Chapter 6 on the non-identity of the mental and the physical for more of my thoughts on this.

> rent thought it, but it is certainly there. There is *something* it is like to be having such thoughts. When I think of a lion, for instance, there seems to be a whiff of leonine quality to my phenomenology." (Chalmers (1996) p.10)

I take it that it's therefore more plausible that conscious occurrent mental states cannot be identified with physical states than non-conscious or standing mental states. I will therefore remain neutral on the latter and focus on the former. All I need to show to make my case is that there are at least some cases where a mental state causes a physical effect in virtue of being mental.

The CEA is problematic not because it denies the possibility of mental causation per se. Mental states causing mental event does not violate the CEA. What does violate the CEA is mental states bringing about physical effects. It doesn't matter to my argument whether there is mental-to-mental causation[17], or if there are such things as unconscious mental states, so long as it is clear that there are some mental states that are causally efficacious in the physical world in virtue of their being mental (such as by being conscious which is necessarily non-physical) in some aspect. If such causation exists then the CEA must either be invalid or as I will argue, unsound.

[17] Although I assume there is.

1.7 Plan

The broad outline of my thesis will run as follows. Chapter 2 will lay out the groundwork. In Section 2.1 I will introduce the physics which leads me to assume the world is indeterministic as well as what it means to be indeterministic, deterministic and probabilistic. Section 2.2 discusses the question of why I think indeterminism at the quantum or microlevel should scale to the macro-level of everyday. To close the chapter I will also discuss in Section 2.3 what I take an event to be by looking at Davidsonian and Kimean definitions. I will do this because I take events to be the relata of causal relations.

In Chapter 3 I will outline five main theories of causation to help spell out what I mean by the term 'mental causation'. These will be counterfactual theories, probability raising theories, interventionist theories, process theories and mechanistic theories. I discuss these theories in particular not only because they are the current more popular ones but because I also hope to put together a general world view under which the existence of mental causation is consistent. Although I will favour counterfactual and probability raising accounts, I do not think my arguments rely on any one interpretation of causation. This chapter will conclude with Section 3.2 on what exactly I take mental causation to be.

Chapter 4 continues in a similar 'world building' vein as I discuss theories of the interpretation of probability. To that end I will favour a best systems analysis of probability although again I hope that the exact theory of probability one holds doesn't effect my arguments. That said, for my argument to be sound I

do require that objective probabilities exist. Of course this doesn't rule out also holding a subjective interpretation. This is because it is possible for objective and subjective probabilities to coexist, they're not mutually exclusive, at least in many views. So in Section 4.1 I briefly discuss the axioms of probability, classical and logical interpretations. My discussion on these will be brief as they are historical theories which are not widely held today but are helpful for following the development of theories of probability. However there are four theories I will go into more detail on. These are subjective or Bayesian theories, frequentist theories, propensity theories and the Best System interpretation. I end this chapter with Section 4.2, an analysis of which theory of probability I prefer and how it can blend with theories of causation. So by the end of this chapter I will have put forward a sketch my world view under which mental causation can be consistently held to exist.

I will spend Chapter 5 laying out my reasons for thinking that mental causation exists. I make three main arguments; the argument from the Mental Manifest Image (Section 5.1), the argument from evolution (Section 5.2) and inference to the best explanation (Section 5.3). The argument from the Mental Manifest Image is that we have good reason to think mental causation exists from our introspective and phenomenological experience of our mental lives. Given this evidence, we should, all else being equal, prefer philosophical theories which include mental causation rather than rule it out. Alternatively, if a theory does rule out mental causation the Mental Manifest Image will place the burden of proof on that theory to explain why we then have these experiences. The

evolutionary argument is fairly self explanatory in that it argues that the reason we and as far as we can tell, many other species evolved mental lives was because mental causation exists. Lastly, in the section on inference to the best explanation I discuss the placebo effect and some psychological evidence for mental causation. To conclude this chapter, section 5.4 will introduce one potential problem regarding the Mental Manifest Image. After all, couldn't it still hold in a deterministic world where there is no mental causation? To counter this point I can only say that I believe we only have evidence from a probabilistic world (as I take ours to be) and therefore we can only speculate about what mental lives would be like in deterministic ones. This amounts to shifting the burden of proof back onto the denier of mental causation to explain our mental phenomenology as we know it.

Having introduced all the key background parts, I will move on in the next four chapters to lay out the CEA as well as its probabilistic analogue and analyse its three premises in more detail. So, in Chapter 6 I finally lay out Kim's CEA in full detail. I will also discuss the non-identity of the mental and the physical in this chapter, a premise which I shall accept. In Section 6.1 will present three reasons why I accept the premise. The first is Levine's (1983) Explanatory Gap Argument which argues that even if we stated every physical fact there is to know about pain, we would still be leaving something very important out of its explanation if this is all we talked about. The second is Nagel's (1974) argument from "What it's like to be a Bat". Finally I will mention Frank Jackson's (1986) Knowledge Argument about an unfortunate Mary locked in a black-and-white room.

In Chapter 7 I examine the second premise of the CEA; that there is no systematic overdetermination. In section 7.1 I go through a few arguments as to why overdetermination should not be philosophically worrying from Sider and Bennett. Section 7.2 shifts to specifically probabilistic overdetermination. Section 7.3 then revisits Sider's and Bennett's arguments to see if they are still sound in a probabilistic setting. I will argue that these arguments can provide a way to reject the CEA in either deterministic or probabilistic worlds even if the reader does not find my other arguments (particularly on causal closure) convincing.

The last two substantial chapters deal with my original arguments. Chapter 8 discusses the first premise of the CEA; the causal closure of physics where I will put forward my reason for thinking causal closure (in both its deterministic and probabilistic guises) doesn't hold in probabilistic worlds. Section 8.1 introduces what causal closure is and provides some arguments for it. I will argue though that we hold causal closure as a result of physicalistic bias. What is held as a result of a bias should be examined to see if we actually have good reasons for believing it. I also consider an argument from Bishop (2006) that causal closure holds only as a typicality or ceteris paribus condition. I then argue in section 8.2 that because we live in what I assume to be a probabilistic world, then the causal sufficiency of physical causes no longer holds. This leaves room for the mental to 'top-up' or indeed the lower the probability of a physical event occurring. I will put my argument into formal terms to add further clarity. I close the chapter by putting my reasons for why I think my view of mental causation

should be preferable to those views which argue it doesn't or can't exist.

Chapter 9 is when I put forward my positive argument in favour of mental causation based on natural kinds. I begin in section 9.1 by introducing the concepts of natural kinds and scientific laws, particularly Lewis' view of perfectly and imperfectly natural kinds (see Lewis (1983)). I also elucidate the intimate connections between the two and with causation. Namely, that natural kinds are taken to feature in laws which underlie causal relationships. Therefore, if mental properties can be considered as natural kinds then they can be rightful candidates to feature in laws. This means they would have causal efficacy even over physical events. I argue that mental properties can be thought of as natural kinds and what's more they can be considered as more natural given the very conjunctive nature that any physical brain kind would have. I also discuss in section 9.2 the debate held between Kim ((1992), (1993*b*)) and Fodor (1997) on mental states and multiple realisability. I also focus here on Yablo's (1992) paper "Mental Causation" and argue that his solution to the problem of mental causation is wrong.

Chapter 10 closes the thesis by arguing that while the deterministic version of the CEA may or may not hold in a deterministic world, neither the deterministic version nor the probabilistic version I put forward hold in probabilistic worlds. Therefore, it cannot stand as a counter argument to the existence of mental causation in probabilistic worlds such as ours. Section 10.1 touches on how what I have to say about mental causation can also apply more broadly to the special sciences. I will in section 10.2 also briefly touch on some unresolved

issues such as where my theory sits in relation to physicalism and whether I come to some peculiar conclusions. I will argue that even if my thesis does lead to what could be considered a peculiar conclusion this is not sufficient reason to dismiss it. In the end, I feel my arguments at the very least place the burden of proof back onto the opponent of mental causation to explain how we can make sense of the Mental Manifest Image given that I think a coherent world view can be put together which can accommodate mental causation.

Before going any further it is now essential that I explain exactly what I mean by such key terms as 'determinism', 'probabilistic' and 'mental causation'.

2

GROUNDWORK

"Do I dare to eat a peach?"

The Love Song of J.Alfred Prufrock -

T.S. Eliot (2010)

While TS Elliot's protagonist ponders, paralysed by indecision, whether he dare eat a peach, perhaps he should have asked himself `*can* I dare to eat a peach'? Do we have any power over our choices; are our mental states ever causally efficacious? My argument will be that the CEA works only on a deterministic[1] picture of the world, by virtue of the fact that its first premise, the causal closure of physics, relies on determinism to hold. Orthodox quantum me-

[1] For example a Newtonian picture of the world, although interestingly Newtonian physics may not actually be deterministic, see Norton (2008).

chanical pictures of the world are fundamentally and in principle probabilistic. Therefore, supposing that we do in fact inhabit some kind of quantum mechanical world, which is what our best current theories and evidence suggest, what consequences does this have for the CEA?

Before beginning substantive philosophical work, groundwork needs to be laid by way of setting out clearly the views and terms I will be using. First, in section 2.1 I introduce physicalism and then define key terms such as 'determinism', 'chance' and 'randomness'. I then move on to discuss orthodox interpretations of Quantum Mechanics.[2] In section 2.2 I will discuss whether it's possible to use quantum mechanic phenomena as evidence for a probabilistic universe given the problem of scaling micro-phenomena to the macro-level. Lastly, in section 2.3 I will introduce two of the most commonly held philosophical theories of events which are Davidsonian and Kimean theories of events. I do this because I take events to be the relata of causal relations.

2.1 Physics and Physicalism

Newtonian physics, while still utilised for some practical purposes, is generally agreed to be fundamentally wrong when it comes to describing how the world really is. The two front runners of our best current partial physical scientific theories are now quantum mechanics at the microscopic level, and General

[2]I will not be discussing non-standard interpretations as they are not always probabilistic. See for example the 'Many Minds' interpretation (Albert & Loewer (1988)) and Everett's 'Many Worlds' interpretation (Everett (1973)).

Relativity at the macro-level. There is currently one largely orthodox, though heavily debated, philosophical way of conceptualising the world: physicalism.

Orthodox quantum mechanics is a fundamentally indeterministic theory of the world.[3] I will explain further exactly what is meant by terms such as 'indeterministic' and 'probabilistic' after introducing what is meant by physicalism. As there are many different definitions of physicalism it will be important to clarify which definition I will use. I hope the type of physicalism which obtains under this definition will be relatively uncontroversial. I will define it as follows:

(Phys) Everything supervenes on the physcial.

In *Mind in a Physical World* (1998) Kim defines mind-body supervenience, as follows:

> "Mental properties supervene on physical properties in the sense that if something instantiates any mental property M at t, there is a physical base property P such that the thing has P at t, and necessarily anything with P at a time has M at that time." ((1998) p.39)

This definition of physicalism is minimal he claims, because he takes it to be the minimal commitment to which every type of physicalist must agree. In this case the relation between the mental and the physical is supervenience, so there is no possible change in the mental without some change in the physical. For

[3] I will briefly discuss Superposition and Heisenberg's Uncertainty Principle in order to explain why this is the case below.

Kim therefore, physicalism only holds if mind-body supervenience holds. I will generalise Kim's definition so that it applies to all properties, not just mental ones.

2.1.1 Determinism, Non-determinism, Probabilistic Indeterminism and Indeterminancy

Before going any further I want to avoid potential conceptual confusion which could arise from terminological similarity. I will set out here how I use the terms 'determinism', 'non-determinism', 'indeterminism' and 'indeterminacy' for the purposes of my thesis. I'll start by defining determinism (D) as;

(D) Every actual event is necessitated by the initial conditions of the world given the complete set of the laws of physics.

Non-Determinism (ND) as I shall use the term is simply a denial of determinism. Thus;

(ND) Determinism is false.

However, there are many different ways a world can be non-deterministic. For example it could be probabilistic, random, utterly lawless or a blend thereof, (I disambiguate the concepts of chance and randomness below in section 2.1.2). I will therefore be using ND as something of a catch all term. As noted above,

orthodox interpretations of quantum mechanics lead us to believe that our world is not strictly deterministic and therefore ND is true of it. But, it is important to note that an orthodox quantum mechanical world is ND in a particular way, namely it is probabilistic. As this is the way that we think the world actually is given our best current physical theories, this is the particular variety of ND that I will be focused on.

Probabilistic indeterminism (ID) then is the thesis that;

(ID) The world is fundamentally probabilistic. Given an initial set of conditions and a completed set of the laws of physics a non-trivial probability distribution[4] could be derived over possible future events.

This is not to say that probabilities cannot alter over time; there will be some events whose unconditional probabilities differ from their probability conditional upon certain possible initial histories of the world. Probabilities at later times can be derived from this first distribution given by the initial conditions and laws.[5]

So far I've been disambiguating concepts which are different shades of the

[4]In other words one which does not just return 1s and 0s as would be the case on a deterministic picture.

[5]Although this is not an entirely uncontroversial point, see for example Ismael (2008). Ismael's claims that "because for any finite string of events, there is always the *possibility* of events to follow ... there is no general way of turning a distribution over finite strings of future events into a distribution over total histories" ((2008) p.301). Without knowing the distribution of probabilities of total histories, we can't assess the accuracy of theories of chance themselves. She gives an example from the actual frequentist theory of probability (for more on this theory please see section 4.1.4). The actual frequentist can never say with certainty what the probability of a given event's occurring is because as long as there is a possibility of future event's occurring, the frequency is always subject to change.

same theme. Indeterminacy (IDy) on the other hand is quite a different concept. Although it's distinct, I mention it here to disambiguate quantum indeterminism from quantum indeterminacy. Indeterminacy is the thesis that;

(IDy) There is no fact of the matter.

An illustration of this comes in The Uncertainty Principle (Heisenberg (1927)). If a particle is in superposition this means that it has no definite state until it is measured. The Uncertainty Principle states that once the position of a particle is known its velocity becomes indeterminate, that is, there is no fact of the matter about the precise velocity of the particle. Rather than being a failure of our observational abilities, this has to do with the indeterminacy of the velocity of the particle in this case. However, if the velocity of the particle was then to be measured there would be a *probabilistic* fact of the matter of its velocity (and an indeterminate one of its position). So when an observer measures a particles velocity, thereby collapsing the wave function and causing it to take a determinate value, she causes its position to fall into superposition - in which it has no determinate location. If she then measures the position of the particle, causing it to have a determinate location, she causes the particles velocity to fall into superposition, where recall, there is no fact of the matter. The result of the collapse of the wave function is probabilistic. Because of this, and given that there is no way out of this cycle, the quantum mechanical world is (at least on this interpretation) in principle probabilistic and in some ways indeterminate.

2.1.2 Chance and Randomness

Often in everyday speech the terms 'chance' and 'randomness' are used interchangeably. These are philosophically distinct concepts however. Following Anthony Eagle's *Chance versus Randomness* (2016) I will make the distinction between process and product. Chance lies in process whereas randomness applies to products of processes. Further, predictability will be a factor in distinguishing the two. Chances are probabilistic, whereas truly random events are strictly unpredictable.[6]

Compare the process of flipping an unbiased coin to the product of a random sampling. The outcome of the coin toss is 50/50 coming up heads or tails. As I've been stressing, processes at the quantum level are intrinsically probabilistic or chancy. Random samples on the other hand need not be the result of chancy processes. Eagle gives the example of using a simple heuristic to collect demographic details. His example is to take the personal details of babies born at any time ending in seven. Because (we assume) there is no connection between the exact time of peoples' birth and their other pertinent demographic details, this simple process can produce random results. There will not be any pattern within the data collected by which you could predict the results. Let's return once more to the coin toss. While you would know for each toss the probability of it coming up heads and tails (that is 0.5), the actual

[6]Indeed one definition of a random sequence is that you would need an algorithm at least as long as the list itself in order to specify the sequence (Eagle (2016)). This is because there is no pattern whatsoever by which the algorithm could compress the sequence. Therefore it must just give a brute list.

sequence which results will be random. This is extremely brief but should be enough to highlight the differences between chance and randomness.

So why is this distinction important for my work? The kind of probabilistic world I am interested in for the purposes of this thesis are chancy but not merely random. That is, I'm not interested in worlds where only the outcomes are random, but ones in which the causal processes that produce them are themselves chancy.

Why am I not interested in truly random worlds? That is those worlds which are either completely lawless or worlds in which causal processes are not deterministic but not probabilistic. Carl Hoefer (2016) outlines another kind of world which is deterministic and merely looks by its product to be random; he calls this kind of world "deterministic chaos". As I've said I don't want to discuss random worlds principally because current orthodox theories of quantum theory suggest our world is not like this and I want to focus on the closest picture we can get to of our world. However, more than this, in a truly random world it might be hard to see how our mental states could be bringing about effects in any meaningful way given that no leading theory of causation accommodates indeterministic but not probabilistic causation. In other words, in truly random words, it would be hard to see how anything as systematic as causation, let alone mental causation, could be said to obtain.

2.2 Micro to Macro Question

There is an important question about why I think that just because quantum mechanics is a fundamentally and in-principle indeterministic theory[7], I should assume that this indeterminism is global or that it scales to the macro-level. This is a fair question as the indeterminism of the world is a key assumption I make in order for my argument to work. Further, for my argument to work the world must be truly indeterministic and not merely epistemologically indeterministic. I will start the section by briefly considering the problem of pessimistic meta-induction, then go on to explain what is meant by the terms 'micro' and 'macro'. I will discuss the indeterminism of the micro-level. I then discuss the issue that it's not obvious that these micro-level phenomena scale to the macro-level. I will then present the positive reasons we have for thinking that the world is indeterministic based on micro-level phenomena. I will conclude that my assumption that the world as a whole is indeterministic is therefore a safe one.

Before I explain what is meant by the micro-level and the macro-level there is a problem which could be raised here about relying on our current scientific theories for knowledge about how our world really is. This problem, known as pessimistic meta-induction, argues that all of the scientific theories we've held until now have been proved wrong, so we have good reason to think that our current theories will also be proved wrong by future science.[8]

[7] At least on the orthodox interpretation that again, I am assuming.
[8] See Wray (2015) for a variety of more detailed definitions of pessimistic meta-induction.

It may seem like a compelling argument, given that we have no access to future physics. However, we only ever have current and past physical theories to work from. Furthermore, it could be hoped that, while ultimately wrong, our current theories are getting closer and closer to the truth. In other words, while they may be wrong, that's not to say they are completely wrong and that nothing about the world can be learned from them. Lastly, there may be some core concepts which have 'stood the test of time' or which span all our current theories, and we can be more confident that these will continue to feature in future physics. Vincente discusses this problem in his paper "Current Physics and the Physical" (2011). So, despite the pessimistic meta-induction argument, current science is our best guide as to how the world really is. It is therefore prudent to consider them when carrying out philosophy albeit with the understanding that they are always subject to change.

So, what exactly is meant by the terms 'micro' and 'macro'. Micro-level entities are the smaller entities known to science, not visible to the naked eye. These kind of entities are currently dealt with by quantum mechanics and are given by the Standard Model of particle physics. The list of micro-level entities include things such as atoms or (even lower) quarks[9].

Why is it not clear that probabilistic phenomena scale to the macro-level? Just because micro-level physical entities behave in in-principle indeterministic ways, that does not mean that macro-objects composed of such entities as a whole also exhibit this trait. In fact, there are some reasons to think this is not so. Take

[9]See for example Chapter 1 in Cottingham and Greenwood (2007).

indeterminacy, another quantum mechanical trait which does not seem to scale to the macro-level. Objects in our everyday life do have a determinate location and velocity even if the particles these objects are comprised of do not. If a bird is flying through the sky we can calculate both its momentum and position at any given moment.

Another phenomenon which seems to appear at the micro-level but not the macro-level is entanglement. Entanglement is the phenomenon by which the state of one particle is correlated with that of another (or others) even when the two are vast differences apart.

In some situations entaglement can give rise to some strange phenomenon. Jeffrey Bub describes this thus;

> "After the particles move apart, there are `matching' correlations between both the positions of the two particles and their momenta: a measurement of either position or momentum on a particular particle will allow the prediction, with certainty, of the outcome of a position measurement or momentum measurement, respectively, on the other particle. These measurements are mutually exclusive: either a position measurement can be performed, or a momentum measurement, but not both simultaneously. The subsequent measurement of momentum, say, after establishing a position correlation, will no longer yield any correlation in the momenta of the two particles. It is as if the position measurement disturbs the correlation between the

momentum values, and conversely." (2017)

Much about this phenomenon is still mysterious to us. However, it is not clear that objects at the macro-level exhibit this behaviour.

Interestingly, it has been noted by Carl Hoefer (2016), that although quantum theories are thought of as the 'friendliest' towards indeterminism, that they might not actually be. This is because there is an interpretation of quantum theory called Bohmian Quantum Mechanics which is fully deterministic.[10] This is problematic because, as Hoefer says, Bohmian Mechanics is "empirically equivalent to standard Copenhagen (orthodox) QM" (2016). What this effectively means is that it is impossible to chose between the two theories at least in so far as explanatory power is concerned. However, as stated, I will assume that the orthodox interpretation is correct.

Are there any reasons to think that, contra previous arguments, such indeterministic phenomena can and do scale? One argument for this rests on the idea that there's a difference between what actually happens and what could possibly happen. Macro-objects are made up of micro-level particles and these particles do entangle. There is therefore the remote possibility that macro-objects could exhibit very strange quantum behaviour. For example is it theoretically possible for my coffee mug to fall straight through this solid wood desk; a phenomenon known as Quantum Tunnelling which has "no counterpart

[10]Bohm got around the measurement problem by adding a "guiding equation" which means that particles do in fact have determinate positions and velocities (see Bohm (1952)). Hence, Bohmian Mechanics is a determinate theory.

in classical physics" (Bowman (2008) p.125). Needless to say this phenomenon is "negligible at the macroscopic level" (Flowers et al. (2016)). This isn't to say that it doesn't happen more often with electrons at the micro-level, indeed "tunnelling is frequent on the nanoscale" (Flowers et al. (2016)). For a macro-object to show this behaviour all the sub-atomic elements within it would have to tunnel at the same time which explains why the chances of this happening are so low. It's logically (and metaphysically) possible, just spectacularly improbable. The chance is nevertheless greater than 0. The real question here is therefore, what does this actually amount to? Macro-objects, when looked at from this angle are technically indeterministic creatures but next to never act like they are. This is not the kind of indeterminism which I require for my argument to go through.

That said, the previous paragraph has been talking purely about physical states. How mental states would slot into this picture is far from clear. Maybe they could act more like the micro-level? Given that it remains far from clear how exactly the brain and nervous system give rise to mental states, it is possible that small quantum mechanical events in the micro-level of the brain could have a more significant effect on the mental states they give rise to rather than the macro-objects they make up. This is purely speculative here however and it's not clear to me why supervening upon rather than being constituted by would make a relevant difference.

A second argument in favour of the world being indeterministic may even come from macro-level theories of the higher-level sciences. The macro-level is the level of everyday or massive objects. These fall under the domain of our other

best current scientific theories; special and general relativity in physics and other special sciences such as biology and chemistry. Hoefer (2016) makes the case that these theories are actually more open to indeterministic interpretations. Furthermore, statistical mechanics is probabilistic. What's more, Loewer (2001) argues that probabilities within statistical mechanics are objective, even on the assumption, taken from Albert (2000) that fundamental dynamics are Newtonian.[11] Even Newtonian Mechanics[12], usually considered the most straightforwardly deterministic of the theories, may be indeterministic.

That said, if it does turn out to be the case that macro-objects can or do exhibit strange quantum mechanical behaviour then it cannot be the case that Newtonian mechanics holds for these objects. This is because there is no way for Newtonian mechanics to adequately explain such observations. This would suggest that the final unified theory of physics (should such a theory be within our grasp) would have to be in principle indeterministic in order to capture the weird quantum mechanical behaviour at both the micro- and macro-level. But, a world scaled up from those probabilities may well not be indeterministic in the way I require (that is, probabilistic as opposed to random). Is there any other reason from the higher-level sciences that can be given to suggest the world is probabilistic in this specific way? Perhaps we can call on them to give us some

[11] Briefly, Loewer modifies Lewis' Best System Theory (see section 4.1.6 for more on this) of laws and probabilities so that it includes a probability distribution over the initial conditions of the world ((2001) p.618). He claims that, as it forms part of the best system, such a distribution counts as a law and as such is objective ((2001) p.619).

[12] This is a best current theory only in the sense that it still has some practical applications. I mention it here for completeness and because it strengthens my point that it is at least not obvious that theories don't suffer from failures of determinism.

verification of the idea?

The two final current physics theories to consider are special and general relativity. Because mental causation and more broadly mental states supervene on the physical brain (and wider central nervous system[13]), quantum mechanical theories would appear to be the more relevant theories. So, it could be argued then that even if relativity theories are probabilistic, then, because we're dealing with phenomena at the neuronal level, this outcome is not pivotal.

Simple interpretations of special relativity are in fact deterministic. However, as the theory becomes more sophisticated then failures of determinism can arise (see, among others, Earman (1986) and Earman and Norton (1987)). General relativity on the other hand is more straightforward. In Hoefer's words; "the simplest way of treating the issue of determinism in GTR would be to state flatly: determinism fails, frequently, and in some of the most interesting models" (2016). Hoefer (2016) argues that one example of how determinism may fail in General Relativity is singularities. There are various different kinds of singularity, perhaps the most well known being 'simple' black holes. Outside the black hole determinism holds, but within them it breaks down. However, naked singularities cause even more problems. They are areas of space which act like black holes (or white holes) but which do not have an event horizon as a barrier between it and the rest of space. Similar to the Newtonian problem of 'space invaders' this means that things can pop into space from naked singularities without any

[13]If content externalism is true then mental states also supervene on a subjects environment. However, it's also not clear how general relativity theory would help explain these states either.

way of predicting such an event. This is a failure of determinism. To summarise neatly, again in Hoefer's words, "the prospects for determinism in GTR as a mathematical theory do not look terribly good" (2016). The question of whether this scales to the level of mental causation remains unanswered though.

So, there are reasons to think that macro-level physical theories are not straightforwardly deterministic as there may be cases where determinism breaks down within them. This on its own is not sufficient for my purposes though. The determinism must break down in the 'correct' way. A random world would likely not exhibit the kind of phenomenon I am trying to capture. That said, such a chaotic world would likely not exhibit any law like activity. It's not obvious that mental causation could exist in such a world so whether or not the CEA holds there becomes somewhat moot. Therefore I think we have reason to think that the world we live in is not utterly random[14] and neither are the theories I have been discussing above. Furthermore, it does look as though in the future, however GTR and QM are reconciled, it is likely to be the case that in principle indeterminism will be retained in order to capture the quantum phenomena we observe. This is because it is hard to see how (or even if you can) accurately describe an indeterministic phenomena (such as wave collapse) using the tools of a deterministic theory.

The next argument I will put forward in defence of indeterministic micro-phenomena scaling to the macro-level comes from Anscombe. In *Causality and Determinism*

[14]In the sense that so many undetermined events occur frequently enough that few reliable predictions can be made.

((1971) p.24) she put forward an example of quantum indeterminism causing an indeterministic outcome in the macro-world similar in spirit to Schrödinger's Cat thought experiment. In it, a bomb has been rigged to a Geiger counter. The Geiger counter will only activate if an atom next to the bomb decays. The decaying of an atom is a quantum event, which means that it is also an in principle indeterministic event. Therefore, whether or not the bomb explodes is also an in principle indeterministic event. Given that bomb explosions are events at the macro-level, this demonstrates how indeterministic micro-level phenomena can scale.

It could be objected that Anscombe's example is a bit contrived. A perhaps more convincing reason to treat biology and chemistry as indeterminate come from the fields of quantum biology and quantum chemistry. Both these fields have a focus on molecules (see, among others, Fleming (2014), Richards (1983) and Marais et al. (2018)). As Marais et al say "all living systems are made up of molecules, and fundamentally all molecules are described by quantum mechanics" ((2018) p.2). Furthermore, all biological entities are "constituted of ions, atoms and/or molecules whose equilibrium properties are accurately determined by quantum theory" ((2018) p.2). Photosynthesis, vision and respiration are all processes which Marais shows have previously been modelled in fundamentally quantum mechanical ways. This shows that fruitful progress has been made by sciences other than physics which focus on the indeterminacy of both chemistry and biology based on their reliance on quantum mechanical processes.

Finally, it has been argued that whether or not indeterministic micro-phenomena scales to the macro level or not is irrelevant as there are still objective probabilities within higher-level sciences. Cohen and Calender (2009) make such an argument in their better best system analysis of laws. I discuss their arguments in detail in section 4.1.6. Barry Lower also makes this argument in "Determinism and Chance" (2001).

To recap, there is reason to think that not only is quantum mechanics potentially an indeterministic theory, it appears as though macro-level theories may be too. I therefore think that we have enough evidence to make the assumption that the world as a whole is indeterministic at the micro-level. Furthermore, whether or not this indeterminism itself scales to the macro we may have other reasons to think that the macro is more than epistemologically indeterministic.

2.3 Events

Before going into any theories of causation there are two important questions which need answering.[15] What are the relata of causation and what are events? These questions are of course linked if you take the relata of causation to be events as I do. It's important to be clear, because what an event is will bear on which events we can genuinely take to be the relata of mental causation. A Davidsonian event for example, may be too coarse grained to meet the criteria for true mental causation. To be a case of genuine mental

[15]I follow Casati & Varzi (2015) and Schneider (2017) in this section

causation it is not enough for an event with a mental property to do the causing. It must be that the event causes the effect *in virtue of* its mental property.[16] I will now lay out a Davidsonian and a Kimean view of events as these are two popular theories of events.[17]

2.3.1 Davidson

Davidson's latest theory of events is coarse grained. While he used to identify events by their cause and effects he later individuated events by their spatio-temporal location.[18] So two events are identical if they share the same space-time region. Therefore for Davidson `the stabbing of Caesar' and `the killing of Caesar' are one and the same event under different descriptions.

In *Mental Events* Davidson (2001*b*) combines three principles which he takes to be true. They are (1) The Principle of Causal Interaction, (2) The Principle of the Nomological Character of Causality and (3) The Anomalism of the Mental. The first principle states that "at least some mental events interact causally with physical events" ((2001*b*) p.208). The second principle states that "where there is causality, there is a law" ((2001*b*) p.208). By this, Davidson means that wherever events are causally related, "they have descriptions that instantiate a law" ((2001*b*) p.215). Importantly, this doesn't mean that such events will instantiate a law under every description. He takes the first two principles to be

[16]Furthermore, as I've mentioned, to be the type of mental causation that violates the CEA, it must be a mental cause of a physical effect.

[17]Furthermore as I am specifically assessing Kim's CEA it is incumbent upon me to fully understand his position.

[18]See Davidson (2001*b*).

assumptions.

The third principle states that "there are no strict deterministic laws on the basis of which mental events can be predicated and explained" ((2001*b*) p.208). On the face of it these principles are inconsistent. If it's assumed that mental and physical events causally interact, and that with causality comes laws, how then can there be no strict psycho-physical laws?

He dissolves this inconsistency by arguing that the physical and mental events are identical to each other and as such share spatio-temporal location. So for example, pain is identical to the firing of C-fibres. Therefore, every mental event is also a physical event which does have laws associated with it. To be the kind of cause I am interested in though, it must be in virtue of the instantiation of a mental property that any event (individuated by its spatio-temporal location) brings about an effect. To clarify, this is not Davidson's view, rather an adaptation of it. For Davidson, the mental can be causally efficacious but "given (his) concept of events and of causality, it makes no sense to speak of an event being a cause "as" anything" ((2005) p.188). Rather, "if causality is a relation between events, it holds between them no matter how they are described ((2005) p.189).

Davidson's "Thinking Causes" (2005) was written as a response to a critique made by Kim (1993*b*) (among others) whereby Kim argued that "under Davidson's anomalous monism, mentality does no causal work" ((1993*b*) p.269). This is because, Kim claims, on Davidson's view, you could alter all the mental proper-

ties of a causal system, or even completely remove them, and it would make no causal difference to that system due to the lack of psychophysical laws. Furthermore, Kim claims, Davidson can't claim that changing an event's mental properties would thereby change its physical properties as this would imply the existence of a psychophysical law. To paraphrase Kim's reading of Davidson; if mental events are only ever causally efficacious in so far as they are physical events, then it is irrelevant that they are mental events and you may as well discard the notion of 'mental'. Thus Kim argues Davidson's Anomalous Monism actually comes "perilously close to out-right eliminativism" ((1993*b*) p.271).

Similarly, Honderich (1982) points out that not all properties of an object are causally relevant in all contexts. He gives the example of weighing some pears. The fact that the pears are green is in no way causally relevant to the scale pointer moving ((1982) p.61). The pears and the pointer stand in a lawlike relation but not because of every property of the pears and every property of the scales. Thus we must update The Principle of the Nomological Character of Causality to The Principle of the Nomological Character of Causally-Relevant Properties. This raises a tension though with the idea that the mental can cause anything as a mental event. To do so it must do so in virtue of its mental properties, but then there must be psychophysical laws linking a mental property to a physical one. So to retain Anomalous Monism is to render the mental epiphenomenal *as* mental (Honderich (1982) p.63). Claiming that a mental event causes a physical event but only so far as it falls under a physical description doesn't save the mental from epiphenomenalism. This is because the

mental in such a case acts as the greeness of the pears did in the weighing example. The mental properties were "necessary to the event's being the event it was, but not necessary to the event's being the cause it was" (Honderich (1982) p.61). If Davidson's account of events cannot be modified to remove the epiphenomenalist worry then it will not suit the needs of my thesis. Alternatively, if you believe with Davidson that "redescribing an event cannot change what it causes, or change the event's causal efficacy" ((2005) p.189) then you would consider the event a mental cause in the correct sense.

There is one last counterexample I will mention to Davidson's view provided by Davidson himself in response to Lemmon (1996).[19] The example is designed to show that there can plausibly be two separate events which share the same spatio-temporal location. Imagine a metal ball which is being heated up at the same time as rotating. Say you stipulate that the warming from the heating process is happening over and above that caused by the rotating. Intuition then suggests that there are two separate events occurring. That is, the event of heating and the event of rotating, rather than one event of heating-rotating. But, given the molecules which are being heated and the molecules which are collectively being rotated share the same space-time location, it seems unclear why this is not just one Davidsonian event.

On a Davidsonian world view, a mental event will always also be a physical event. His account of events will not suffice for my purposes if you think this renders the mental epiphenomenal. However, if you are not persuaded by

[19]The counterexample can be found in (Davidson (2001*a*) p.178).

these arguments (although I admit that I am) then Davidson's conception of events will work for my purposes. His is not the only conception though, so now I will move onto Kim's view of events which may be more palatable to those who are convinced by epiphenomenalist worries about Davidson's account.

2.3.2 Kim

Kim (1993*c*) has a much more fine grained conception of events. He views them as a triple <P,o,t>. That is, the instantiation of a property P by an object o at a time t.[20] It is more fine grained as it individuates events by the property they instantiate and therefore the property of 'stabbing Caesar' and the property of 'killing Caesar' produce different events (at least on Kim's view of properties). This is because the property of being a stabbing and the property of being a killing are different properties. Not all stabbings are fatal and there are many more ways of killing people than by stabbing. Kim argues that despite the fact that in Caesar's case the stabbing did amount to a killing, they are still not the same event. He presses this by making the point that "to explain Brutus' killing Caesar (why Brutus killed Caesar) is not the same as to explain Brutus' stabbing Caesar (why Brutus stabbed Caesar)" ((1966) footnote, p.232). Therefore, a Kimean view of events allows mental and physical events to be non-identical because an object can instantiate both a mental and a physical property.

So in summary, when I talk about events bringing about effects I specifically

[20]Some people prefer to call such triples facts or states of affairs, for example Casati and Varzi discuss this in (2015), in which case my view can be modified to take these as the relata of causation instead.

mean a (Davidsonian) mental event bringing about a physical event in virtue of its mental property or because it is a genuine (Kimean) mental event (unless otherwise specified). Insofar as this criteria is met, it doesn't make any difference for my purposes which theory of events is endorsed. As long as the cause is either a mental event in the Kimean sense, or is a Davidsonian event which is only efficacious in virtue of its mental property, then this suffices for my argument. I will use both of these locutions throughout but anytime I do, the other is substitutable. With these clarifications made I will move onto introducing some of the most popular theories of causation.

3

Theories of Causation

"But whether or not this was cause and effect I couldn't make out"

Cain's Jawbone - Edward Powys

Mathers (1934)

Other than 'not correlation', what is causation? As this is a thesis on mental causation, it will be vital to make it clear what mental causation *is*. To this end I will now introduce and discuss some major theories of causation.

In "Two Concepts of Causation" Ned Hall (2004) claims that causation comes in "two basic and fundamentally different varieties". The dependence concept on the one hand and the production concept on the other. Difference making views include probabilistic theories and manipulation views along with coun-

terfactual theories. Production theories include process views but also include mechanistic views.[1]

Hall argues for this duality by claiming that certain intuitive theses about causation conflict with each other. The theses are: Dependence, Omissions, Transitivity, Intrinsicness and Locality. The Dependence thesis holds that counterfactual dependence between distinct events is sufficient for causation. Omissions states that failures of the occurrence of events can cause and be caused. Transitivity means that if *a* causes *b* and *b* causes *c* then *a* causes *c*. Intrinsicness is the thesis that the causal structure of a process is determined by its intrinsic character (along with laws). And lastly, Locality states that causes and their effects are connected spatiotemporally by continuous causal intermediates.[2] Roughly speaking, Hall claims that dependence forms one kind of causation and transitivity, intrinsicness and locality are hallmarks of production. Sometimes the two come apart which is why dependence and the other theses can sometimes come into conflict. Furthermore Hall argues that treating the two as separate concepts can help us explain problem cases, such as causation by omission, which traditional unifying theories of causation have struggled with.

While the two often coincide there can be cases of dependence without production and production without dependence. For dependence without production Hall gives the example of Billy, Suzy and the Enemy fighter jet ((2004) p.241). Suzy is on her way to bomb an enemy city with only Billy to escort her.

[1] They also include information based views but I will leave these to one side.
[2] See ((2004) pp.225-226) for Hall's definitions of these terms.

Enemy fighter jet would shoot Suzy down and prevent the bombing but fails to do so because Billy shoots Enemy down first. Billy in some sense is a cause of the bombing in that he helped prevent a preventer. But he is not even anywhere near the bombing at the time it occurs (Suzy having flown away leaving the dogfight behind her). So there is a sense in which he does not cause the bombing. Hence a case of dependence without production.

Conversely there are cases of production without dependence. Hall again uses Billy and Suzy to illustrate, giving the standard bottle smashing example ((2004) p.235). Billy and Suzy both throw a rock at a by-standing bottle. Because Suzy throws first, her rock hits the bottle and smashes it before Billy's can. However, as Billy had also thrown his rock, his throw would have been the cause of the bottle's smashing had Suzy thought better of throwing hers or had her rock missed its target. Thus the bottle's smashing has no causal dependence on Suzy's throw (because the bottle's smashing does not counterfactually depend on Suzy's throwing) despite the fact that her throw is the actual cause of the bottle's smashing. This is a case of production causation without counterfactual dependence. Suzy produces the smashing by throwing even though there is no dependence given that Billy's stone would have done just as well.

One argument which could be made against Hall is that his view isn't parsimonious. On the parsimony argument it would be better to have one unified conception of causation rather than two if at all possible. Hall acknowledges this argument and defends his thesis against parsimony type arguments. Methodologcially he says, we have reason to prefer a unified conception of causation,

but we have no a priori reason for thinking this is so ((2004) pp.254-255). Many attempts have been made to produce an unified account of causation that can deal with the wide variety of problem cases and counterexamples. Given that there still is not a fully satisfactory theory goes to show, so Hall argues, that we should broaden our thinking away from unified conceptions. There is after all no a priori reason to think metaphysics should be simple much though we'd methodologically prefer it to be so.

It has been speculated by some, that all theories of causation 'bottom out' or fundamentally rest on difference making concepts. Hitchcock made such arguments about process accounts (particularly Salmon's and Dowe's accounts.) for example in (1995), (1996) and (2004*a*) where he argues that Salmon's account cannot do without the notion of counterfactual dependence. A similar argument has been discussed in relation to mechanistic theories of causation but I will defer discussing this until section 3.1.5 where I introduce these theories. So it may be possible to do without production type accounts in favour of difference-making ones or it may be possible that difference-making is the more fundamental type of causation. However, these thoughts remain highly speculative and is not a requirement for my arguments that this be the case.

Rather, I want to remain neutral on the issue of whether there are two different types of causation and by extension two types of mental causation. While I don't have to endorse such a view for my argument, I do not rule out that it's possible to understand mental causation through difference making type views or through production type views via psychophysical processes and

mechanisms.

However, one requirement my thesis does make of a theory of causation is that it be able to handle cases of probabilistic causation. This is not really a restriction in the end though, as any complete theory of causation, of whichever stripe, will provide for an analysis of probabilistic cases.

3.1 Five Theories of Causation

Of course, there are far too many theories of causation to cover here[3], in what is really a quick overview rather than a detailed examination. Therefore, I shall only cover five sets of theories; counterfactual, probability raising, interventionist, process and mechanistic theories.[4] I picked these based on their popularity within the field and their applicability to chancy or probabilistic causation. Probability raising accounts for example, seem particularly apt for discussing probabilistic causation.

These theories are often interrelated to each other in various ways but distinguishing them clearly will not be necessary for my purposes. I wish to remain as neutral as possible in regards to theories of causation so that what I have to say can have the widest possible appeal. Luckily, I believe my arguments can be adapted to accommodate which ever theory of causation is preferred and so

[3]For example, I will not be talking at any point about regularity theories of causation and Hume will only be mentioned here. See (Hume & Norton (2009) and Hume (1988)).

[4]As I will discuss more below, counterfactual, probability raising and interventionist accounts are not necessarily mutually exclusive.

it won't harm my argument to not endorse one in particular. That having been said, I do want to demonstrate that my views can be consistently held within a coherent world view. To that end I will make use of a counterfactual probability raising theory.

3.1.1 Counterfactual Theories of Causation

Perhaps the first kind of theory which springs to mind when causation is mentioned is the counterfactual theory of causation, and particularly that of David Lewis ((1973), (1986*b*)). Contemporary counterfactual theories including Lewis' make use of possible world semantics in order to assess the truth or falsity of any given counterfactual statement. Lewis does this through the idea of cross world similarity. So, any counterfactual statement of the form 'if A had not occurred then B would not have occurred' is true if and only if there is no non-A world or some non-A world in which B does not occur is closer to the actual world than any non-A world in which B occurs.

Lewis makes the point that although causal relata are (often taken to be) events and counterfactuals are presented as propositions, this is not problematic. This is because for every proposition can be paired with a corresponding event:

> "To any possible event *e*, there corresponds the proposition *O(e)* that holds at all and only those worlds where *e* occurs. This *O(e)* is the proposition that *e* occurs. ... Counterfactual dependence among events is simply counterfactual dependence among the

corresponding propositions." ((1973) p.562)

The most straightforward of these theories is that for any two events A and B, A causes B if and only if had A not occurred then B would not have occurred. More fully it would be correct to say, A causes B if and only if, had A not occurred then B would not have occurred, and, had A occurred, then B would have occurred.[56] Causal dependence in Lewis' terms can then be understood as counterfactual dependence among distinct events:

> "Let c_1, c_2, ... and e_1, e_2, ... be distinct possible events such that no two of the *c*'s and no two of the *e*'s are compossible. Then I say that the family e_1, e_2, ... of events *depends causally* on the family c_1, c_2, ... iff the family $O(e_1)$, $O(e_2)$, ... of propositions depends counterfactually on the family $O(c_1)$, $O(c_2)$, ... As we say it: whether e_1 or e_2 or ... occurs depends on whether c_1 or c_2 or ... occurs. ((1973) p.562)

Likewise, causal dependence can hold between single events. In such cases, causal dependence between single events c and e is defined as when the family O(e), ¬O(e) counterfactually depends on the family O(c), ¬O(c).[7]

So to put this into terms of mental causation, I will start with a Kimean conception

[5]In the instance that A and B are both actual events the second of these counterfactuals is automatically true according to Lewis' semantics. This is because the closest A world is the actual world and it is also a B world. This relies on what he calls 'strong centering' whereby the actual world is closer to the actual world than any other is ((2001) pp.14-15).

[6]For Lewis though, A and B need only to be connected by a chain of counterfactual dependence (see Lewis (1973)). This is help deal with cases of early preemption. As its unlikely that mental causation will involve systematic cases of early preemption, I will not explore this point further.

[7]I may slip between using the terms 'counterfactual dependence' and 'causal dependence'.

of mental events. In that case mental event *m*'s causing physical event *p* can be understood as the counterfactual dependence of *p* on *m*. On the other hand, if you take a Davidsonian coarser grained view of events then it is the mental property of the event which must be causally relevant. We can generalise Lewis' account to accommodate this by saying that *m* causes *p* in virtue of mental property M iff (where M is an accidental property of *m*) had *m* not instantiated M, then *p* wouldn't have occurred or (where M is an essential property of *m*) had *m* not occurred, then *p* wouldn't have occurred.

For example, for the mental event *t* (thirst) and the physical event *d* (of reaching for my drink), had *t* not occurred (or had *t* occurred but not been a thirst) then *d* would not have occurred and if *t* had occurred then *d* would have occurred. If I had not been thirsty, then I wouldn't have reached for the drink. Looking to the worlds in which *t* did not occur, the closest worlds to the actual world are ones in which *d* does not occur. Therefore, it is true to say that had *t* not occurred then *d* would not occur and that therefore *t* is the cause of *d*. This would be a case of counterfactual causation.

There are some problems with counterfactual theories however, for example the problems of early and late pre-emption in which some cases counterfactual analyses can fail to count the cause as such. Early pre-emption cases involve scenarios where the process which would have led from the pre-empted cause to the effect is cut short before the effect occurs. Assassin cases are typically used to illustrate this. Take Hitchcock's ((2007) p.499) example where Assassin poisons Victim's drink leading to Victim's death. Backup would have poisoned

the coffee if Assassin hadn't so Victim would had died even if Assassin had not been the actual killer. Therefore there is no causal dependence between Assassin's actions and Victim's death, though Assassin's actions are the cause of Victim's death.

Lewis got around this problem by invoking his idea of casual chains which he defines as follows; "let *c*, *d*, *e*, ... be a finite sequence of actual particular events such that *d* depends causally on *c*, *e* on *d*, and so on throughout" ((1973) p.563). There must be a causal chain between one event and another in order for the former event to cause the latter. There is a causal chain between Assassin's actions and Victim's death. There is an intermediary event, lets say, the Victim drinking the poison on which Victim's death is causally dependent (because by that stage Backup had decided it was safe to not intervene) and which depends on Assassin's action of poisoning. There is no chain between Backup's action and Victim's death on the other hand. So there is a causal chain linking Assassin's (but not Backup's) actions to Victim's death so Assassin's (and not Backup's) action was the cause of Victim's death.

The classic case of late pre-emption from Ned Hall (2004) is that of Suzy, Billy and a bottle as described above in section 3. A possible reply to the problematic nature of late pre-emption cases is that events are modally fragile and therefore the smashing which would have occurred had Billy's throw been successful would have been a different smashing from the one which actually occurred after Suzy's throw. This is because the smashing resulting from Billy's throw would have occurred at a slightly different time and in a slightly different manner,

perhaps a different size of hole would be created for example. Lewis does not think this is the correct way to analyse the situation (See Postscript E *Redundant Causation* in Lewis (1986*b*)) however it is the one which intuitively makes sense to me. In his *Postscript to Causation* Lewis (1986*b*) gives the example of death by poison. The poison in question kills its victim much more quickly if taken on an empty stomach. When taken after food the poison is much more slow acting and painful. So, is it fair to say that the two deaths are different versions of the same event (in which case whether or not the victim ate beforehand is largely irrelevant) or different events? If the latter, then it would be true to say that the victim's eating dinner before ingesting the poison was actually part of the cause of the death, as it lead to this specific death as opposed to a much quicker and more painless one. This is because the specific death would counterfactually depend on eating the dinner. Lewis claims that it seems counterintuitive though to say that the victim's eating dinner was (at least part of) the cause of their death.

I personally do not find this way of thinking about the relata of the causal relation to be that counterintuitive. Indeed, there is perhaps a reason why someone may have the opposite intuition to me; that is, the intuition that the dinner is not a cause of the death. Hitchcock & Knobe (2009) analysed some experimental philosophy which suggested that subjects tend to conflate cause and attribution of moral responsibility. Subjects in the Knobe & Fraser (2008) experiment were presented with different scenarios where two peoples' actions were required to bring about the effect but only one of whom's actions violated

a norm in some way. There was a statistical difference in the attribution of causation with the agent violating the norm being considered the cause much more frequently than the morally 'neutral' agent. So when someone is deemed to have acted in a morally unacceptable way[8], their action is more likely to be deemed the cause and is likely to be attributed more causal weight. This makes sense given that, often, violation of a norm (particularly moral but also statistical) will lead to a negative outcome. To return to the poisoning case, this could help explain the widely held intuition that the poison and not the dinner is a cause of the death. The poisoning is a violation of a norm (which does indeed lead to a negative outcome) whereas eating dinner is not. Therefore *on the face of it*, it seems that it's the poisoning which is a cause of the death and the eating of the dinner is not.

A second factor that might lie behind the intuition that the poisoning but not the eating of the dinner is a cause is a conflation between token and type causation. Again this would make sense given that on a day to day basis, we have to make generalisations about causation. This links to Hitchcock and Knobe's claim that we attribute cause to that which we can most easily intervene on (See Hitchcock & Knobe (2009) pp.606-607). So, in the dinner case, although the dinner may be a token cause, generally dinner is not the type of thing which brings about death. Poison on the other hand is both a token cause in this case and a type of thing which causes death. We must therefore be

[8]Or, more generally speaking, when they violate either a "prescriptive norm" or "statistical norm" (Hitchcock & Knobe (2009) p.597).

careful around poison in a way we don't generally have to be careful around dinner, so we attribute the cause to the poison. The poison is the thing we could most easily intervene on and thereby prevent not just this death, but any death type event, from occurring.[9]

In a sentence, I do not think it's fair to say more causation must mean spurious causation. Our everyday ways of speaking about the fragility of events can be variable and context dependant. However, when we're constructing a consistent account of causation, a more standardised and consistent way of assessing event fragility will be required. So, I'm not sure how much we can apply intuitions from everyday ways of speaking into our philosophical analysis at least without closer consideration. This is my favoured approach to the problem of pre-emption although of course more could be said on this topic and other solutions to this problem have been suggested within the counterfactual tradition. For example, see, Lewis (2000), Yablo (2002) or Hitchcock (2001*a*)

How does a counterfactual approach to causation work in the case of probabilistic causation? In his "Postscript to `Causation'" Lewis (1986*b*) discusses just this. Roughly, for any, actually-occurring events *a* and *b*, *a* caused *b* if and only if, had *a* not occurred, the probability of *b* occurring would have been lower

[9]There is a similar issue with hastener/delayer intuitions. There is an asymmetry between whether hasteners or delayers are attributed as causes whereby hasteners generally are and delayers generally aren't. Bennett (1987) gives the example of heavy rains delaying the forest fire from May till June. The heavy rains, like the dinner, is a token cause of this particular fire, but not in general a cause of fires, and similarly is not usually attributed as a cause in peoples' intuitions despite the fact that it is a cause in this case of this particular June fire.

than it is in the actual world. This formulation is a combination of counterfactual and probability raising accounts which I introduce below in section 3.1.2.

There is a class of counterexample which is problematic for the probabilistic counterfactual approach provided by Menzies ((1989) pp.645-646). Menzies' example involves two systems which produce the same effect. He talks about two systems of neurons which I will label system A and system B. System A is more reliable than system B however. On one occasion, both systems fire but system B, the less reliable system, inhibits system A which then switches off. On this occasion system B cooperates and produces the effect. So, despite lowering the chances of bringing about the effect, system B is a cause of the event's coming about.

A more intuitive example comes from Dorothy Edgington ((1997) p.420). Two people are on a deer-hunt but with only one gun and are down to one bullet. The first person is a crack shot while the second is a novice. If the second person doesn't shoot the first person will. In the end the second person takes the shot and does manage to kill the deer whilst lowering the probability of this event occurring.

While my account of causation does not presuppose this, my preferred solution to this case would be to say that the deer died a different death than the one it would have died had the first person shot (for example it would have occurred at a different time). Lewis' solution was to say that there was a chain of probability-raising between the second person's shot and the deer's death.

The second person's shot raised the probability of the bullet flying through the air between the second person and the deer. Once the bullet is flying through the air the first person can no longer take that shot. So the bullet flying through the air raises the probability of the deer's death.

It is important to note that, although I can't go into them here, there are many other arguments which have been put forward[10] in response to these problems with counterfactual theories.

3.1.2 The Conditional Probability Approach to Probability-Raising

As this thesis is not only about causation, but specifically probabilistic causation, it would make sense to discuss probability raising theories of causation.[11] The conditional probability approach is an alternative way of understanding probability raising. In this sense it is an alternative way of doing the same thing as the counterfactual based approach to probability raising. Lewis discusses this distinction in "Chancy Causation" in his "Postscripts to `Causation'" ((1986*b*) pp.175-184) where he puts forward some reasons for preferring the counterfactual approach. For example, one reason is that conditional probabilities "go undefined if the denominator is 0" ((1986*b*) p.178). As he points out, this would turn out to be especially problematic in deterministic worlds where probabilities are either 1 or 0. For now, I will put critiques of the approach to one side and introduce it more fully.

[10]See for examples, Menzies (1989), Hitchcock (2001*b*) and Fenton-Glynn (2017).
[11]The following exposition is influenced by Hitchcock (2012).

The idea is that some event *a* is the cause of further event *b* if and only if *a*'s occurring raises the probability of *b*'s occurring. To put this more formally *a* is a cause of *b* iff P(*O(b)* | *O(a)*) > P(*O(b)* | ¬*O(a)*) where *O(b)* is the proposition that *b* occurs and *O(a)* is the proposition that *a* occurs. In prose, *a* causes *b* if and only if the probability of *b* occurring is higher given *a* occurs than the probability that *b* occurs given that *a* does not occur.

More accurately for my purposes, event *e* brings about further event *e** in virtue of instantiating M if and only if event *e* raises the probability of *e**s occurring in virtue of instantiating M. Let E* stand for the proposition that `*e\** occurs' and let I{e,M} stand for the proposition that `*e* occurs and instantiates M'. The inequality then is as follows; *e* is a cause of *e** in virtue of instantiating M iff P(E* | I{e,M}) > P(E* | ¬I{e,M}).

There is a problem for the conditional probability approach to accounts of causation which involves two events being correlated due to sharing a common cause. This can then produce the illusion of causation where none actually exists. The famous example of this, originally given by Hans Reichenbach (1956), involves atmospheric pressure, a barometer and a storm. The atmospheric pressure is a common cause both of the barometer reading changing and of the storm occurring. Because the barometer reading changes just before the storm hits, it may be thought that the reading on the barometer changing is a cause of the storm, although obviously this is not the case. Say A is the proposition that the atmospheric pressure is changing, B is the proposition that the barometer reading changes and S is the proposition that the storm occurs.

Then P(B | A) > P(B | ¬A) and P(S | A) > P(S | ¬A). But the inequality P(S | B) > P(S | ¬B) also holds. That is, the probability of a storm occurring, given the change in barometer reading, is higher than the probability of a storm occurring given no change in the reading. This is obviously because there is no change in barometer reading without the atmospheric change which also actually causes the storm, but there's nothing in this inequality to tell us that.

Reichenbach (1956) however, came up with a solution to this problem; screening off. When two events are screened off by an earlier event, they are not causally related. The idea is to hold other conditions fixed when assessing the inequalities in order to try to isolate the specific thing which is actually doing the causing. What does it mean to hold something fixed? It means including the fixed variable among the things being conditioned on. So, say we hold atmospheric pressure variable fixed. Does the inequality P(S | A & B) > P(S | A & ¬B) hold? No, it will not because the atmospheric pressure will cause the storm whether or not the barometer reading changes. However, holding the barometer reading fixed, the inequality P(S | A & B) > P(S | ¬A & B) will hold. Using this method (and so long as the cause comes earlier than the two screened off events), it becomes obvious that the atmospheric pressure is what is actually causing the storm. It also puts us in a position to be able to tell that the storm and the barometer reading share a common cause.

There are some counterexamples to probability raising accounts. For example there is a class of counterexample in which a cause lowers the probability of the events coming about (see for example Hitchcock (2004*b*)). Menzies' two

systems case given in section 3.1.1 above would be an example of such a case. Indeed, probability-lowering causes are problematic for both conditional probability and counterfactual approaches, and the solutions suggested to fix this issue likewise may work for either approach[12].

There are also problematic cases of probability raising non-causation (and once again these apply to both conditional probability and counterfactual approaches). Due to space restrictions I will not go into detail, but it is important to note that there are a variety of arguments which have been put forward in response to these problems.[13] This omission is permissible because these debates are slightly tangential to my main argument. The problem cases usually deal with (probability-raising non-causation) cases in which a cause brings about a process which raises the probability of the effect occurring but that process gets broken and therefore doesn't run to completion. Thus, a cause raises the probability of a process for an effect. Schaffer (2001) details various examples of such cases. To take one, say Frank and Pam both have bricks and are standing near the same window. Fred is just about to throw his brick when Pam independently throws hers and smashes the window. As Fred was just about to throw, assuming there was a greater than 0 probability that Pam would either not throw or would miss her target, he raised the probability of the window smashing even though his throw 'fizzled out' ((2001) p.81). To deal with these cases Schaffer proposes a theory of causation which combines probability

[12]Hitchcock, for example, suggests a possible solution in (2001*b*).
[13]See for examples Hitchcock (2001*b*), Fenton-Glynn (2009) and Kvart (2004).

raising and process views.

However, I think I can somewhat sidestep the problem. This is because, although there may be cases of 'fizzled' mental causation, with most mental causation there is little reason to suspect that these type of problem cases will arise systematically. This is because there seems to be robust systematic causal relationships between mental and physical states where the former cause the latter. If there is systematic mental causation of physical events, which I argue there is, then that suggests the causation in play is not spurious.

3.1.3 Interventionist Theories of Causation

The third set of theories I will discuss here are interventionist accounts[14], particularly as I will make reference to these later in section 5.3 when I discuss the placebo effect.[15] Interventionist accounts relate to counterfactual and probability raising accounts of causation. For example, you could think of interventionist accounts as being counterfactual accounts which promote a certain type of semantics for counterfactuals. The theories suggest that we should evaluate the causally relevant counterfactuals relative to those worlds in which their antecedents are realised by interventions.

Further, probabilistic interventionist theories could be considered a version of

[14] I followed Woodward (2013) in my exposition of this section.

[15] Indeed one strength of such theories is that they are made wide use of outside philosophical circles, for example in medicine. This would lead us to believe that, like Newtonian Physics, whether they are correct or not, they are at least prima facia effective at getting 'correct enough' answers. Such putative empirical success seems like a good reason to me to consider such theories carefully.

probability raising accounts (where probability raising is understood in terms of interventionist counterfactuals). So, when an intervention on a given variable raises the probability of an event occurring (which could be represented by another variable's taking a certain value) then the variable being intervened on can be considered a cause. What is the relationship between variables and events? After all, I have stated I will take events, not variables, to be the relata of causation. There is no conflict here though because variables can be used to represent whether an event occurs or not.

Interventionist accounts hold that A is a cause of B if and only if the correlation between A and B holds even after manipulation or intervention. I take the following simple example from Hitchcock (2019). It takes the form of a Structural Equations Model.[16] A Structural Equation Model "$\mathcal{M}$ is an ordered set <$\mathcal{V}$,$\mathcal{E}$>, where $\mathcal{V}$ is a set of variables, and $\mathcal{E}$ is a set of structural equations" (Fenton-Glynn (2017) p.4) where the variables represent potential causal relata. The models can be used to show how the variables relate to each other and can be manipulated or intervened upon to highlight which of these relations is causal. The models can also be presented graphically by using arrows to represent relations between the variables. There are three variables in this model:

B = 1 if Billy throws his rock, 0 if he doesn't

S = 1 is Suzy throws her rock, 0 if she doesn't

W = 1 if the window shatters, 0 if it doesn't

[16]You don't have to be an interventionist in order to make use of Structural Equation Models and vice versa. However, often Structural Equation Models are understood using an interventionist semantics and so I discuss them here.

Let's say that Suzy throws her rock and Billy doesn't and suppose the model has the following equations:

B = 0

S = 1

W = max(B,S)

We hold variable B fixed at its actual value (representing the event that Billy doesn't throw his rock). We can now vary the value of variable S from 1 (Suzy does throw her rock) to 0 (she doesn't). This will result in a change in the value of variable W from 1 (the window does smash) to 0 (the window remains intact). So Suzy's throw is a cause of the window smashing. This is a deterministic example here but the same point can be made in probabilistic terms. If an intervention on a variable raises the probability that a second variable will take a certain value then we have a causal relationship.[17]

I will be sticking to interventionist accounts and not referring to manipulability accounts which make reference to human agency in their explanation of causation.[18] This is not least due to the complications which would no doubt occur when using human agency to talk about the causation involved in human agency. However, I believe a more philosophically important reason exists to

[17] Although there may be some cases, for example pre-emption cases, where we may need to hold some other variables fixed.

[18] See for example von Wright (1971).

not make use of manipulability theories of causation which is their reduction of causation to human agency. Briefly, I believe this is fundamentally mistaken and anthropocentric. Surely there was plenty of causation happening before there were even humans around to have agency. Indeed, the theory of evolution, which I believe gives us part of our reason for thinking that mental states are causally efficacious (see section 5.2 for my argument for this) must rely on the notion of causation while explaining how humans came into being.[19] Therefore I think to reduce causation to human agency is a mistake it is best to avoid.

Interventions as Woodward defines them on the other hand do not depend on human agency. Woodward (2003) characterises an intervention as "an idealised experimental manipulation carried out on some variable *X* for the purpose of ascertaining whether changes in *X* are causally related to changes in some other variable *Y*" ((2003) p.94) where the aforementioned manipulation may or may not be a human action. So while a human manipulation can be an intervention for Woodward, not all interventions are human manipulations.

For interventionists, A is a cause of B if and only if the correlation between A and B holds even after intervention. Thus if B and C both share a common cause A, then there will be correlation between B and C. However, this correlation will not hold under intervention. Take the barometer example. Although the storm and the barometer reading are correlated, this correlation breaks down

[19]It could be of course that the evolution of human agency brought about a new kind of causation, distinct from the causation which had been in play up until that point. This new type of causation could be what manipulability theories are tapping into. Against this I will only say that I see no reason which this should be the case and parsimony warns against just this sort of proliferation of kinds.

upon intervention. Say we intervene (though again, interventions need not be human manipulation) to fix the barometer reading to 'storm'. Because of this intervention, the barometer reading becomes independent of the atmospheric pressure and therefore independent of whether the storm occurs or not. However, despite our fixing the barometer to read 'storm' the probability of the storm coming about remains the same as it would if we had intervened to set the barometer to any other reading. This demonstrates that the barometer reading is not a cause of the storm.

As Woodward (2013) points out, interventionist accounts can deal with some pre-emption cases which some counterfactual theories can't handle.[20] Take for example the problematic gunman case. The first gunman who shoots and kills their victim is the cause of the death, but there is no counterfactual dependence because the second gunman would have shot and killed the target if the first gunman had thought better of it. This type of case can be tricky for counterfactual theories because they seem to be cases of causation without counterfactual dependence. On the interventionist accounts, holding the actual action (non-shooting) of the second gunman fixed by means of an intervention, intervening on the first gunman's actions, that is, preventing the first gunman from shooting, brings about a change in the effect, that is the target is not shot dead. So, in this case, the first gunman's shooting would qualify as a cause despite the lack of counterfactual dependence.

[20]Although some counterfactual theories can handle such cases, see for example Lewis ((1973), (2000)) or Yablo's "De Facto Dependence" (2002).

Again, space prevents me from going into more detail but there have been many defences put forward for interventionist theories.[21] Rather than explore these in greater detail however, I will now move onto process theories.

3.1.4 Process Theories of Causation

There are many variations on process theories so I will focus on Dowe's Conserved Quantity Theory of Causation as the most popular version of such theories.[22]

The idea behind conserved quantity theories of causation is that causation is the exchange of some conserved quantity from one object to another. Usually the conserved quantity in question is something like energy or momentum. For example Fair (1979) characterises it as "energy-momentum transference in the technical sense of physics" ((1979) p.219). What counts as a conserved quantity is taken from physics which therefore makes such theories quite empirically based. If a non-conserved quantity is exchanged then this is pseudo-causation. Dowe ((1995) p.324) gives shadows as an example of a pseudo-object which is not capable of acting causally. That is because it only possess non-conserved-quantities such as shape and size. They can change their size and shape but as they possess no conserved quantities, they are incapable of causing things. If two shadows cross, they will both leave with the same conserved quantities as they arrived; which is to say none. Rather it is the surface they are cast

[21] See for example, Pearl (2000), Halpern (2011), Halpern and Pearl (2005), Woodward (2003), Hitchcock (2001*a*) and Hitchcock and Woodward (2003).

[22] I followed Dowe (2008) in my discussion of this topic.

upon which is the genuine object, that is to say, an entity which can possess conserved quantities.

So what is a causal process on such a view? As it is objects which possess conserved quantities, it is the world line of an object (a causal interaction then being an interaction of two or more of these world lines in which a conserved quantity is transferred). A world line of an object is the set of all of the space-time points in its history. An object is anything specified in the ontology of the best current scientific theories. A conserved quantity is a property which is universally conserved such as energy, momentum or charge.

I want to question the causal power of pseudo-objects however. Shadows are, on Dowe's view, unable to have causal power. But, I have reacted to shadows in the past, usually by jumping at them, which means they have caused me to do things.[23] This would suggest that either shadows do possess a conserved quantity of some kind (which defies empirical evidence), or that transfer of conserved quantity is not required for causation. This is similar to the problem of causation by omission or absence which conserved quantity theorist have trouble explaining. Dowe (2001) gives the example of a father's inattention causing a child's accident. Which conserved quantity was transferred by `not paying attention to the child'? Schaffer gives the famous example (first given by Hart and Honoré ((1985) p.38)) of "the gardener's failure to water the flowers (absence) caused them to die" ((2000) p.295).

[23]Lewis also argues that absences can be causes in "Void and Object" (2004).

Dowe deals with this problem by positing another, secondary, type of causation he names causation* or quasi-causation which relies on counterfactual dependence. How does this work with my shadow problem? Dowe could say one of two things. He could, and actually does, say the shadow is a case of causation* and that had the light been there I would not have been scared. Alternatively, Dowe could reply that, although it's a convenient way to speak as though the shadow did the causing in this case, it was actually the ground the shadow was cast on which made me jump. I do not find either of these replies particularly satisfying however due to space I will have to put this issue to one side.

Mental causation could potentially be problematic for conserved quantity theories because it is unclear how mental states can possess conserved properties and which conservation laws would govern such cases. This wouldn't be a problem if mental states are identical to brain states because then the brain state could be the object which possess the relevant conserved quantities (whatever they turn out to be). However, I will argue against this identity (see Chapter 6.1). Furthermore, as I'm not embracing a substance dualism there is no mental substance which could act as the object to possess the conserved quantity. It's then not clear what would be possessing the conserved quantity in the mental causation case.

Assuming for now that mental states could be conserved quantities, the process theorist would likely need to find a new conserved quantity for them. It's further unclear how this new conserved quantity would interact with the kinds of conserved quantities which have already been mentioned; energy, momentum

and so on. This is David Papineau's (2002) objection. In his appendix he discusses the history of the search for `vital' or `special mental' forces which presumably could act as the required conserved quantity. However, "detailed physiological investigation failed to uncover evidence of anything except familiar physical forces" ((2002) p.254). Furthermore, "detailed modern research has failed to uncover any such anomalous physical processes" ((2002) p.253) as could even hint at the existence of previously undiscovered special mental forces.

Hitchcock (2009) argues that Salmon and Dowe's conserved quantity theories fail to meet three adequacy criteria; firstly that the theory should correctly classify causal and pseudocausal processes[24], secondly that the theory shouldn't be circular and thirdly that the appeal to conserved quantities shouldn't be redundant. Regarding the circularity criterion, a conserved quantity theory would be circular if it defines `object' and `possession' in terms of causal process and interactions ((2009) p.76). As to the redundancy criterion, a conserved quantity theory would fail to meet this if "the concept of an `object' does so much work in the theory" ((2009) p.76) that the appeal to conserved quantity adds nothing.

Hitchcock demonstrates this by use of counterexample cases. His first example of a pseudocausal process is actually provided by Salmon and is called `the spot of light' ((1984) pp.141-142). A rotating light source is placed in the centre of a large astrodome. The spotlight is spun so quickly that the spot of light

[24]Where this is limited to empirical cases and doesn't have to account for intuitions on magical or physically implausible cases.

illuminated on the wall travels faster than the speed of light. The spot of light on the wall is therefore a pesudocausal process.

Hitchcock then creates a gerrymandered case from Salmon's example by adapting the spot of light case. Thus, the `light gerrymander' is "the mereological sum consisting of the illuminated patches of wall taken during the time in which they are illuminated" ((2009) p.79). Hitchcock argues this light gerrymander, which is a pseudoprocess, could be seen as an actual causal process with a conserved quantity, energy, which the light transfers to the wall making the illuminated sections slightly warmer than the non-illuminated sections.

According to Dowe, what separates such gerrymanders from legitimate examples of causation (the example used in this paper is a sound wave) is that the former are not objects while the latter is, because sound waves appear in the ontology given by science. Dowe explicitly rules out `timewise gerrymanders' as real objects which is "a putative object defined over a time interval where the definition changes over time" ((2000) p.99). The light gerrymander is also a timewise gerrymander and therefore is not an object and therefore can't partake in genuine causal processes. However, Hitchcock points out that it's hard to see why the sound wave example won't also be a timewise gerrymander. To see why, consider the description of a sound wave Hitchcock provides:

> "For $t_1 \leq t < t_2$ the sound wave is molecule *a*
> For $t_2 \leq t < t_3$ the sound wave is molecule *b*
> For $t_3 \leq t < t_4$ the sound wave is molecule *c*, etc" ((2009) p.81)

The description of the sound wave seems to be changing over time as the wave travels through different molecules. What distinguishes the changes over time in the sound wave case and a 'true' timewise gerrymander? So, it seems Dowe's theory at least struggles to differentiate pseudo from genuine causal processes.

For Dowe, the spots of light are not causal processes because they don't posses any conserved quantities. In the 'spot of light' case it is the wall which actually posses the conserved quantities. Science will provide the answers as to which objects posses which quantities. To this end, Hitchcock conducted a small survey of four physicists asking each the question "whether a spot of light could possess energy" ((2009) p.88). The answers he received were far from unanimous on this, suggesting, according to Hitchcock, that the answer was far from scientifically obvious. It's harder to see therefore how Dowe can defend his claim that the spots of light do not possess their conserved quantities. Furthermore, according to Hitchcock, Dowe cannot get around this problem by arguing that the spots of light are not objects as this would render his appeal to conserved quantities redundant. This would be problematic if true, as the central concept of Conserved Quantity Theories is their appeal to conserved quantities. But why would such a move, if Dowe were to make it, render the appeal to conserved quantities redundant? Hitchcock claims this is because if Dowe were to claim that the spots of light aren't causal because they don't have conserved quantities because they're not objects, then Dowe is relying on the concept of 'object' to do all the work. In that case the definition of a causal process "could be replaced with the simpler principle 'a *causal process*

is a world line of an object'." ((2009) p.84).

3.1.5 Mechanistic Theories of Causation

There is another way of viewing causation and laws which is related to the manipulation and interventionist theories of causation, notably that of Woodward ((2003), (2013)) and to process theories of causation, particularly those of Salmon ((1980), (1984)) and Dowe ((1995), (2000)). These are mechanistic accounts. In this section I will explain how 'mechanisms' have been characterised in recent literature, what role they're supposed to play in causal theories and I will explain one major problem with mechanisms playing one particular role that has been posited for them: namely removing the need for counterfactual dependence in causation.

What is a mechanism? There are many different definitions of what a mechanism is[25] therefore different views on how exactly they play their roles (See Ioanndis and Psillos (2017)). Illari and Glennan[26] have however defined what they call minimal mechanism:

> "A mechanism for a phenomenon consists of entities (or parts) whose activities and interactions are organised so as to be responsible for the phenomena" ((2017) p.2)

The key idea Illari and Glennan say is that parts and interactions of the mech-

[25]See for example Glennan (2017), Illari and Williamson (2011), and Psillos (2011) among others.

[26]This definition is however taken from Glennan (2017). Note, Illari and Williamson (2011) have a similar definition.

anism are what has the power to bring about the effect. There are many examples of mechanisms that could be given (see for example Abrams (2017) who goes into detail about how bacterial chemotaxis in Ecoli is a mechanism.). Glennan gives the example of a toilet first in ((1996) pp.56-58) and then again, "when the handle is pulled, water is released from the storage tank into the bowl, and the storage tank is refilled" ((2011) p.803). In this case the parts are the handle, the tank and the water and the interaction of these parts bring about the emptying of the tank and the flushing of the toilet.

One role that have been posited (for example by Glennan (2011)) for mechanisms is as a basis for a singularist view of causation which removes the need for counterfactual dependence. This works because of the hierarchical nature of mechanisms. That is, there can be sub-mechanisms which make up parts of mechanisms which in turn can be made up of sub-sub-mechanisms and so on. This means that you don't need a counterfactual story to explain how a given mechanism works, it's 'mechanisms all the way down'.

There is a problem when you hit the fundamental level however. At this point you cannot call on sub-mechanisms any more. By definition you are at the lowest level. As mentioned above, some people therefore argue that counterfactual dependence is still required to ground this lowest level of mechanisms. In other words, while mechanistic theories may provide some causal analysis, there comes a point at which mechanisms aren't sufficient, a point at which there cannot be another sub-mechanism. At that point, perhaps, a difference making account is needed to show why the sub-mechanism is a cause. Ioannidis and

Psillos discuss just this question in their "Mechanisms, Counterfactuals and Laws" (2017). So it looks as though the mechanist may ultimately have to 'bottom out' at a counterfactual view. At least the burden of proof is on the mechanist to explain why this is not so.

3.2 Mental Causation

So, what is it that I mean when I talk about mental causation? The answer will be different depending on exactly which theory of causation is endorsed. I intend to be as open as possible in which theory of causation to endorse for two interrelated reasons. Firstly, it is because I hope that my arguments can be adapted to be compatible with various different theories of causation. Secondly, I want my views to remain as plausible and as available to everyone as possible, so I want to avoid tying myself to any one theory of causation if possible.

In general by mental causation I mean that a mental event, such as a feeling, for example thirst, is the cause of another physical event, say taking a drink. It is important that the mental event is either a true mental event (in a Kimean sense) or is an event (in a Davidsonian sense) which is causally efficacious in virtue of its mental property. So, whenever I speak of events in the Kimean sense, what I say can be rephrased into the Davidsonian sense and vice versa (although for space reasons I may not always explicitly do so).

Mental causation on a counterfactual interpretation of causation amounts to a

counterfactual causal dependence from a mental event to another event of the form `had *m* not occurred, *e* would not have occurred (or the probability of *e*'s occurring would have been lower). To put this into `mental terms' again; had event *e* not instantiated mental property M then further event *e** would not have occurred (or the probability of its occurring would have been lower) and if *e* had instantiated M then *e** would have occurred.

Alternatively on the conditional probability approach to probability raising, mental causation occurs when an event raises the probability of a physical event in virtue of its being a mental event or instantiating a mental property.

In the next chapter I will set out some theories of probability in order to build a more comprehensive world view in which my views could consistently and coherently be held. As I have already stated, I take many of our best scientific theories to be both probabilistic and our best guides as to how the world is. It is therefore important to understand how probabilities should be understood and interpreted to see how they can be reconciled with the phenomenon of probabilistic mental causation. I hope to show that there is a coherent world view which incorporates theories of causation and probabilities in which mental causation is philosophically unproblematic (at least in regards to the CEA).

4

THEORIES OF PROBABILITY

"We demand rigidly defined areas of doubt and uncertainty!"

The Hitchhiker's Guide to the Galaxy - Douglas Adams (1995)

This chapter introduces popular theories of probability. The final section deals with building an overall world view under which I claim our best scientific theories can be adequately philosophically reconciled with theories of causation, theories of probability and most importantly with the phenomenon of probabilistic mental causation.

4.1 What is Probability?

Key to my thesis is the concept of probabilistic indeterminism and while I have devoted time to discussing indeterminism, I haven't as of yet discussed the concept of probability per se.[1] I will correct this omission here by going through various theories of what probabilities actually are. This is necessary as it is important for my argument that objective probabilities do exist and that probabilities are not merely epistemic (see Section 4.1.3 for more on this). Roughly speaking, epistemic or subjective interpretations take probability to represent a subject's degree of belief. This speaks to the subject's knowledge as opposed to what's true or not in the world, though of course these will be related. Objective interpretations on the other hand *do* represent what is taken to be true in the external world because probabilities exist in the external world. This is important because if there are to be objective probabilistic causal relations in the world (as I claim there are) then presumably objective probabilities would be required to underwrite those relations. It is hard to see how my arguments could hold if the world lacked such objective probabilities. Objective and subjective interpretations are not mutually exclusive though. For example, even in a world where there such things as objective probabilities, subjective probabilities which track their objective cousins are very useful.

The structure of this section will be as follows: I will introduce the axioms of probability which will arise at various times throughout the chapter, I will then

[1] I followed Hájek (2012) in my exposition of this section

discuss the advantages and problems with six different interpretations of probability; analyse which of those work within the structure of various theories of causation and then conclude by indicating which interpretations I will make use for the remainder of my thesis.[2] The six theories I will cover are; classical, logical, subjective or Bayesian, frequency, propensity and best system.

4.1.1 Axioms of Probability

Before I go on to discuss the various theories of probability I will introduce the axioms which underpin many theories as well as the ideas of conditional probability and conditionalising. So, what are the axioms of probability? They were laid down by Kolmogorov in 1933 (see (1956)). Note, the three following principles are those taken to be the minimum set for any particular system of probability. I take the following formulation from Bradley ((2015) p.28).

Where X and Y are propositions, B is a set of propositions which is closed under conjunction and negation and P() is a function from B onto the real numbers, the following should hold in order for P() to obey the axioms of probability:

1. Non-negativity: $P(X) \geq 0$ for all X in B

2. Normalisation: P(T) = 1 for any tautology T in B

3. Finite Additivity: P(X v Y) = P(X) + P(Y) for all X and Y in B such that X is incompatible with Y

[2] Although, I think everything I have to say can apply to whichever theory is preferred, so long as that interpretation is objective. I won't have much to say about subjective approaches as these are not mutually exclusive with, and so can be held alongside, objective accounts.

In prose, non-negativity means that any proposition must have a probability greater than or equal to 0. It is not possible to have a probability less than 0. Normalisation says that any tautology[3] should have probability 1. Lastly, finite additivity says that if X and Y are mutually exclusive, then the probability of either X or Y should be the sum of the individual probabilities.

Now I will discuss some of the theories of probability which have been suggested throughout the years starting with classical and logical theories.

4.1.2 Classical and Logical Probability

Classical probability and logical probability have both largely been replaced by more contemporary theories. But, as they are the forerunners to more contemporary accounts I will briefly discuss them and why they have since been replaced.

Classical Interpretation

The Classical Interpretation of probability assigns an equal probability to each possible outcome whenever we have a group of mutually exclusive and jointly exhaustive outcomes (and no overriding evidence to support one outcome rather than another); this is known as the Principle of Indifference. Laplace

[3]Bradley mentions that this isn't uncontroversial if probability is understood as subjective, as it seems to imply that everyone should have credence 1 in every tautology including the ones they don't know about ((2015) pp.28-29).

states his theory of probability as consisting in "reducing all the events of the same kind to a certain number of cases equally possible" ((1902) p.6). So, as there are six faces on a dice then, assuming we have no evidence to suggest otherwise, there is a probability of $\frac{1}{6}$ of said dice landing on any given number. Similarly, as a coin has two faces, and ruling out the possibility that the coin could land on its edge, there is a $\frac{1}{2}$ chance of getting say a tails on flipping a fair coin.

I have presented an overly simplistic view of classical probability and there is much more that can be said about it. But already some problems present themselves. For example, in the dice or coin case it is fairly clear how many outcomes are possible although even within the coin example a simplifying assumption has to be made ruling out the coin landing on its edge. Furthermore, it intuitively seems that each outcome does have an equal chance of coming about. Arguably it's a good theory within its domain of fair probability games but inapplicable outside that. This makes sense given that the classical interpretation was constructed with such games in mind.

Everyday probability reasoning on the other hand usually involves scenarios where it's not always possible to calculate what all the outcomes might be and what weightings to give them. For example, if I throw a rock into the air in the middle of a busy street what are the potential outcomes? It could land harmlessly back on the ground, it could end up hitting me or another person or it could end up smashing a window. But these are not the only outcomes, it could hit a bird, a drone, or any number of things. Now, lets apply the principle

of indifference and assign each outcome I've listed here (lands harmlessly, hits a person, smashes a window, hits a bird or hits a drone or none of the above) a $\frac{1}{6}$ chance of happening. This doesn't seem right, surely the chance of hitting a person is higher than hitting a bird, numerous though pigeons are. And the probability of hitting a drone must be lower than both of these.

Bertrand's paradox (1888) demonstrates a problem with applying the Principle of Indifference. He laid out a problem; take a circle with a triangle inside where the points of the triangle touch the edge of the circle. What is the probability that a random chord of the circle is longer than the sides of the triangle? He showed that using different methods for calculating the probability produced different answers. Van Frassen (1989) provided a similar example with his cube factory ((1989) p.303). The factory produces cubes with edges up to and including 2cm long. What is the probability that any given cube made in the factory has an edge between 0 and 1cm long? It seems to be $\frac{1}{2}$ as entailed by the Principle of Indifference. However, "the problem could have been stated in different words, but logically equivalent form" ((1989) p.303). A cube with edge length between 0 and 1cm is equivalent to having a side face area of between 0 and $1cm^2$ and a volume between 0 and $1cm^3$. Having edge length between 0 and 2cm is equivalent to having side area between 0 and $4cm^2$ is equivalent to having volume between 0 and $8cm^3$. So what's the probability of any given cube produced at the factory has an edge length of 1cm or less? The Principle of Indifference seems to imply it's $\frac{1}{2}$ looking at edge length, $\frac{1}{4}$ looking at side area and $\frac{1}{8}$ looking at volume. Different but equivalent ways of

describing the problem leads to different and incompatible probabilities which cannot all obtain at the same time.

These examples show that the same event can have incompatible probabilities assigned to it depending on how you choose to divide up the possibility space. But one event shouldn't have more than one probability of occurring at any given time on pain of inconsistency. The Principle of Indifference leads to these inconsistencies because each way of dividing the outcome space into seemingly equal possibilities results in different, incompatible probability assignments. Therefore the principle gives you no way to know which way of dividing the space is superior. Evidence is then needed to supplement the principle in order to avoid this. Therefore, the classical interpretation is at best a partial theory of probability.

Logical Interpretation

The logical interpretation assigns probabilities based on possible outcomes. However, and importantly, the probabilities do not have to be assigned equally in accordance with the Principle of Indifference. This makes it more applicable to a wider range of cases where some outcomes are more likely than others. Probabilities are instead assigned on the basis of the amount of evidence is available for each outcome.[4] So the probability of an event's occurring will change depending on how much evidence you have for (or against) it. For

[4]A distinction between classical and logical theories is that the classical interpretation is considered more of a forerunner to objective accounts whereas the logical interpretation is more naturally thought of as the forerunner to subjective accounts because the probability of an event occurring is calculated relative to the evidence for it.

example, say I want to work out the probability that it's currently raining where I am. Let's say the only piece of evidence I have regarding whether it's raining or not is that I'm currently in Scotland where it rains 50% of the time. There is therefore a 50% probability that it's raining. I add to my evidence by looking at the calendar and noting that it's January which is a rainy month. Given my new evidence base there's an 80% chance it's raining so I reach for my umbrella. If I looked outside the window and saw it actually wasn't raining, then the probability relative to my new evidence base would be drastically reduced.

Carnap (1951) constructed the most sophisticated system of logical probability.[5] His intention was to build a system which took the logical power of deduction and 'translated' it into induction, hence the name *logical* probability. The general goal of such theories is "to encapsulate in full generality the degree of support or confirmation that a piece of evidence E confers upon a given hypothesis H" Hájek (2012). In other words, Carnap's idea was that evidence can lend objective support to a hypothesis to different degrees. This information can then be systematised into a theory of logical probability.

Much could be said about this interpretation but I will leave the topic here. This is for two reasons. Firstly, this account has now been largely superceded by the Bayesian account. Secondly, the logical interpretation is an epistemic account of probability. As my argument requires an objective theory of probability I will not be focusing on subjective accounts as these two varieties of theory are not

[5]Other proponents of logical interpretations include Johnson (1921), Keynes (1921) and Jeffreys (1948).

mutually exclusive. It is for this same reason that I will not focus too much on the Bayesian Theory of probability despite this being an important theory.

4.1.3 Subjective or Bayesian Probability

Bayesian and subjective accounts do not take probability to be an objective feature of the world rather as an epistemological phenomenon. So probability is not a way of representing the way things are in the world, but rather the subject's degree of belief (or credence or confidence) in given outcomes occurring. This raises the question of what exactly is a 'degree of belief'? Often they are cashed out in terms of betting behaviour.[6] So, the higher a degree of belief you have the higher value a bet you'd be willing to take on a given proposition's being true.[7]

This isn't to say that these probabilities aren't supposed to track chances in the world. You want to assign higher probabilities to events that are actually more likely to occur and likewise lower probabilities to events less likely to occur. But the probability in itself is not a feature of the external world, rather it is a feature of our lack of knowledge. There is actually a spectrum of subjective or Bayseian theories which span from the entirely subjective epistemological side to the more objective probability tracking side.

On the extreme subjective end are what Bradley ((2015) p.61) calls "subjectivists".

[6] See for example, Cohen and Hansel (1956).

[7] Assuming that you're free, un-coerced and have no reason to misrepresent your degree of belief and so on.

They believe that degrees of belief must be probabilities (so they must obey the axioms of probability) and these beliefs should be updated by conditionalization. However, beyond these minimal requirements there are "no *further* rationality constraints" ((2015) p.61, emphasis in original).[8]

The more objective leaning accounts on the other hand may agree that the probability is an expression of a degree of belief but that this is tracking objective probabilities out there in the world. As such an objective Bayesians "hold that there is a fact of the matter about what someone *ought* to believe given their evidence" ((2015) p.63, emphasis in original). This places further rationality constraints upon them for example the Principle of Indifference or Lewis' Principal Principle.

It is important to note that one could endorse a Bayesian or other subjective account of probability while also endorsing an objective account. The two are perfectly compatible as long as your metaphysics doesn't explicitly rule out or the other. Indeed the Principal Principle (Lewis (1987*b*), (1994)) roughly speaking, states that rational credence in an event is equal to its objective chance where this is known. Lewis' definition runs as follows;

> "Let C be any reasonable initial credence function. Let t be any time. Let x be any real number in the unit interval. Let X be the proposition that the chance, at time t, of A's holding equals x. Let

[8]Though, it should be noted, this is an extreme view and one I'm not sure anyone actually endorses.

> *E* be any proposition compatible with *X* that is admissible[9] at time *t*. Then *C(A | XE) = x*." ((1987*b*) p.87)

That is not to say that objective and subjective probability cannot come apart. As Lewis says, "canons of reasonable belief need not be counsels of perfection" ((1987*b*) p.85). There can be situations where the objective chance of an events occurring and the correct level of credence you should put in the event occurring can be different. Take an example involving a loaded coin which the agent doesn't know is loaded. This means the probability of landing heads is actually 0.33. The agent however, lacking this information, assumes that the coin is fair so they assume a 0.5 credence to it landing heads. This would be the correct credence to assign if the coin was fair which to the best of their knowledge it is, so they are not irrational in their degrees of belief. If, on the other hand they had been told (by a trustworthy and knowledgeable source) prior to the toss that the coin was loaded such that the probability of heads was 0.33 and they didn't update their belief in line with this information, they would potentially be violating one of the main principles of Bayesianism, conditioning on evidence, or the Principal Principle (and possibly both).

While there are many different kinds of Bayesian or subjective accounts, there are generally taken to be two main tenets which proponents of such theories will all hold which I will discuss in the next section. There are many other constraints which different variations accept or reject (for example they may differ on what

[9]Admissible information being any which is not `crystal ball' type information which says how future events turn out, see ((1994) p.483).

constraints an agent's degrees of belief should meet) but the two cornerstones are that subjects must obey the axioms of probability and that beliefs are updated by conditioning on evidence.

Two Main Tenets of Bayesianism

The first general premise of Bayesianism is that agents must obey the axioms of probability in order to be rational. Agents which fail to obey the axioms are irrational. Unfortunately many people do not understand[10] or obey the axioms which leaves them open to bad betting behaviour in which losses are guaranteed (this is known as a Dutch book).

The second principle all Bayesians hold is that beliefs should be updated and conditioned on new evidence. The only rational way to update your beliefs is to follow condtionalisation[11]. That is "an agent's beliefs after learning E should equal her earlier beliefs *conditional* on E" (Bradley (2015) p.51). It's important that we are able to update our beliefs over time as we gather more evidence. We can formalise this as $P_B(A) = P(A \mid B)$ where A and B are the propositions that events A and B occur respectively[12] and B has a probability greater than 0.

[10]Understanding the axioms is not a requirement, as you can obey them without understanding them. However, I would think it would be easier to explicitly obey them, for example in betting decisions, if you do understand them.

[11]There's a more general principle of conditionalisation called Jeffrey Conditionalisation. This definition is taken from Bradley ((2015) p.53):

> *Jeffrey Conditionalisation*
> $P_E(H) = P(H \mid E)P(E) + P(H \mid E)P(E)$.

Many Bayseians would allow that you can update by Jeffrey Conditionalisation even if you are not certain about the evidential proposition E.

[12]There is of course a distinction between events and propositions but every event will have a

This assumption allows a subject to update their degree of belief in an event occurring in relation to new evidence. P(A) is the initial degree of belief in the proposition that event A occurs is true. P_B(A) is the posterior probability (your credence in A after learning B). P(A | B) is the prior probability in the proposition that A occurs given B occurring.

As Bayesianism is a subjective account it will not suffice for the needs of my thesis. Thus I will not discuss it further and will move on to discussing objective theories of probability starting with the frequentist approach.

4.1.4 Frequentist Interpretations of Probability

There are two main types of the frequency interpretation of probability. The first I will discuss is the finite approach wherein actual frequencies are taken to *be* the probabilities. This account suffers from some serious problems which sparked the development of the second main variety; the hypothetical frequency interpretation. On this interpretation, rather than the actual outcome of a finite series of events, the limiting frequency of a hypothetical infinite series is the probability.

The Finite Frequency Interpretation

The finite frequency interpretation of probability is an objective theory. The difference is that it assigns probabilities based on actual outcomes rather than

corresponding proposition stating whether that event occurs or not. So, when I speak of events occurring I am more strictly speaking about its corresponding proposition although I may not always state this explicitly.

possible outcomes. The classic example is coin tosses. Say a coin is flipped 100 times and the results recorded. Heads occurred 48 times and tails occurred 52 times. The frequency within this finite reference class is therefore 0.48 for heads and 0.52 for tails. If we were to flip the coin more and more times, although not an infinite amount of times, we would expect the result to trend towards 50/50. But the probability assigned to a given outcome just is the frequency with which it occurred. So if we flipped the coin another 900 times and the results came in 480 for heads and 520 for tails across the 1000 flips, the probability for flipping heads just is 0.48 and the probability of flipping tails just is 0.52. The same could be said of a fair die. We would expect each number to occur approximately $\frac{1}{6}$th of the time given a large enough, but not infinite, series of rolls.

One problem with the finite frequentist approach is that within a finite reference class, you can get 'anomalous' results. Say for example, that we did a second round of coin flipping. This time, out of our 100 coin flips we got 12 heads and 88 tails. This is perfectly possible if a tad inconvenient and gives us a very different result to the first 100 coin flips. Of course, the larger you make the reference class, the more we would expect it is that the results tend to 50/50, so in this way we could solve our problem. Why would we expect this? We would expect this because the law of large numbers[13] would suggest that the larger a number of trials the higher the probability that the results should tend to the expected value or fall within a narrow range around this value. In other words,

[13]Révész ((1968) p.8) notes that there are various related laws of large numbers but they all share the idea of convergence to an average.

the more times you run the trial the more likely you are to achieve approximately the expected outcome. However, speaking about becoming 'more likely' is tricky in the context of establishing a theory of probability which defines what 'likely' means without risking circularity. After all we can't say what likely means independently of a theory of probability, exactly what we're trying to pin down.

The upshot is that the finite frequentist approach doesn't always track our intuitions about probability very well (although as always with intuitions, they have the capacity to be false friends). So, should we be searching for a theory which explains these intuitions (presuming one could be found) or should we accept finite frequentism and its counterintuitive results in some cases?

There are certain 'one-time' or 'black swan' (Taleb (2008)) events which only occur once or very rarely and in the case of black swans have a large effect when they do. What is the frequency of such an outcome within a larger reference class? Other more mundane events which only happen rarely would equally prove problematic for the same reasons if not quite deserving the title of 'black swan'. Take Lewis' 'unobtainium' example ((1994) p.477). Unobtainium^{346} is so difficult to make that only two atoms of it has ever existed. The first has a 4.8 microsecond lifetime whereas the second atom's lifetime was 6.1 microseconds long. What can we surmise about the true half-lives of unobtainium atoms from these frequencies? Lewis argues next to nothing.

Along the same lines there are events which may never actually come to pass[14]

[14]Lewis' unobtainium has an isotope: unobtainium^{349}. This isotope is even rarer than its counterpart and there is no atom of it. As the frequency of its decay within a given time period

(say me intentionally dyeing my hair neon green) which nevertheless do have a probability of occurring attached to them. But if I never do it, then what is the number of actual outcomes in order to assign said probability?

A further problem with this account is the reference class problem (see Venn (1876), Reichenbach ((1949) p.374) and Hájek (2007). It has to do with the reference class against which you count the frequency of an events occurring. A choice must be made as to which reference class is appropriate for comparison. One possible answer is that the appropriate reference class for tosses of a coin is the class of other tosses of that same coin (or maybe coins relevantly similar to it). As soon as we move into more complex scenarios however, the answer becomes less clear cut. Medicine is a field where we can get many realistic examples of this problem for example when trying to calculate the probability of a given person suffering a given disease. Say I want to know what my probability of catching the flu is. Which reference class do I use to calculate the frequency in which it arises and thereby find the probability? Should I use the class of all people? The class of all people living in London? All people living in London within my age range and social class? An indefinite list of potential reference classes could be given. And, while some of these sound obviously to be more reasonable classes than others, it's not clear which the *best* class would be. Equally compelling reasons could potentially be given by proponents of various classes. But why is this matter so problematic?

is 0/0, this means it's undefined. "If there's any truth about its chance of decay, this undefined frequency cannot be the truthmaker" ((1994) p.477).

This is problematic because each reference class will lead to different frequencies and therefore to different probabilities of me catching the self same disease in each case. Perhaps the answer would be to get more and more specific. Keep adding variables to the reference class until you have covered every possibly relevant base. However this means you will also keep adding variable until you end up in the position where the reference class which is applicable to me *is me*. This is problematic because the objective chance now depends on whether I get the flu (in which case the chance is 1) or not (in which case the chance is 0). But the objective chance seems like it should be something other than 1 or 0.

Perhaps the lesson of the previous paragraph is that there are no chances that pertain to individuals. Only those that pertain to people as members of a group. As Von Mises said "The phrase `probability of death' when it refers to a single person, has no meaning at all for us" ((1939) p.11). But this is of no help given that it seems wrong to say there is no objective chance of an individual catching flu (or dying) at any given time. There surely is a chance I could get ill tomorrow and it is surely greater than 0. So this solution does not work, particularly not if we want to provide a theory of token causation for which we'll require a good understanding of single case probabilities.

Given what quantum mechanics tell us about the world, I think we have good reason to think the world is in principle probabilistic and therefore our interpretation of probability should allow for this and explain it. But, there are some potential problems here for the finite frequentist approach in relation to quan-

tum mechanics. Hájek (1996) makes the point that a relative frequency can never take an irrational number for a value. This is problematic because most numbers between 0 and 1 are irrational and quantum mechanical probabilities can take such values. Hájek states, "for example, according to finite frequentism, the radioactive law for radium is false for all time periods that have irrational probabilities for decay - which is to say that it is false almost everywhere" ((1996) p.224).

Hypothetical Frequency Interpretation

The hypothetical frequency interpretation solves some problems raised by the finite frequency interpretation, for example the one-time event problem. This is because, under this view, the probability of a given event is equal to the limiting frequency of an infinite series of events of a relevant type. For example, if we wanted to calculate the probability of a coin landing heads the relevant event type which we would run an infinite series of would be coin flips. This makes the probability a counterfactual matter as opposed to an actual one (unless the event in question is of a type which actually has infinitely many instances). This may be a problem in and of itself if you find appeal in the empirical grounding of the frequentist approach. Besides this, there are other problems which can come from moving to infinite series of trials.

One such problem is that of ordering as discussed by Hájek ((2009) p.218-220). He gives the example of tossing a coin (so the `correct' limiting frequency which should result is obviously 0.5) on a train which is moving back and forwards in

an east-west direction. He plots the resulting heads and tails on a graph with the temporal dimension on the y-axis and spatial dimension on the x-axis. He then asks which order we should put the results in. In the example he gives the temporal ordering produces one set of results, the spatial results another. And because the two orderings are different, they might result in different limiting frequencies[15]. This is possible because there is no one privileged dimension which should take priority when it comes to ordering. Intuitively, the temporal order seems to take preference, but there is nothing more than intuition to back up this choice. To illustrate this, take the example of simultaneous flips. In this world, rather than flipping one coin infinitely (lending to a temporal reading), infinite coins are all flipped simultaneously. There is no temporal ordering to be had here as all the coins are flipped at the same time.

Another question for the hypothetical frequentist is how do you 'run' an infinite series of trials? It's impossible to actually run it which is why it becomes a hypothetical, counterfactual issue. Assessing these counterfactuals doesn't seem difficult to do in the simple cases such as coin tosses and dice rolls with intuition to fall back on, but becomes much less clear as soon as more complex scenarios are considered.

Ultimately the biggest problem for the frequentist interpretation of probability is that even in an infinitely large series of tosses of a physically unbiased coin

[15]He mentions you can make the problem even thornier by adding a lift to the train and adding another spatial dimension (up and down as opposed to east and west) to the mix ((2009) p.219). This would produce another ordering and another limiting frequency further compounding the problem.

(or other event), it is possible to not get perfect 50/50 results. This means that according to this interpretation, the probability of a heads or tails would not be 0.5. If the theory cannot guarantee us the right results (or alternatively leads to such a counterintuitive result we have to rethink a lot of what we believe to know about probability) then this looks the worse for the theory. In both the finite and hypothetical frequency accounts our intuitions suggest that the truthmakers for probability claims diverge from the frequencies. Lewis makes this point when he discusses the decay frequency of unobtainium ((1994) p.477) discussed above. Again then, the question is, should we accept this counterintuitiveness or search for a theory which can explain and incorporate our intuitions?

So, the frequentist interpretation of probability suffers from some problems. Perhaps the propensity interpretation, which was actually put forward to try to solve some of these problems will fare better and therefore, if successful, be a better overall approach.

4.1.5 Propensity Theory of Probability

Popper ((1957), (1959)) put forward a propensity theory specifically to handle the problem of quantum single case probabilities[16]. Probability on this interpretation is understood as a physical propensity or disposition of a certain kind of physical scenario to lead to another physical scenario of a given kind. This deals with the frequentist single case problem because the probability of a given event is not *identified* with the number of actual outcomes. In so far as it achieves this,

[16]Propensity theories were also posited to help solve the reference class problem.

it therefore looks like a better theory[17]. Broadly speaking there are two kinds of propensity theory distinguished by Gillies (2000); the long-run interpretations and single case interpretations.

The way probabilities are calculated on this approach does make use of actual outcomes as a proxy for, or as evidence of, the propensities, at least when it comes to assessing what probability the propensity produces.[18] So, rather than merely counting the number of times an outcome occurs within a certain reference class, an experimental set up is required in which repeatable experiments can be run. The limiting frequency of the results of these experiments reflects the probability. Take the coin flipping example. A repeatable experiment must be set up. In this case, this is simple, it is just the flipping of the coin. Therefore the chance of heads or tails is 0.5. The probability here is not to be identified with the frequency but rather with a dispositional property instantiated in the physical experimental set up. In Popper's words, the probability is "*a property of the generating conditions*" ((1959) p.34, emphasis in original) which he took to be the "*whole physical situation*" ((1990) p.17, emphasis in original).[19] Whatever the specifics though, the disposition is a physical property.

What is a disposition? They could be understood in counterfactual terms. The

[17]There are many and various exact interpretations of propensity theory so it is tricky to treat them en masse. Therefore, what I have to say here will be general.

[18]The actual probability could diverge from the frequency which can be an advantage in some cases but would be tricky for the propensity theorist if it happened systematically. Luckily, there is no reason to think it would happen in a widespread and systematic manner.

[19]Whatever experimental set up conditions exactly instantiates the property will depend upon the exact variant theory.

simple conditional analysis[20] says that "an object is disposed to *M* when *C* iff it would *M* if it were the case that *C*" (Choi and Fara (2018)). For example, a glass has the dispositional property of fragility, based on its physical microstructure, which means that, were it to drop, it would smash. This is the reductionist view.

The picture is not as clear as this however, due to problems such as finkish dispositions. In his paper `Finkish Dispositions' Lewis (1997) attributes the idea to C.B. Martin (1994). One type of finkish disposition is one in which "stimulis *s* itself might chance to be the very thing that would cause the disposition to give response *r* to stimulus *s* to go away" ((1997) p.144). The converse situation also produces a finkish disposition. That is, something which doesn't already have a disposition to respond in a certain way to a given stimulus may gain a disposition only when the stimulus is present. Lewis gives the example of a finkishly fragile object as being one which is fragile until it is struck, at which time it doesn't break ((1997) p.144). Perhaps a real life finkish disposition can be found in Non-Newtonian fluids. Oobleck[21] is an example of such a fluid. It has the disposition to have objects sink into it (objects will sink if placed on top of such a fluid) unless those objects hit its surface with force at which time it becomes solid.

Lewis suggests you could resist Martin's counterexamples to the simple counterfactual analysis by focusing on the timings of the dispositions disappearing (or appearing). If the stimulus causes the disposition to disappear any less than

[20] Held for example by Ryle (1988), Goodman (1983) and Quine (2014).
[21] Oobleck is a mixture of one part water to two parts cornstarch. See Hoover (2018).

instantaneously, then there is (maybe) time for the response to manifest. If on the other hand the change is instantaneous then we can argue the case is "fantastic" ((1997) p.140).

Johnston (1992) and Bird (1998) provide a further problem for analyses of dispositions in the form of maskers or antidotes respectively. The idea is that it is possible to prevent dispositions from manifesting in the presence of the appropriate stimulus. Bird gives the example of medical antidotes ((1998) p.228). If you take an antidote quickly enough then it prevents the poison's disposition to harm you.

These problems for the conditional analysis of dispositions could be problems for propensities if they are supposed to behave in similar ways. One way to solve some of these problems would be to think of dispositions and propensities as brute, however some may find this unsatisfying or undermotivated. However, there may be more reason to think this tactic might work better for theories of dispositions rather than for propensities. An argument could be made that there may be more motivation for considering dispositions to be brute than propensities. This is because propensities are `tailor made' entities in a sense that they were posited to fill a specific role in probability. It could be thought that this role does not mesh with wider scientific views making the positing of such brute entities seem ad hoc. I talk more about this below.

What does it means for a suitable stimulus to be present in order to manifest a disposition or propensity? This is an important question but one beyond the

scope of this thesis. But, however this question is answered, I presume that other causal factors can effect whether a given disposition or propensity manifests. For example, the glass may not smash when I drop it despite its disposition to do so if I previously package it in bubble wrap. For another example it's hard to think of a situation where salt put into unsaturated water wouldn't dissolve.

The propensity theorist claims that there is an analogous type of dispositional properties which ground probabilities. Rather than being deterministic in the same way as dispositions though, they are probabilistic in producing outcomes. Propensity theorists are, for the most part, explicitly realist about the propensities.

As noted, one difficulty in discussing propensity interpretations is that there are almost as many different theories as there are people endorsing them. I will now try to distinguish them.

Long-Run, Single Case and Other Sub-Theories

Gillies (2000) distinguished two broad types of propensity theories; the long-run interpretations and single case interpretations. However, there are fairly distinct theories even within these two sub-divisions.

Every propensity theory adheres to the same basic structure. A stimulus occurs which leads to a disposition bringing about a manifestation. The various sub-theories differ on how to understand this basic structure[22]. Therefore there are three broad questions which any propensity theory has to answer. The first is

[22]Thanks to Luke Fenton-Glynn for his exegesis of the propensity theory literature.

what possesses the propensity? Is it the single-trial or is it a long-run trial? Is it an object, a set of objects or an event? The second question is, what exactly is the probability? Is it the propensity itself? Or is it the manifestation brought about by the probabilistic disposition. Thirdly, what exactly is the manifestation? Is it the outcome itself or is it a probability distribution?

Take a coin toss. Does the propensity to give rise to a certain outcome lie in the single toss of the coin, the single coin itself or in a long-run trial of coin tosses? Is the probability the propensity the coin has to manifest heads or tails? And is the manifestation the result itself or the probability distribution of outcomes?

The answers to these questions logically allow for many combinations. However, not all of the combinations are plausible. For example, if you take probability to be the propensity then it's implausible to say the manifestation is a probability distribution. Rather, the manifestation must be an outcome. Conversely, if the probability *is* the manifestation then presumably the manifestation isn't also an event or sequence of events. That leaves four broad views available. The (1) single case theory in which probability is a propensity which manifests in an outcome and the (2) single case theory in which probability is the manifestation of a propensity. The (3) long-run theory in which probability is a propensity to produce a series of outcomes (a frequency) and (4) the long-run theory in which a probability distribution over different sequences of outcomes or frequencies is the manifestation.

Problems For The Propensity Interpretation

There are many potential arguments which could be made against propensity theories (see for example, Eagle (2004)). Does the propensity theorist solve the problems that the frequency theorist struggles with? I will discuss claims that propensity theories do not actually solve either the reference class or single case problems. I will then discuss Humphreys' paradox and questions around the mystery of the propensity property. I will end this section by concluding that propensity theories do not look promising as an interpretation of probability.

Do propensity theories suffer from the reference class problem? Propensity theorists try to over come it by placing the probability in a property instantiated in a physical set up. According to Eagle this set up would then "contain all the statistically relevant features of the situation, and would thus uniquely classify each single event into a probability space" ((2004) p.393). Eagle, however, argues that neither long-run nor single case theories actually manage to avoid the problem. This is perhaps clearer in the long-run trial version as it more straightforwardly inherits the problem from the frequency analysis. Take the coin example once again. What exactly counts as the experimental set up? You'll need to know this in order to know what counts as a repetition of it. Is it the class of all coin tosses? The class of all fair coin tosses? The class of all fair 10p coin tosses? Intuition on the matter won't do, we need some kind of guidance provided by the theory. The addition of a propensity to the mix doesn't help because there is nothing in the theory to tell us which things exactly instantiate

the property.

Now I will discuss whether propensity theories solve the single case problem. As mentioned above the frequentist Von Mises acknowledged this problem and dealt with it by denying that such single case probabilities were meaningful. Propensity theorists don't have to do this. What arises in fact is almost the opposite problem, what Eagle calls the 'generalisation failure' problem.[23] This problem applies to propensity theories which take the single case propensity as brute. Series can then 'inherit' their probabilities from the single case it is a trial of. If single case chances are considered brute or fundamental then the question becomes "how should we classify (which are) the statistically relevant properties" (Eagle (2004) p.395) of that single case. In other words, it becomes impossible to generalise from the single case because it's impossible to pick out which properties are the relevant ones to generalise. Howson points out that in the single case every property is a statistically relevant property. Which means there is no way to hold them all fixed and still abstract away from them. Ultimately the reference class for the single case just is the single case. Take again the example of my catching the flu and let's say it has a well defined probability. Which class of trials could this generalise to? The class of all humans? The class of all PhD students? The class which just contains me? Which properties of the case must be held fixed and which can be abstracted from? Because the single case is taken to be brute there's no way to distinguish relevant and irrelevant properties and therefore it becomes its own reference class. It seems

[23]Though this point was originally raised by Howson (1984).

then that propensity analyses do suffer from a single case problem, albeit in a different form.

Humphreys' (1985) paradox is a problem for those propensity theories which treat probabilities as identical to propensities (in other words with theories which claim the probability *is* the propensity with the power to manifest an outcome). Dispositions are treated as causal type phenomena. They display asymmetry in the same way as causal phenomena do. A fragile glass has the disposition to break when dropped but no disposition to be dropped when broken.

Analogously with propensities. There may be some propensities which do display a symmetrical structure, but these will be special cases and the vast majority are asymmetric. Humphreys gives the example of smoking and lung cancer ((1985) p.559). There is a propensity for smoking to produce lung cancer and you can work out what this probability is. Further, using Bayes Theorem you can calculate the probability of having smoked given the presence of lung cancer. But there is no propensity for lung cancer to make people have smoked in the past.

Humphreys also gives the compelling example of a man who is made very grumpy in hot weather ((1985) p.565). The hot weather has the propensity to make the man shout at his wife but there is no inverse propensity for the man shouting at his wife to cause hot weather. Propensity accounts which posit a close relationship with causation therefore fail to obey the axioms of probability. However, as noted above this only applies to those theories in which probability

is the propensity. Those theories which claim the probability is the manifested distribution do not face this problem precisely because probability in these theories is an outcome and not something that plays a causal type role.

There is one problem which applies to all kinds of propensity theories but not to frequency interpretations. That is there is a lack of independent empirical evidence that these dispositions exist in the sense that propensity theories claim. The propensity theorists claim is substantive. The propensities are real physical properties. However, they do not independently occur in our best scientific theories. The evidence we have for their existence is the frequencies they're used to explain. And, as an adequate best systems theory demonstrates[24], they are not necessary for giving an explanation of the phenomena we see. Another way to put this is to say that the propensity property, the property a majority of propensity theories explicitly hold is real and not reducible, is mysterious. What this property is and how it actually explains probability distributions or probabilistic outcomes is mysterious. Unlike standard dispositions such as fragility, they are not really enmeshed in our wider scientific view point. There are no laws about propensities. They appear to be more like ad hoc epicycles introduced to fix a specific problem. Indeed, Popper anticipates that a frequentist may make the claim that "propensities are thus introduced in order to help us to explain, and to predict, the statistical properties of certain sequences; and *this is their sole function*" ((1959) p.30, emphasis in original). This is not a knock-down

[24]And as I will go on to say in Section 4.1.6, I believe an adequate best system theory can be found.

problem in itself, but does suggest the ontological cost might not be worth it. At least there is a burden of proof placed on the propensity theorist to explain why the explanatory value of a propensity property outweighs this ontological cost.

Though propensity theories do make headway on some of the problems faced by other theories, they do not actually avoid all of the frequency analysis problems and indeed raise difficult questions of their own, I would argue that they are not preferable theories, at least when compared to their frequentist cousins. Especially when the mysterious nature of the propensity property is taken into account. There is one last type of theory I want to examine which I will move onto now and which I believe to be the most promising.

4.1.6 Best System Probability

The newest theory of probability on this list (created by Lewis (1987*a*) but based on work by Mill (1846) and Ramsey (1978)) is specifically designed to deal with some of the problems which frequentist and propensity accounts face. Like them it is an objective theory. Lewis intended his best system theory to be an analysis of both laws and chances. A theory or system is a collection of true axioms which pertain to events occurring throughout space-time. Laws are the "generalisations that appear as axioms" ((1987*b*) p.128) and theorems in the best system. Chances are those probabilities which are given by the laws of the best system.

So what are the theoretical virtues Lewis uses to assess scientific theories? The

best theory is the true theory with the best balance between simplicity, strength and fit. Simplicity can be a difficult virtue to define. In this case its designed to be an ontological principle about the number and types of things you have to posit along with the complexity of the laws of the system. The strength of a theory is its explanatory power. Fit is included among the virtues in order to introduce chance into the best system theory. In order to ascertain the 'fit' of a given theory, you have to calculate how probable the actual history of the world is given the probabilities posited by the theory. So Theory A, may have the consequence that it renders the actual history of the world very unlikely. That theory would have a poor fit. Theory B on the other hand results in the actual history of the world being more probable (relative[25] to Theory A) so it fits that history better. Therefore Theory B is preferable to Theory A in relation to fit. If Theory B also has at least as much, if not more explanatory power and simplicity, then it is preferable to Theory A.

Of course, deciding which theory achieves this balance best is a matter of subjective judgement to a degree although, "not just anything goes" ((1994) p.479). For example, Lewis' own modal realism (see 1986*a*) may be considered very unparsimonious but Lewis himself argues that as he is not positing new kinds of things it is not in fact as ontologically profligate as it may appear[26]. In so far as defining what 'simplicity', 'strength' and 'fit' mean (as well as how they

[25]As Lewis notes, "It may well turn out that no otherwise satisfactory system makes the chance of the actual course of history very high; for this chance will come out as a product of chances for astronomically many chance events" ((1994) p.480).

[26]To clarify, I don't mean to imply here that modal realism is a Lewisian system. I'm just drawing an analogy here to theories which may be more parsimonious than they seem at first glance.

should be balanced) this is a problem for Lewis. I do not think it is a fatal problem however. There is broad agreement on such terms and expecting exact precision may be unreasonable. The fact that there is broad scientific agreement on theory choice suggests that agreeing on how to understand and balance the Lewisian virtues in practice is possible and not fundamentally problematic. Furthermore, "our standards of simplicity and strengths and balance are only partly a matter of psychology" ((1994) p.479), there are also objective measures we can apply.[27] To make a crude example, a system with ten rules is objectively simpler than one with 1,000 all else held equal.

Furthermore, I don't think it is necessary to define such terms absolutely, they only need have meaning relatively. Lewis himself considered this problem. But he relied on the idea that "If nature is kind to us, the problem needn't arise" (1994 p.479). In other words, hopefully one theory would so obviously come out as frontrunner that the ambiguity would be unproblematic. I find this convincing. The history of science has followed this pattern. Quantum Mechanics and General Relativity both appeared (and continue to appear) as front runners given our epistemic state from the mid 20th century to now. Newtonian Mechanics was the clear front runner previously. There's no obvious reason why this pattern shouldn't continue as we evolve and increase our evidence base. Indeed, there's no reason to think that this wouldn't be true relative to a complete evidence base (should we ever be able to find such a

[27] It should be noted also, that once the best system has been chosen the laws and chances given by the system are objective. Nature, Lewis says, "determines what's true about the laws and chances" ((1994) pp.481-482).

thing or anything close to it).

So, in summary, compared to some of the difficulties faced by other interpretations of probability, this ambiguity seems less worrying to me. It looks as though nature *is* kind to us.

How does this produce an interpretation of probability? You can simply read off the probabilities assigned by the laws of the Best Theory. If your best theory is a deterministic one then all the laws within will return trivial probabilities. If however your theory contains probabilistic laws, then it will return non-trivial probabilities. So far it seems to be better suited to my needs than Bayesian accounts in that its an objective theory. It also seems to have an advantage over propensity accounts in that it does not have to posit a new (potentially ad hoc) property. And, as regards frequentist accounts, the Best Systems approach seems advantageous in that it does not rely on frequencies in quite the same way.

Furthermore, in some cases (see the unobtainium case below as an example) the Best Systems approach can provide a probability where trials would be hard or impossible to run, and as such should be able to handle single cases. This is because the single case probability will not be based on a single case trial or on a propensity property only ascertainable through (a) trial(s). It will be based on what the laws of the theory suggest the probability is[28]. To return to

[28]Hájek (2012) also suggests that this may solve the frequentist problem that probabilities cannot be irrational. This is because laws are not limited to producing rational probabilities in the way that frequency methods are. Lewis also makes this point in ((1994) p.477-478).

Lewis' ((1994) p.477-478) Unobtainium example as described above in Section 4.1.4.

> "There are general laws of radioactive decay that apply to all atoms... Unobtainium atoms have their chances of decay not in virtue of decay frequencies for unobtainium, but rather in virtue of these general laws." ((1994) p.477-478)

The general laws are not just derived from unobtainium but from all the examples of atoms which radioactively decay. So, the probability of unobtainium decay is 'nested' in these more general scientific laws for which we have more evidence.

Because probabilities are given by the network of laws there's more room for divergence between the probabilities and the frequencies. This gives the Best System approach more wiggle room when it comes to assessing the probability of a hard to run trial (unobtainium) or where a finite frequentist account may give a counterintuitive result. Recall the case in which a coin was flipped 1000 times but stubbornly refused to land 500 times on heads and 500 times on tails. The finite frequentist must say that the probability of landing heads just is the frequency with which heads landed (despite this not being 0.5). The Best Systems analysis would not suffer from this problem because the laws can provide the probability of 0.5 while diverging from the actual frequency in this case. This is because the Best System Analysis can provide a general law which is the best balance between strength, simplicity and fit across a class of molecules or isotopes. It is possible for the actual frequencies to diverge from

this general law without detracting too much from the strength, simplicity and fit of the general law.

Frigg and Hoefer (2015) provide an example of a general law from which actual frequencies may diverge. It relates to cases such as coin flips (although presumably it could be altered to include more cases such as dice rolls) where a "simple rule stating $Pr(H) = Pr(T) = \frac{1}{2}$" ((2015) p.560) is plausibly held to be the strongest, simplest law describing coin flipping behaviour. In prose, the rule states that the probability of landing heads equals the probability of landing tails equals a half. It wins inclusion into a best-system, they argue, because the "derivation-simplicity" ((2015) p.560) it provides outweighs the small drop in simplicity caused by adding a new rule to the system. A law from which probabilities can be easily derived has greater derivation-simplicity than a more accurate law which it's harder to derive the probabilities from. Frigg and Hoefer compare a system which includes '$Pr(H) = Pr(T) = \frac{1}{2}$' as a single law relating to coin flips to a system which calculates the probabilities of coin flip outcomes by factoring in the quantum mechanical and micro-level laws such that "only a Laplace's demon could actually calculate the chance of landing Heads" ((2015) p.560). A small loss of accuracy can be outweighed by a large gain in the derivation-simplicity. This is permissible in Frigg and Hoefer's view because Best Systems are "'guides to life' for epistemically limited beings" ((2015) p.560) so a variety of dimensions of simplicity should be considered.[29]

[29]Other forms of simplicity they discuss include "numerical simplicity" ((2015) p.560) which is the number of laws a given system has and "simplicity of formulation" ((2015) p.561). This regards how simply the law can be formulated in any given natural language such that a law which

An Apparent Contradiction

I have already set to one side the problems for Lewis' account based on interpreting 'simplicity' and other such terms. I have also set aside the potential problem of idealism. What other problems are there for this theory?

Lewis discusses one in "Humean Supervenience Dubugged" ((1994) p.482-483); the problem of undermining. If probabilities are given by the laws of a best theory then they are true because they supervene on the whole of history, not just the past and present[30]. As this is the case, future outcomes must have some effect on present probabilities. Each possible 'future history' has a greater than 0 probability of coming about. To highlight the problem Lewis suggests we take one of these alternative (i.e. non-actual) future histories, *F*, which produces different present chances than the actual future history of the world does. *F* is not the actual future history by stipulation but at the present there is a greater than 0 chance of it coming about. This means that at the present time, there is a chance that events could unfold such that the present chances would be different to what they actually are. This is what Lewis means when he says that "present chances *undermine* themselves" ((1994) p.482). Lewis again;

> "Although there is a certain chance that this future (i.e. *F*) will come about, there is no chance that it will come about while still having the same present chance it actually has. It's not that if this future came

can be stated in one line has greater simplicity of formulation than one which takes many lines.

[30]This is because the 'fit' of a theory is based on its fit with the whole of history. In other words, how well it fits with future outcomes not just past and present outcomes.

> about, the truth about the present would change retrospectively. Rather, it would never have been what it actually is, and would always have been something different." ((1994) pp.482-483)

In so far as this is merely weird or counterintuitive Lewis does not consider this problematic. However, it is problematic in so far as it is ruled out by the Principal Principle. The principle states that our credences should match the objective probabilities where these are known. Take *F*. According to the Principal Principle then someone's credence (*C*) in *F* based on how chances actually are at the present time (*E*) should be $C(F \mid E) \neq 0$. This is because there is a greater than 0 chance of *F* coming about in the present even though it differs from the actual future history. But remember that if *F* were to come about then it would change the present chances, say to *E'*. This means that *F* is inconsistent with *E* which means that $C(F \mid E) = 0$. $C(F \mid E) \neq 0$ and $C(F \mid E) = 0$ cannot both hold at the same time, so this results in a contradiction by the lights of the Principal Principle. This is a huge problem for the best systems interpretation of probability.

Lewis' Solution

Fortunately, Lewis provides a solution to this problem. Undermining is only problematic *if* the Principal Principle holds in this case. It is however, argues Lewis, wrong to apply it to this scenario. This is because *E* includes information about the future which renders the Principal Principle inadmissible when calculating our credences in the present. Present chances on the best systems view will always include information on future chances because present chances su-

pervene on events throughout the whole of spacetime including the future. Therefore any use of the present chances will always be inadmissible and the Principal Principle inapplicable. This is problematic though as Lewis believes that the Principal Principle is a foundational principle (as the name suggests) in understanding probability. So, if it can never apply, this tends to suggest that we will never properly understand probability.

Lewis himself thought he had hit bedrock with this problem until he was given an idea which he credits ((1994) p.473) to Michael Thau[31]. The kernel of the idea is that admissibility admits of degrees and is a relative matter. It doesn't have to be a binary decision as to whether the Principal Principle applies or does not apply in every case. Rather, we can admit that there will always be some inadmissible future-information component of current chances but that this information is of such little use that it renders the Principal Principle only slightly inapplicable. Furthermore, if the future information carried by the current chances applies mainly to future A then the Principal Principle may be inapplicable there, but not inadmissible as regards future B. Given these two vital realisations we can reinstate the Principal Principle as, if not entirely admissible, then admissible enough in the right circumstances. And this means that so far as these circumstances are met, the C(A | E) = P(A) closely enough to save our everyday reasoning on probability.

A set of circumstances which would still rule out the use of the Principal Principal is the undermining scenarios which started the problem. But this is as it should

[31]See Thau's "Undermining and admissibility" (1994).

be. No application of the Principal Principle in this scenario means no credence which means no contradiction[32].

I have now considered all the interpretations of probability I wanted to cover. I find the best system analysis to be the most convincing of the theories given. But before I can conclude this chapter I want to summarise the pros and cons I have just gone through. I then want to analyse exactly what my chosen interpretation of probability would mean in relation to the rest of my thesis.

4.2 Analysis

So, which interpretation of probability do I think looks the strongest for my purposes? Classical and logical accounts have long been superseded by more sophisticated accounts. Bayesian accounts are useful but do not meet my needs for the reason that they interpret probability as being subjective. Frequentist approaches have some appeal in the sense that they equate probability with actual outcomes, but suffer from issues such as the reference class and single case problems. As a good account of what probability is should be able to interpret the quantum phenomena which best describes the world, then these problems are too concerning to overlook. Propensity accounts

[32]There is a problem still however. In so far as we have weakened the Principal Principle we have raised the question as to how chance and credence do exactly relate. Both Lewis and independently Ned Hall (1994) make the same suggestion. For space reasons I will minimise detail but the idea is to update the old principle from C(A | HT) = P(A) to the new principle C(A | HT) = P(A | T) where H is the history and T is the theory of the world. "By conditionalising credence or chance on T, we ignore undermining futures" (Lewis (1994) p.487). The new principle is what the old principle is an approximation of.

then look more promising as they were developed specifically to handle the quantum single case problem. However, they suffer from problems such as the ontological burden of explaining new properties with only limited explanatory benefit. Therefore, the account which I think is the strongest and explains the most phenomena with the least problems is Lewis' best system interpretation.

How does this interpretation of probability fit with various theories of causation; say probability raising, counterfactual theories or Kim's preferred process theory[33].

Probability Raising and Counterfactual Causation and Best System Theories

A comprehensive theory of probabilistic causation will need to appeal to the concepts of probability raising or probabilistic processes or mechanisms. Furthermore, as I've already noted, they will need to appeal to an objectivist theory probability. The Best System account fits the bill in this regard.

There is an issue here though because the laws of the best system are fundamental and therefore `encode' probability at that level. All of our everyday probability reasoning however occurs by definition at a higher level. How do probabilities appear at those higher levels? Possibly the higher level laws could `inherit' their probabilities from entailment from the lower levels. Take ice melting as an example. There is no fundamental ice melting law. But it can be constructed from other fundamental laws from statistical mechanics. So although

[33]I assume that an analogous story could be told for interventionist accounts so for space I will not discuss them here.

the higher level sciences may not have their laws encoded in their own right in the best system, they may still feature in the sense that they, and the probabilities they give, will be deducible from it. Loewer (2001) argues that it may not be possible to deduce statistical mechanical probabilities from quantum mechanics. However, Loewer claims that Lewis' account can be modified by adding a "probability distribution over initial conditions among the axioms of a candidate Best System" ((2001) p.618). This can, according to Loewer, be done with little cost to simplicity at the gain of great fit. Now the 'Albert package' of the fundamental postulates of statistical mechanics can be used as a best system. The Albert Package comprises of three postulates taken from Albert (2000);

> "1. The Newtonian law of motion (which is that $F = ma$).
> 2. The *Past-Hypothesis* (which is that the world first came into being in whatever particular low-entropy highly condensed big-bang sort of macrocondition it is that the normal inferential procedures of cosmology will eventually present to us).
> 3. The *Statistical Postulate* (which is that the right probability-distribution to use for making inferences about the past and the future is the one that's uniform, on the standard measure, over those regions of phase space which are compatible with whatever other information - either in the form of *laws* or in the form of *contingent emirical* facts - we happen to have)." ((2000) p.96)

Schaffer argues however that the Albert package is problematic for use as

a Lewisian style best system as it makes reference to non-perfectly natural properties such as 'low entropy'. "Hence the Albert package *is not even in the running* for the Lewis laws" ((2007) p.130). To counter this problem we could however weaken Lewis' requirement that predicates have to be perfectly as opposed to reasonably natural or on the more perfectly natural end of the spectrum.

So, to summarise, I think a best system interpretation of probability and a probability raising counterfactual theory of causation form the most coherent picture when combined.

5

Positive Arguments for Mental Causation

"The dolphins had always believed that they were far more intelligent than man"

- Douglas Adams (1995)

In this chapter I will lay out three of my positive arguments for the existence of mental causation.[1] The first (laid out in Section 5.1) is the argument from the Mental Manifest Image, the second (laid out in Section 5.2) is an argument from evolution and the third (laid out in Section 5.3) is an inference to the best

[1] My final positive argument will be presented in Chapter 9 on natural kinds.

explanation. I will end in Section 5.4 by introducing and replying to a potential problem with my view.

5.1 The Manifest Image

I shall now introduce my gloss on the Manifest Image (MI) (Sellars (1963)) which is ostensibly in tension with the straightforward physicalist picture:

(MI) The world is how it appears to be to us and we should try to accommodate this in our philosophical theories in so far as is possible given our best current scientific theories.

Thus its good practice to prefer theories which, all other things being equal, accommodate our everyday view of the world over those which are inconsistent with it.

More specifically though, I will be working with the Mental Manifest Image (MMI):

(MMI) Things are how they appear to us in our introspections and mental phenomenology and we should try to accommodate this in our philosophical theories in so far as is possible given our best current scientific theories.

To reiterate, the MMI means that, all else being equal we should prefer scientific theories which maintain and explain our introspective lives as they appear to be, over ones which deny or are inconsistent with our introspective evidence. Part of our everyday, introspective and phenomenological experience is that mental causation occurs and brings about physical effects. You feel pain so you move, you feel thirsty so you drink and it appears to us that these pains and thirsts are the causes. Given that this is the case, we should try to incorporate mental causation into our philosophical world views. Furthermore, given two philosophical theories, one which does and one which doesn't accommodate mental causation, we should prefer the one which does accommodate mental causation.

5.2 The Evolutionary Argument

Aside from the pull of everyday introspection, an evolutionary argument could be given for, at least prima facia, thinking that mental states can be causally efficacious. A staggering number of species, with varying degrees of biological relatedness to each other, have evolved consciousness[2], and this convergent evolution gives us reason to believe that there must be something at work behind this. That so many creatures could evolve to become conscious is then either a huge coincidence, perhaps a by-product or free rider, or it is useful for survival and reproduction and was thus selected for which means it must

[2]Naturally to different degrees, but almost indisputably to some greater or lesser extent. Many animals do not possess self-consciousness, for instance, but it would be more of a surprise to me if at least some non-human animals don't have pain qualia, for example, than if they do.

have causal efficacy. Given the variation and genetic distance in the many examples of this phenomenon, it would seem unlikely to be the former.[3]

One example of a mental state which plausibly has evolved because it helps survival and reproduction is disgust.[4] It seems to have clear evolutionary benefit; those who avoid items which provoke a disgusted response are less likely to get infections or illnesses which could potentially harm or kill them.

As stated above, I take it as given that many creatures have at least some form of consciousness. More specifically, many mammals can be said to have conscious mental states (I take it that of the non-human animals, mammals comprise the most obviously and least controversially conscious). Animals as diverse as cats, elephants, humans and whales, all shared an ancestor but evolved along divergent lines from that ancestor on. That they all are conscious then is perhaps less surprising if their last common ancestor was itself conscious. What remains surprising though is that all these mammalian lines could develop a level of complex consciousness given that it seems unlikely that the last common ancestor of this group would have had this. With no direct proof I can not say that it didn't, but it does seem highly unlikely.

Intriguingly, even some non-mammals can have complex intelligences. Take octopuses for example. Katherine Courage (2013) explains why octopuses are a good example of non-mammalian intelligence;

[3]The sceptic could still press the point that consciousness is a necessary result of the kind of central nervous system which confers evolutionary advantage. However the burden of proof would be on them to show how the neural and not the mental states do all the causal work.

[4]See Curtis et al. (2004).

> "chimpanzees are, like humans, primates. Dolphins are mammals. Even clever crows and ravens are vertebrates. But our last common ancestor with the octopus was probably some kind of wormlike creature with eye spots that lived as many as 75 million years ago, the octopus has a sophisticated intelligence that emerged from an almost entirely different genetic foundation." (2013)

She quotes Peter Godfrey-Smith; "octopuses are the closest thing we have" (2013) to alien intelligence. This makes octopuses a useful example for pressing my point that consciousness likely evolved because it has a causal impact rather than merely free riding on some other adaptation. The argument runs as follows. Both humans and octopuses have complex intelligences and both possess consciousness. Our last common ancestor, while perhaps conscious, almost certainly was neither as intelligent nor as sophisticatedly conscious as either humans or octopuses today. Humans and octopuses diverged evolutionarily so long ago that each has since evolved in very different ways and are now only distantly related. Given this genealogical distance it suggests that consciousness emerged in both cases because it was useful in survival and reproduction and not merely as a free rider or coincidence.[5] This suggests it does make some kind of causal difference. This makes it likely that this is a case of convergent evolution and not a simple shared inheritance.

[5]There will also possibly be energy costs associated with having conscious mental states which would speak against conscious mental states being free riders which adds weight to my argument. However, there is an obvious reply in that it could actually be the underlying brain states which require the energy.

Take the New Caledonian Crow as another example. This is one of only a few species of bird we have so far discovered which use tools.[6] In the New Caledonian Crow's case they trim branches carefully down to hooks to pick insects out of trees. As a bird, humans and crows last common ancestor lies many generations back.[7] Continuing with the assumption that such an early organism can only have had basic consciousness, it seems again we have another case of not merely shared inheritance, but genuinely convergent evolution.

As a final example, take the Leafcutter Ant.[8] This extraordinary ant is known to use agriculture to harvest the fungus it eats. It does this by collecting leaves, carrying them back to the nest and waiting for the fungus to grow. Ant societies are commonly known to be complex, but this is the only example we know of where ants use agriculture. It could be argued in this case that it is not the individual ants which possess the intelligence, but rather the colony of ants as a whole, perhaps in a 'Chinese Brain' style[9]. However, I think what this example shows most clearly is how little we know about non-human animals and their behaviour. Furthermore, it highlights that there may be intelligence in places humans previously would have doubted. This leads me to believe that the more

[6]See Hunt & Gray (2003) and Weir et al. (2002) for more on this fascinating crow.

[7]"The last common ancestor of birds and mammals lived some 300 million years ago, at a time when the six-layered neocortex, which gives rise to sophisticated cognition in primates, had not yet developed" Veit & Nieder (2013).

[8]See Hölldobler & Wilson (2011)

[9]This argument is from Block ((2007*a*) p.71). The thought experiment runs as follows. If every person in China was given a radio that could appropriately connect them to their fellow inhabitant could they act as the mind in an artificial body?

we learn, the more (complex) consciousness we will uncover, all of which lends support to my argument from convergent evolution.

It might be argued that I am conflating consciousness and intelligence or, at least that I'm relying too heavily on intelligence as a proxy for consciousness. To an extent this is true, particularly when it comes to my discussion of ant colonies and Block's 'Chinese Brain'. It could well be that entities such as ant colonies display what might be or look like intelligence without that entity also being conscious. Artificial intelligence displayed by computers also plausibly (even probably) is not (at least for now) attended by conscious experience. Discussion of the exact relationship between intelligence and consciousness, while fascinating, would take us too far afield to consider more deeply. Suffice it to say then, even if some of my speculative examples relating to ants, crows or other animals further away from humans on the evolutionary tree don't convince you, I think there are non-human animals which almost indisputably are conscious, for example, chimpanzees. When it comes to such animals, I think the burden of proof falls on those who argue they have absolutely no conscious experience rather than those who do.

There are many more examples I could give here of putative cases of genuine convergent evolution which would support my case, however for space reasons I will now move on to my third argument for the existence of mental causation; inference to the best explanation and the placebo effect.

5.3 Inference to the Best Explanation

My third argument for the existence of mental causation comes from empirical evidence which leads to an inference to the best explanation. These come from the medical sciences, particularly from psychology. It seems a much better explanation that mental properties are causally efficacious in the case of placebos and talking therapies than the sets of underlying neuron firings. In fact, I take the causal efficacy of mental events to be the best explanation of these phenomena, at least given our current understanding of the brain.

I will start by examining the placebo effect. A placebo is a medical treatment which has casual efficacy based not on some chemical mechanism, but rather one based on expectation. Thus the same drug could be administered to the same patient under different names and be efficacious in one case while failing in the latter (or can be more effective in the former than the latter). The difference in the two cases is the mental state of belief in the first case that it will be efficacious and the belief in the second case that it won't[10]. Alternatively, a sugar pill can be substituted in place of a drug with a known effective mechanism. If the patient believes that the sugar pill is medicine, then by the placebo effect, they can derive curative benefit[11]. In this case, a treatment which is known to have no medically or chemically relevant properties actually proves

[10]The two sets of underlying brain states will also be different but as I will go on to argue it's hard to see what difference in the two sets could be causing an effect whereas the difference in mental states does make explanatory sense.

[11]Although it's also important to note that orthodox medicines also carry their own placebo effect as well. It's almost impossible to avoid.

effective. For example, in one study a placebo marked as an active drug was statistically as effective as the active drug marked as a placebo (Kam-Hansen et al. (2014))[12].

Let's press on the distinctness of the causal mechanisms some more. Say you have two drugs. Effectron (E) is a drug which blocks pain receptors and therefore reduces pain by a chemical mechanism. Pretendtron (P) is a totally chemically inert drug which, by definition, does not interact chemically with the brain in any relevant way. However, when P is represented as being effective its use is shown to reduce pain. It therefore works by a mechanism which must be something other than an orthodox chemical mechanism. As the difference maker in the P case is how the drug is represented to the patient, the putative mechanism at work here is a psychological one based on belief and expectation. This at least places a burden of proof onto someone who wants to deny mental causation to explain why their purely physical mechanism is to be preferred.

What's more, the placebo effect has been shown to be robust. It has been shown there are predictable patterns in response to different types of placebo. For example, studies have shown that a placebo injection of water is more effective than a sugar pill (Goldacre (2009) p.70). The packaging (and price) which medicines are sold in (for) has also been shown to affect their perceived

[12]For another example of a study which found a positive placebo effect see Kaptchuk et al. (2010). Ben Goldacre (2009) also summarises the results of many placebo studies in his very helpful chapter *The Placebo Effect*. However, it is important to bear in mind this is obviously a very small sample of a much larger literature.

efficacy which goes some way to explaining how you can buy the same paracetamol for 30p or £3 (Goldacre (2009) p.71). All this sums to suggest that there is some systematic phenomenon in play with the placebo effect, not mere coincidence.

Furthermore, and in some ways most persuasively, there is evidence to suggest that in some cases, intervening directly on a person's mental states is at least as effective as intervening directly on their physical brain states[13]. This is the case with certain mental health conditions such as depression and anxiety. There are both drug and talking therapies available to treat these conditions. These can both be effective independently, however, it is believed that a combination of treatments is the most effective. In some patients though, only talking therapy proves effective, no amount of drugs will bring about the desired result:

> "CT (Cognitive Therapy) is the best-known and most widely tested of a family of cognitive behavioural interventions. Like ADM (antidepressant medications), it is a safe and efficacious treatment for acute episodes of major depressive disorder. CT is based on the premise that inaccurate beliefs and maladaptive information processing (forming the basis for repetitive negative thinking) have a causal role in depression. This `cognitive model` posits that when maladaptive thinking is corrected, both acute distress and the risk for subsequent symptom return will be reduced. Contrasting with the lack of evidence of enduring effects of ADMs is the substantiation of claims that CT provides

[13]Which would make the mental state the cause on an interventionist account of causation.

> protection against relapse and, possibly, recurrence" (DeRubeis et al. (2008) p.790)

Indeed, the success of counselling as psychological treatment, and the increasing demands for counselling services, speaks (excuse the pun) to how effective intervening directly on mental states to achieve a desired goal is.

Of course, its not entirely impossible that this is not a function of our relative lack of knowledge about the mind, brain chemistry and how the two interact. Future medical science may indeed have the drugs or technology required to most effectively manipulate brain (and mental) states. That said, even in a future where psychiatric treatment has entirely dispensed with all talking therapy, it works, and that's reason enough to examine it more closely. Mental causes would still occur even if we were able to pinpoint more effective physical causes. Even if in the future, another, purely physical, psychiatric approach is put forward, this may be compatible with the existence of mental causation. It is, and always will be the case that mental treatments such as talking therapy work and so mental causation should be carefully considered. Again, this line of argument places a burden of proof on the denier of mental causation to explain away this apparent causal efficacy.

To put my point in a different way, say that C is the effect of being cured, D is the known mechanism of drug efficacy and D* is the mental state of expecting the drug to work. If it's the case that the mental state is doing some additional causal work then the following inequality would hold; $P(C \mid D) < P(C \mid D \& D^*)$. This

is exactly the kind of result placebo studies indicate. Given that the mental state here seems to raise the probability of the event coming about even holding fixed the ordinary biochemical mechanism, it would count as a cause on a picture of probabilistic causation. So it seems we have empirical evidence of the mental bringing about physical effects.

But, is this a case where the only difference is the mental state? It certainly seems so in an everyday sense of talking about the placebo effect. Even the medical literature appears to take it for granted that it is the feeling of expectation, or belief in efficacy, which is making the difference in the two cases. Of course, in both the placebo and the specific pain case it could be the brain state underlying the mental state which is making the medical difference. The placebo effect in and of itself then, cannot be used as a knock-down argument against the CEA. It does not prove that mental states can have non-overdetermining effects over and above their underlying physical state but it is compelling and adds to the burden or proof argument against the opponent of mental causation.

It seems implausible that it's only the associated physical brain states which are making the difference here. Can this implausibility be strengthened by further argument? By examining the placebo effect in the context of scientific laws I argue that we have reason to think the mental is doing causal work. I devote Chapter 9 to the details of this argument so for now I will give an outline only. In order to play a role in any scientific law, the phenomenon in question needs to

be a natural kind[14]. Take the counselling case again. What would the neural correlate of a conversation be? A mass of neural firings. Take the placebo case again. What would the neural correlate of an expectation be? Again, a mass of neural firings. Not plausibly a natural kind, or at least not a particularly perfectly natural one. Maybe, the mental state could play the required natural kind role here? And if mental states provide a better candidate for natural kinds, and therefore are more suited to feature in laws, then is this not reason to think that they may be playing a causal role?

In summary, the medical community treats the placebo effect as though it works based on expectation. This places the burden of proof onto those who argue against the placebo effect and mental causation more generally.

5.4 A Potential Problem

There is a potential problem with my arguments for mental causation that I would like to address before continuing[15]). It might be thought strange that the difference between mental causation existing and not existing is just the difference between whether determinism holds or we live in a probabilistic world. After all, why would the MMI not hold in a deterministic world? If there's only room for mental causation in a probabilistic world, then any phenomenal experience of mental causation would be an illusion. Similarly, consciousness

[14]See for example Goodman's debate on 'grue' and 'bleen' in ((1983), (2000)). For related discussion see Kim (1992) on the relationship between jade, jadeite and nephrite.

[15]I will also return to this in the Conclusion, see Section 10.2.2

could have evolved in a deterministic world, perhaps by free riding off its physical base. That said, it's not clear that mental causation can't exist in deterministic worlds. Indeed, none of the arguments I referenced in Chapter 5 depend upon the world being probabilistic. Furthermore, the argument I will go on to make in Chapter 9 centres on natural kinds and could plausibly hold in either a deterministic or a probabilistic world.

However, what reason is there to think that in a deterministic world, we would have the same MMI? There would be a different physical picture so why not a corresponding different MMI? In fact, given that the physical facts would be different, should it not be expected that the mental facts would be too? Why would we have the phenomenal experience of mental causation if mental causation didn't exist? So, perhaps in a deterministic world, our MMI would be different and we wouldn't have the phenomenal experience of mental causation.

I have made the assumption that we live in a probabilistic world, therefore I have to assume that we only have access to introspective experience of a probabilistic world. If someone was to experience our MMI, it would be an illusion in such a deterministic world. It seems unlikely however that there would be such a widespread, systematic illusion so it seems likely that people in deterministic worlds would have different MMIs to us. We have no reason to think that beings in a deterministic world would feel as though they have the same experiences of autonomy or freedom of the will. In summary, either beings in deterministic worlds share our MMI but it is an illusion, or, they have a different MMI.

So the best explanation of why the MMI holds true, is that ours is a probabilistic world in which mental causation does in fact exist. Our MMI depends on mental causation and in turn my argument for mental causation which casts the veracity of the thesis of causal closure into doubt, depends on the world being probabilistic. So why think that our MMI would be the same in a deterministic world? And, importantly, if deterministic beings do not have an MMI which includes experiences of mental causation, then there is no apparent tension which needs philosophical explanation.

6

The Causal Exclusion Argument And Non-Identity Premise

"O wad some pow'r the giftie gie
us, to see oursels as others see us!"

To a Louse - Robert Burns (1788)

To deny the veracity of the MMI and our introspective experiences of our own mental states is to deny some very direct and compelling evidence, and would be a huge bullet to bite. It is my aim therefore to argue that the CEA is unsound. I have briefly introduced Kim's version of the argument above but now it is time for me to set out the CEA a little more clearly. The deterministic CEA runs as follows:

(P1) Causal Closure of Physics

Every physical event has a sufficient physical cause.

(P2) No Systematic Overdetermination

It is not systematically the case that there are multiple minimally sufficient causes of any given event which exist simultaneously.

(P3) Non-Identity of the Physical and the Mental

There are physical states and mental states and these are not identical to each other.

(C) There is no Mental Causation

Mental states are not causes of physical effects.

Causal closure (P1) is the principle that for any physical effect, there is a sufficient physical cause. Kim states "if you pick any physical event and trace out its causal ancestry or posterity, that will never take you outside the physical domain" ((1998) p.40). While no conclusive argument can be given for holding causal closure, it seems natural that a physicalist would be loathe to give it up. This is because Kim notes, giving up on causal closure means giving up on the completeness of physics.

(P2) states that there is not any systematic overdetermination in the world. This means that there should not be more than one sufficient cause at any given time for any given event in a widespread way. If physical states and mental states are both causally efficacious (and not identical) then this could

potentially result in widespread systematic overdetermination. This is because a wide range of events would have more than one non-identical sufficient cause; the physical brain state and the mental state. Kim argues that while occasional cases of overdetermination happen (think of firing squad cases[1]), this would be strange at a systematic level. If mental causation was always a case of overdetermination then that would qualify as systematic.

Lastly, in (P3), Kim makes the anti-reductionist stipulation that there is no identity between the mental and the physical. If the above three premises hold then the CEA goes through and (C1) follows. If the mental is adding nothing to our theories about the world, because it is doing no causal work, then there is no need to posit it.

Kim rejects (P3), the Non-Identity premise, in order to dissolve the CEA. In support of his reductionism, Kim offers the analogy with properties from the special sciences. Like mental properties, special science properties supervene on basic physical properties. Kim argues that the reduction of special science properties to basic physical properties seems unproblematic.[2] Why not then, Kim suggests, hold the same for mental properties? I will not take the same route as Kim in my disagreement with the CEA as I will accept the non-identity premise. I give three arguments as to why in the second half of this chapter.

So why is causal closure (P1) necessary for the CEA to hold? It is required

[1]Although perhaps a more fine grained view of the effects in such cases would overcome any overdetermination worries.

[2]Although this is not a stance agreed upon by everyone, for example see Fodor ((1974), (1997)).

because if mental causation exists, then mental events must be efficacious precisely because they are mental events (in a Kimean sense of events) or, if you take a more coarse grained view of events (in a Davidsonian sense), because of their mental aspect[3]. And if causal closure didn't hold then the causal efficacy of these non-physical states would not create widespread systematic overdetermination. Note again that causal closure says nothing about mental states causing mental events and thus that phenomenon is not the target of the CEA or my discussion. Mental to mental causation is not the phenomenon at hand.

Say, while stipulating non-identity, that the CEA is valid. If you do believe that mental states are causally efficacious then either causal closure (P1) or no overdetermination (P2) must be false. If you take an event, say my drinking some water, and trace the causal ancestry of my drinking (a physical event), in order to find a sufficient cause, you will have to look outside of the physical realm to my thirst (a mental event). This violates causal closure. However, if causal closure did hold, then the physical basis would be sufficient and the mental cause, if it exists, would be superfluous. The event of my drinking would therefore be overdetermined. This would of course apply to all examples of mental causation. The causal exclusion argument would then, if correct, lead us to believe that the mental has no causal power on the physical. This is because, if every event has a sufficient physical cause, any and all mental causes would be superfluous (and is indeed ruled out by (P2)).

[3]See Section 2.3 for my discussion of this issue.

But why think widespread and systematic overdetermination doesn't happen? Firstly, intuitively it seems odd (Kim calls it "implausible" ((1998) p.44). Secondly, parsimony would suggest that, all else being equal, we should prefer theories which do not allow for the overdetermination of effects rather than those which do. Or more precisely, we should prefer theories which do not posit more than one sufficient cause for a given effect if one will do. Neither of these are knock down arguments however. Intuition and parsimony are both good guides, but not final words. I will examine this premise in more detail in Chapter 7.

The CEA appears to require our world be deterministic. That is because causal closure (in the sense of physical effects having *sufficient* physical causes as opposed to a probability fixing sense) only seems to hold in a deterministic world, if it holds at all. As I've stated though, it seems unlikely that the world is indeed fundamentally deterministic. It seems that to be as charitable to the CEA as I can be while assessing it in a probabilistic setting, I will have to rephrase it into probabilistic terms. This is done in the probabilistic analogue CEA':

(P1') Probabilistic Causal Closure of Physics
Every physical event has a physical cause which is sufficient to fix its probability.

(P2') No Systematic Overdetermination of Probabilities
It is not systematically the case that there are multiple sets of events that are minimally sufficient to fix the probability of a further event which exist simultaneously.

(P3) Non-Identity of the Physical and the Mental

There are physical states and mental states and these are not identical to each other.

(C) There is no Mental Causation
Mental states are not causes of physical effects.

There is no difference in conclusion between the two versions of the argument and likewise no difference in the third premise. The first premise of the analogue differs from its deterministic version in that, rather than fixing whether an event occurs or not, the physical is sufficient to fix the probability of an event occurring or not. Likewise, in the second premise, rather than it being events which are overdetermined, it's the probability of the events occurring which are or are not overdetermined.

To try to make my comparison clearer I have set out a table below which summarises the validity and soundness of the CEA and a potential analogous CEA in both a deterministic and probabilistic setting.

Original CEA	**Analogous CEA**
Deterministic World	
✓ Valid ? Sound	
Probabilistic World	
? Valid X Sound	? Valid ? Sound

Table 1: Comparing the CEA and the Analogue CEA

The table shows that in a deterministic world the Original CEA is valid but potentially unsound. This is because there are some arguments which can be put forward which cast doubt on the CEA even in deterministic worlds. These include (among others) the arguments I laid out in Chapter 5 and the argument from Natural Kinds which I will lay out in Chapter 9 not to mention Kim's own refutation of the non-identity premise.

Furthermore in a deterministic world, the Analogous CEA would collapse into the original as a special case. This is because in a deterministic world, all probabilities are trivially 1 or 0. In that case P(1') "every physical event has a physical cause which is sufficient to fix its probability" amounts to "every physical event has a physical cause which is sufficient to fix its probability to 1". This is the same as (P1) "Every physical event has a sufficient physical cause". Likewise, (P2') amounts to (P2) when the probabilities are 1 or 0.

Presumably then, if the original CEA is valid but potentially unsound in deterministic worlds, then so is the analogous version. However, it is probabilistic worlds I'm most interested in so I will not delve too deeply into the validity or soundness of the CEA in deterministic worlds.

On the other hand, when analysed in a probabilistic world, I argue the original CEA is unsound because the Thesis of Causal Closure is false in such worlds. I will discuss this issue fully in Chapter 8. I will therefore leave discussion of whether the Analogue CEA is valid and sound in a probabilistic world until then.

Perhaps counterintuitively I will now discuss the final premise of the CEA before

moving on to the first and second. This is because I accept this premise and therefore I'll have the least to say about it.

6.1 The Non-Identity of the Mental and the Physical

For the second half of this Chapter I will introduce the premise that there is no identity between the mental and the physical. I will spell out what exactly this premise means and why Kim rejects it. I will then give three reasons as to why I accept the premise.

Kim rejects this premise allowing him to get round his own argument and avoid the unpalatable conclusion that no mental state or property can bring about physical events. How does denying this premise avoid the unwanted conclusion? It works because if the mental and the physical are identical then they can both be playing a causal role without any overdetermination taking place and without violating the principle of causal closure.

Kim is therefore a reductive physicalist. However, many philosophers find this conclusion to be just as unpalatable as the conclusion Kim manages to avoid. Hence, the many different types and kinds of physicalisms and the many and varied attempts to work around the CEA by different methods, including mine. Why might someone find the idea of the reduction of the mental to the physical unpalatable? In order to answer this question I will now spell out what it means for the mental to be identical with the physical.

An early and well known proponent of Identity Theory was J.J.C. Smart (1959).

His was a *type* identity theory (as opposed to a *token* identity theory). The idea behind identity theory is that there is nothing over and above the mental than the physical state associated with it. In token identity theory each particular mental state is identical to some brain state. Whereas type identity theory says that each type of mental state or property has its type of identical brain state.[4] Analogies have been put forward to water being identical to H_2O (Kripke (1980)) and lightning being identical to electrical discharge (Smart (1959) p.145). Whether such paradigmatic cases of physical reduction apply to the mind is of course a further and contested issue. According to identity theories, it may be possible (whether or not this may ever be the case in practice) to reduce every mental state to its underlying physical brain state. The famous empirical example given is 'pain' and 'C-fibres firing'.[5]. So every case of pain is a C-fibre firing and every case of C-fibre firing is a case of pain[6]

How does this premise work when placed into a probabilistic setting? As stated above, there is little difference between the two settings for this premise. Whether or not the world, and causation is deterministic or probabilistic, all this premise states is that the mental and the physical are not identical. Or, in other

[4]It should be noted that you'd have to adopt a type identity theory in order to avoid the conclusion of the CEA.

[5]See Rorty (1965) for example.

[6]Empirically this is now known to be extremely simplistic both because C-fibres are responsible for other sensations than pain and because there are other fibres associated with pain sensations. See (Puccetti (1977) p.303). Scientists now believe they have narrowed down at least two kinds of fibres the firings of which produce difference kinds of pain sensation. They call these C-fibre and Aδ-fibre firing respectively. A study "identified 'pricking', 'dull' and 'pressing' as distinguishing best between Aδ mediated (punctate pressure) and C fibre mediated (blunt pressure) pain sensations" (Beissner et al. (2010) p.3). This in itself is not important to the identity theorist as it does nothing to show that their theory is wrong, but is empirically interesting.

words, the former is not reducible to the latter. This could be the case, or not the case, in either a probabilistic or a deterministic world. Therefore, the three arguments I'm about to present do not depend on what kind of world we're in. The three arguments are well known; Levine's (1983) 'Explanatory Gap', Nagel's (1974) 'What it's Likeness' and Jackson's (1986) 'Knowledge Argument'. I will only briefly summarise each as I want to focus my discussion on the other two premises of the CEA.

6.1.1 The Explanatory Gap

One reason to think that Identity Theory may be wrong is the 'explanatory gap' (Levine (1983)). Levine's argument is that reducing phenomenal mental states to mere physical brain states misses something about the mental states in question. The thing this reduction misses is precisely the phenomenal quality of the mental states. Pain is a great example of such a mental state. There is nothing in Identity Theory, or indeed in *any* reductive physicalist theory which explains why pain feels the way it does. There is nothing in Identity Theory, or again, in *any* physicalist theory, which explains why this mental state is associated with this brain state.

Levine contrasts pain with heat. For the type identity theorist, heat is nothing over and above the motion of molecules. Trying to imagine heat without the movement of molecules is not possible in the way imagining pain without c-fibres firing is. And this is because;

> "The experience of pain, the sensation of pain, counts as pain itself. We cannot make the distinction here, as we can with heat, between the way it appears to us and the phenomenon itself." ((1983) p.355)

What does Levine mean when he distinguishes the way heat appears to us and the phenomenon of heat itself? Levine ((1983) p.358) notes that there is a phenomenal aspect of heat, how it feels when you warm your hands over a fire for example. But then there is the phenomenon of heat itself which just *is* the movement of molecules. There is no explanatory gap with regards to the phenomenon of heat as once you know all the physics of the situation, you understand everything there is to understand about the phenomenon. In regards to the experience of heat, there is an explanatory gap, but that's not surprising according to Levine since "it is precisely phenomenal properties - how it is for us to be in certain mental (including perceptual) states - which seem to resist physical (including functional) explanations" ((1983) p.358).

In the case of pain however, there is no way to make this distinction. The phenomenon of pain *is* how it feels to us. There is no analogous physical story (involving c-fibres firing and so on) which will completely explain the phenomenon of pain because, unlike in the heat example, the *feeling* of pain requires explanation. The physical description therefore leaves something out and is not identical with pain.

Weaker versions of the argument claim that this issue is merely practical or technical, not that there is *in principle* no systematic way of explaining the

gap.[7]

Why does a non-reductionist view not suffer from this problem? This is because a non-reductionist can explain the phenomenal character of pain as brute or basic[8]. If qualia such as pain are basic or fundamental, then we should not expect a further explanation to be forthcoming and the gap is no longer problematic. Another way to look at this is that there seems to be nothing mysterious about heat being nothing more than the movement of molecules. On the other hand there does seem to be something mysterious about the phenomenon of pain just being the firing of c-fibers (or any other neuronal firings for that matter). This mystery implies that something has been left unexplained by the physical explanation alone.

In summary, if there is, as Levine argues, an explanatory gap between the mental and the physical, then they cannot be identical.

6.1.2 Nagel and 'What It's Likeness'

The second argument I reference to motivate the belief that the mental and the physical are not identical is from Nagel (1974). Nagel argues that the mind-body problem is uniquely different from other kinds of reductionist questions. This is because of the phenomenon of consciousness. Nagel's argument is that

[7]There are two options to sidestep the problem however. You could deny identity between the mental and the physical and explain phenomenal states as basic or you could take an eliminative stance as Levine suggests. I want to resist taking an eliminative stance myself as I believe this would run contrary to the MMI, a methodological principle which guides this whole thesis. Thus, I think to take an eliminativist path would be to misrepresent the world.

[8]Although it should be noted this isn't the route Levine takes; he takes an eliminative stance.

reductionist accounts fail to explain consciousness adequately. Although there may be more to the story "fundamentally an organism has conscious mental states if and only if there is something it is like to *be* that organism - something it is like *for* the organism" ((1974) p.166). Nagel claims that an objective reductionist or physicalist account will in principle be unable to capture the "single point of view" ((1974) p.167) which is essentially connected to the subjective conscious experience. Nagel uses bats to illustrate his point. As they echo-locate, and as we assume that they have phenomenal experience, there must be something it is like to sense by echo-location. But, what it must be like will be so different from anything humans can experience based on our sense modalities that we would be unable to imagine it. Therefore, there is a fact and that fact is inaccessible (and possibly even inexpressible) for us. As physicalism and reductionism are both objective theories (moving further from individual perspectives) they are both unsuitable for capturing the essentially subjective nature of phenomenal experience.

So, if Nagel is right and there is something about the mental that the physical story misses, then the two cannot be identical.

6.1.3 The Knowledge Argument

Frank Jackson's influential paper 'What Mary Didn't Know' (1986) was designed to question our overwhelming physicalist preoccupation. The argument goes that Mary is a future 'superscientist' in that she knows all there is to know about a by now complete physical science. However, sadly for poor Mary, she has

been trapped in one room for her entire life, and even more sadly, that room is entirely black and white. Putting to one side the possibility of this (let alone the moral implications) let's assume she has never before observed colour. One day, Mary is released from her room and on that day observes colours for the first time. When she sees a rose for the first time she learns something that she, in principle, could never know in her room; what red actually looks like. The argument is that her physical education, complete though it was stipulated to be, actually left something out about the world, that thing being qualia, or the phenomenal qualities to the physical structures. As qualia are essential to certain mental states, and further, as qualia are non-physical, there are certain types of (phenomenal) mental states which are non-physical. Because it was stipulated that Mary knew all there was to know about the physical, when she left the room and learnt something new, she must have learnt something non-physical. Therefore the mental and the physical can not be identical.

6.2 Accepting the Premise

So to summarise, why do I think it is the case that the mental and the physical are not identical? I am convinced by the arguments put forward by Levine, Jackson and Nagel that there is something that reductive accounts of the mental miss about the phenomenon. Of course, there is some connection between the mental and the physical, however, I agree with a majority of contemporary philosophers of mind that the relationship is not one of identity.

Rather, although I don't wish to commit to a precise view of this relationship, I assume it will be a supervenient or nomic connection.

7

CAUSATION AND OVERDETERMINATION

"It all depends if you've smeddum or not"

Smeddum - Lewis Grassic Gibbon

(2001)

An event can be said to be overdetermined if it has more than one distinct, sufficient set of causes all obtaining at the same time.[1] More specifically;

> "say that a set of events A *overdetermines* event b if and only if (i) b would still have occurred if any member of A had not occurred while all the others had, (ii) b would not have occurred if none of

[1] A quick note on the epigraph for this chapter. Smeddum is a Scots word for grit or determination.

> the members of A had occurred, and (iii) all members of A have an equally good (or bad) claim to be a cause of b." (Kroedel (2008) p.128-129)

I will now examine arguments from both Sider and Bennett that it is not problematic for metaphysical theories if they posit mental causation as systematically overdetermining its effects. As I'm interested in probabilistic worlds, I will then discuss whether overdetermination occurs in those and whether the aforementioned arguments still hold in such worlds. I will conclude that it is not clear that overdetermination is problematic and that questioning (P2) or (P2'), the No Overdetermination premise of the CEA, remains a viable way of defending the causal efficacy of mental states.

7.1 Denying The Problem Of Overdetermination

Perhaps overdetermination is in fact a common and widespread but unproblematic phenomenon? If this is the case then the CEA would be unsound. Mental states could then be causally efficacious along side physical states without this being a problem even if causal closure and non-identity both hold.

7.1.1 Sider

Sider (2003) says he doesn't understand why overdetermination should be seen to be a bad thing. He thinks it's only natural to want to say a given event can

be caused by both a macro-object as well as the micro-parts of the object.[2] Kim thinks such views are "at best extremely odd" ((1993*a*) p.247) but Sider says he isn't sure what this actually means. So, he outlines three different possible objections to overdetermination; (i) metaphysical incoherence, (ii) coincidence and (iii) epistemic doubts.[3] He concludes by explaining why these objections are not actually persuasive.

Regarding (i), metaphysical incoherence, this is the idea that overdetermination is incompatible with theories of causation. However, Sider argues that no currently popular account of causation (he lists counterfactual, covering law, probabilistic and primitivist accounts) is actually incompatible with overdetermination.[4] Sider provides some rough examples to make his point. A window smashing can counterfactually depend on both a macro-object and its atoms given that the macro-object and the macro-object's atoms stand in a mereological relationship. An effect can have its probability of occurring raised by both a mental state and a physical state. To merely state that theories of causation *shouldn't* allow overdetermination would be arbitrary and would rule

[2]As Sider notes this kind of relationship is different from but analogous to the relationship between the mental and the physical. I will discuss whether this difference is relevant later in this section.

[3]Sider actually also includes a fourth argument which he terms a "phantom complaint" ((2003) p.721) so I will not discuss it here.

[4]Furthermore, they shouldn't be given that at least occasional examples of overdetermination (think firing squad cases) do happen. It's important to note that Kim is only objecting to widespread and systematic overdetermination. You might question why systematic overdetermination should be any more objectionable than occasional overdetermination? After all, if a metaphysical theory isn't inconsistent with occasional overdetermination then why would it be inconsistent with widespread overdetermination? Kim's problem likely lies more with the level of coincidence that widespread overdetermination implies, which is not present in the occasional case. Sider also raises this point which I discuss later in this section.

out all the aforementioned theories.

As an example of a contemporary theory of causation which can accommodate overdetermination take Hitchcock's (2007) Structural Equation Analysis of causation which is designed to handle cases of pre-emption and overdetermination. Recall the following simple example from Hitchcock (2019) that I laid out in Section 3.1.3. There are three variables in the model:

B = 1 if Billy throws his rock, 0 if he doesn't
S = 1 if Suzy throws her rock, 0 if she doesn't
W = 1 if the window shatters, 0 if it doesn't

Let's suppose this time that both Billy and Suzy throw their rocks. So, the model has the following equations:

B = 1
S = 1
W = max(B,S)

The equation for W states that if either Billy or Suzy throw their rock, then the window will shatter. Interventions are used to determine which variables are actual causes in the models they appear in. The actual value of B was 1, so Hitchcock suggests that we hold this value fixed and see if the effect (window smashing) still occurs when we vary the value of S. The actual value of S was 1 so to vary it we set it to 0. Now the maximum value of B and S is 1 therefore

the window still shatters. This shows that Suzy's throwing the rock didn't cause the window to smash. The same can be said of Billy's throw. This can also be represented in a directed causal graph[5]:

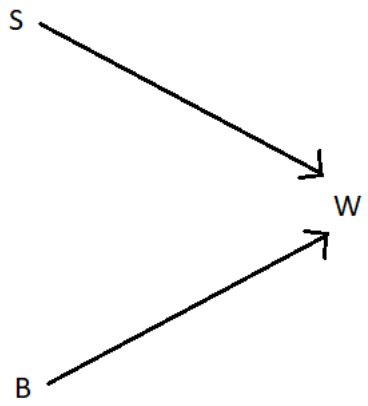

Figure 7.1 An Example of Overdetermination

The direction of the arrows show the direction of the causal relationships.[6] We can see this graph represents the above equations as there is an arrow leading from B to W and likewise from S to W representing that were either to throw their rocks the window would smash. We can now see that holding fixed one or the other doesn't prevent the smashing of the window by removing one or other of the paths. To do this we can perform "arrow breaking" interventions (Woodward (2015) p.316) on Billy and Suzy's throws as below:

[5]For more on causal graphs, see Hitchcock (2001*a*).
[6]Although arrows in causal graphs don't necessarily show actual causal relations.

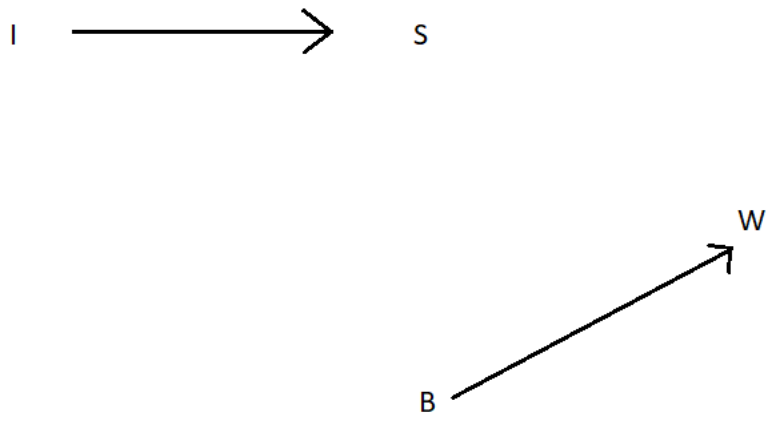

Figure 7.2 representing an arrow breaking intervention (I) on Suzy's throw. A parallel diagram could be made to show an arrow breaking intervention on Billy's throw which leaves Suzy's arrow unbroken.

By intervening to break the arrow we can show that the window smashing does not depend on either throw, and thus that this is a case of overdetermination. However, using these models, a general counterfactually based analysis of causation can be given. The definition given below is a simplified version of the one Hitchcock gives in Hitchcock (2001*a*).

> $X = x$ is an actual cause of $Y = y$ in world w just in case:
> X and Y are distinct variables representing metaphysically independent events,
> $X = x$ and $Y = y$ in w,

> There exists a path from X to Y in an apt model[7] and a pair of possible values x' and y' such that if X had the value $X = x'$ and all variables not on the path had their actual values, then it would have been the case that $Y = y'$ in w.

It should be noted this analysis is overly simple[8] for illustrative purposes. As such it won't be able to handle overdetermination cases. To see why this is so recall that holding Billy's throw fixed at 1 and varying Suzy's throw will still result in the window smashing so Suzy is not the cause. Symmetrically, holding Suzy's throw fixed at its actual value and varying Billy's throw will show that Billy is not the cause.

The reason that the simple Structural Equation Model given here can't handle cases of overdetermination is that interventions must fix variables at their *actual* values. But, Fenton-Glynn points out that we could "consider a liberalisation" ((M.S.*a*) p.86) such that "we may sometimes *vary* features of the actual situation" ((M.S.*a*) p.86).[9] Indeed Hitchcock himself considers this idea in (2001*a*). Halpern and Pearl (2011) also consider making this move. If we did this then we could set Billy's throw to the (non-actual) value of 0 and vary Suzy's throw. This recovers the result that Suzy's throw was a cause of the window smashing. Again, a symmetrical argument will mean that the window smashing is caused by Billy's throw. I will discuss how Structural Equation Model approaches can be

[7] Discussing exactly what makes a model apt goes slightly beyond my remit here but for more on this see Halpern and Hitchcock (2011) and Blanchard and Schaffer (2017).

[8] See Fenton-Glynn (2017), Halpern (2016), Halpern & Pearl (2011) and Hitchcock (2001*a*) among others for examples of sophisticated Structural Equation Analyses of causation.

[9] Though this is not a move Fenton-Glynn would himself endorse.

expanded to handle overdetermination in the probabilistic setting in Section 7.3.1 below.

Briefly, another reason to think there is no incompatibility between theories of causation and overdetermining mental causes comes from Campbell (2010) and his concept of a control variable. Control variables are those variables "in terms of which you can parametrize the relation between cause and effect" ((2010) p.21) in a given system. So you will have a cause variable which you can intervene on and an outcome or effect variable. The outcome variable can be monitored to see whether and how it varies under interventions of the cause variable. These can allow you to plot a function from cause to outcome.

To summarise Campbell's argument very briefly, he argues that psychological variables are the most appropriate variables, at the most appropriate level for describing the causal behaviour of human beings. The micro-physical level is too fine grained, including too much redundant information. Campbell notes that someone could argue that it will always be possible to simply gerrymander a physical variable that can act as a control variable for any given phenomenon of interest. He replies to this by saying "it is not a priori that this will be possible, even though everything else supervenes on the physical" ((2010) p.26). He gives economic variables as an example. It is not clear that it's possible to gerrymander together a phsyical variable which can capture notions like interest rates. And he argues, it's even less clear that this is something it would be useful to do if you're interested in studying and controlling interest rates. It would be much more profitable to look for a control variable at a higher level. Likewise, when

studying psychology, it would make more sense to look for control variables at the psychological rather than physical level.

Sider admits there may be legitimate worries about distinguishing epiphenomenalism from genuine cases of overdetermining causation. He gives the example of a singer smashing a glass with the pitch of their voice. The pitch of the singer's voice is causing the smashing unlike the word the singer is saying which is purely epiphenomenal. He points out that there's a difference between this case and its ilk and a case of mental causation. Sider admits that it's hard to pin point exactly what separates the two kinds of case. But, he wants to stress that the search for an answer to this question shouldn't lead to the conclusion that the mental should be causally excluded. Too often, Sider says, the lack of an answer as to why mental states are not epiphenomenal leads to denying mental causation. Rather, "the burden to produce an analysis (showing why mental events are indeed epiphenomenal) is on the objector to mental causation" (Sider (2003) p.722).

Indeed, arguments I have made in Chapter 5 do give us reason to defend mental causation and which supports the idea that the burden is on the objector to explain away these reasons. Sider also mentions that "ordinary belief generates a Moorean pressure to postulate composites" ((2003) p.724) that is, macro-objects (not that every composite will form a macro-object). The MMI I presented in Section 5.1 can be used to make this same point with respect to mental states and mental causation. So, short of an analysis to back it up which can explain why these arguments are wrong, the metaphysical coherence

argument fails to establish the absence of mental causation.

The second objection that Sider considers relies on the concept of coincidence. It's important to note that this only works against *systematic* overdetermination given that it's unobjectionable that there should be occasional coincidences. Sider gives an example with that staple of overdetermination discourse: hitmen. Suppose that every time someone was shot, it was by two shooters overdetermining the death. This wild coincidence would require an explanation, one which couldn't plausibly be given. But the relationship between objects and their parts or mental and physical states are different from the relationships between hitmen. So even though the overdetermination in question would be systematic it does not seem to be coincidental given the close connections at play between an object and its parts or between a mental state and its underlying brain state. Therefore, it should not be surprising, nor objectionable, that both an object and its atoms cause a smashing. In the mental case, there is a nomological relationship between a mental state and its underlying brain state, therefore it is arguably not problematic that each should cause the same event.

The last objection to overdetermination that Sider considers is epistemic, focusing on how we come to know about macro-objects and mental properties or events. The idea is to throw doubt over the existence of the overdetermining entities. To take the macro-object case, if both an object and its constituent parts cause the window smashing, then we may have parsimony reasons to doubt that either the macro-object exits or that the macro-object's atoms exist;

it's "epistemically redundant" ((2003) p.723) to posit both. If we posit that the atomic level exists, then we don't need to posit the macro-object in order to explain why the window smashed. Therefore we can't use the evidence of the window smashing to argue that the macro-object exists.

Sider admits this is a more powerful objection to overdetermination than either that from metaphysical incoherence or coincidence, but even then it's limited in it's power. What it does show is that we cannot rely on simple causal arguments to argue for the existence of macro-entities or mental causation, we need compelling independent reasons. That is, we don't need macro-entities or mental causation to account for certain of our experiences of the world, therefore we cannot use our experiences of the world to argue for their existence. If this were the only evidence we had, or reason to think that these things existed then this would be a compelling argument that overdetermination is troublesome.

In the mental case at least though we have another kind of direct evidence of mental states; introspection. And we can use this kind of experience to argue the case, not only for mental states but also for mental causation, indeed this is at the heart of my argument from the MMI. Sider argues that likewise, few people actually *do* rest their arguments for the existence of macro-entities on the kind of simple causal argument which the epistemic argument has force against. Given that we do have compelling reasons, aside from the simple causal argument, to think marco-objects and mental causation exist, the epistemic argument against overdetermination loses its force. Therefore,

there is no argument left that overdetermination is problematic. Again, as I've outlined in Chapter 5, we do in fact have compelling reasons for positing mental causation.

In conclusion, Sider (and I) reject the first two arguments he considers against overdetermination. Furthermore, the best that can be said for the argument from epistemic doubt is that it means those positing the existence of mental causation or macro-entities need to give more than a simple casual argument for the existence of these phenomenon. It just so happens that we have some such arguments (outlined in Chapter 5 and Chapter 9) in favour of the existence of mental causation and moreover these reasons work in probabilistic worlds, though none rely on the world being probabilistic. First, what I have termed the Mental Manifest Image, secondly, arguments from evolution and lastly the argument that mental states may play natural kind roles in certain scientific theories better than physical ones.

I think these lines of arguments combine to throw significant doubt on the no overdetermination premise of the CEA in both deterministic and probabilistic worlds. With pressure from the positive arguments I mentioned above (such as the MMI or argument from evolution) this could potentially be reason to reject the no overdetermination premise. If the no overdetermination premise doesn't hold, then mental states could be causally efficacious even in deterministic worlds. This is because the CEA would then be unsound because overdetermination does occur but is not problematic and doesn't exclude other, physical, causes. This is to my benefit in two potential ways. Firstly, if you don't agree with

my assumption that the world is probabilistic, this line of argument is still available to me. Secondly, even if you agree with me that the world is probabilistic, you may still not agree with the argument I will go on to make in Chapter 8 that Probabilistic Causal Closure doesn't hold. In that case again this line of argumentation is still open.

7.1.2 Bennett

Another tactic for arguing that overdetermination is benign comes from Bennett. Her (2003) paper `Why The Exclusion Problem Seems Intractable and How, Just Maybe, To Tract It' outlines her Theory of Compatibalism. Her focus is overdetermination, in particular, in showing that there is no (bad) overdetermination involved in mental causation. Her aim is to show that the mental and the physical do not overdetermine their effects or do not do so in a bad way. Bennett is equivocal on the two locutions calling it "just a terminological issue" ((2003) p.474).[10]

She notes that the exclusion problem in and of itself does nothing to say that the mental is not `fit' for causing things. It merely claims that there is nothing left for it to cause once the physical has been taken into account ((2003) p.471). So, Bennett claims, once she has shown that there is no (problematic) overdetermination between the mental and the physical, she does not need to further argue that the mental is capable of being causally efficacious ((2003) p.472).

[10]She repeats this point in footnote 3 of `Exclusion Again' ((2008) p.281)

Given Bennett's focus on overdetermination to defuse the CEA, she herself argues that we need to be clear on what overdetermination is. In order to do this, she puts forward a counterfactual test[11] which gives a necessary, though not sufficient, condition[12] for overdetermination (though she notes that one does not need to hold a counterfactual theory of causation to avail themselves of the test). It runs as follows:

> "*e* is overdetermined by *c1* and *c2* only if
> (O1) if *c1* had happened without *c2*, *e* would still have happened:
> $(c1 \ \& \ \neg c2) \ \Box \rightarrow e$, and
> (O2) if *c2* had happened without *c1*, *e* would still have happened:
> $(c2 \ \& \ \neg c1) \ \Box \rightarrow e$." ((2003) p.476)

or putting the test into terms of a mental state (*m*) and a physical state (*p*):

> "(O1) if *m* had happened without *p*, *e* would still have happened ($m \ \& \ \neg p) \ \Box \rightarrow e$, and
> (O2) if *p* had happened without *m*, *e* would still have happened ($p \ \& \ \neg m) \ \Box \rightarrow e$." ((2003) p.480)

In order for an event to be overdetermined she claims, both these counterfactuals (O1) and (O2) will have to be non-vacuously true. So the compatiblist has two strategies open to her, she can either argue that one or both these

[11] Bennett's test for overdetermination is similar to Kroedel's (2008) definition which I quoted earlier.

[12] Bennett only requires a necessary condition for her compatiblist argument. Therefore she doesn't add the requirement that *c1* and *c2* are causally sufficient for the effect, see ((2003) p.477).

counterfactuals is false or that one or both of these counterfactuals is vacuous.

Bennett argues it's hard to see how the compatiblist could argue that either counterfactual is false without risking the causal efficacy of the mental or the physical. For example, if one claims that (O1) is false then it is not the case that if *m* had happened without *p* then *e* would have happened. In other words, *m* on its own is not sufficient to cause *e*. This calls the phenomenon of mental causation into question. If one claims that (O2) is false then it is not the case that if *p* had happened without *m* then *e* would have happened. Or in other words, *p* in itself would not be sufficient to cause *e*. And this is a denial of closure. So, Bennett concludes, the compatiblist cannot rely on arguing the overdetermination counterfactuals are false without risking denying causal closure or mental causation itself.

Therefore, she reasons, we must shift our attention to the vacuousness of the counterfactuals. If the compatiblist can argue that it's impossible to have the mental event or property without the physical or vice versa, then she can show that one of the counterfactuals is vacuous and that therefore there is no overdetermination. Why does the vacuousness of the counterfactual mean there is no overdetermination? This relies on the idea of the distinctness of the two causes. If it is not possible for one of the causes to happen without the other then it must mean that the mental and the physical causes are not distinct from one another despite not being identical. This diffuses the problem of overdetermination.

Bennett argues that this counterfactual test can show how the `tighter than causal' connection between the mental and the physical could defuse the exclusion problem. "If one of the causes *guarantees* the existence of the other, there is no issue about skipping over some worlds to get to one where the antecedent of the relevant overdetermination counterfactual holds" ((2003) p.479, emphasis in original). In other words "if one of the causes necessitates the other, if it is at least metaphysically impossible for the one to occur without the other, then one of the overdetermination counterfactuals will come out *vacuous*" ((2003) p.479, emphasis in original).

Bennett argues that it is unlikely that the compatiblist will want to claim (O1) is vacuous due to multiple realisation. If there are multiple ways that the mental can be realised by the physical then it is not the case that *m* could not have happened without *p*. It could have happened with *p** instead.

Maybe though the antecedent of (O1) should actually be read as supposing that event *m* (or property M) occurs (or is instantiated) without any relevantly *p*-like event (or relevant P-like property). Perhaps (O1) is vacuous on this reading? Bennett thinks not, because reading (O1) in this way implies that physicalism is necessarily true rather than merely contingently true ((2003) p.484). Bennett points out that although there may be no mental events without associated physical events in this world, there may be in other worlds (see ((2003) p.484)). Therefore, (O1) is not vacuous.

There is more hope for claiming vacuity with (O2) because of the asymmetric

`upwards' necessitation between the physical and the mental. This is the idea "that the physical necessitates the mental, even though the mental does not return the favour" ((2003) p.484). But, Bennett argues, it is not impossible to find scenarios where this might be true. She gives the example of a C-fibre firing occurring in a petri dish. Bennett points out that some functionalists may be able to make such a claim. For some functionalists[13], a mental property is a second order property of having a first-order property which plays a particular role. Take Bennett's example of C-fibres in a petri dish. Such C-fibres could instantiate the property of *C-fibre firing* while not also instantiating the property of *having a pain role property*.[14]

An analogous tactic the compatiblist can take is to argue that "although *p could* occur without *m*, it would no longer cause *e* if it did. That is, although there *are* worlds in which *p* occurs without *m*, they are different enough from the actual world that we have little or no reason to think that *e* would occur here" ((2003) p.487, emphasis in original). It is the case that if *m* had not occurred this entails such a change in circumstances that *p* could still occur but could no longer cause the same things. This is to say that (O2) is actually false.

This does not undercut the causally sufficiency of *p*. All it means is that "the conditions that must hold for *p* to bring about *e* - physical conditions, note - *are*

[13]Ned Block ((2007*b*) footnote 4, (2013)) distinguishes two kinds of functionalism. Taking the example of pain, some functionalists (also known as `realizer functionalists') identify pain with the realizer of the functional role it fulfils (Lewis ((1966), (1999)) is an example of such a theorist) whereas some (also known as `role functionalists') identify pain with the second-order property of having the first-order property which fulfils the functional role. Bennett's point applies to the latter group.

[14]Though, to clarify, I don't wish to endorse a functionalist account.

basically the same as the conditions in which p necessitates m. So if *p were* to occur without *m*, those conditions would not hold - and *p* would not, or at least might not, cause *e*. And that does not mean that *p* does not *actually* cause *e*." ((2003) p.488-489, emphasis in original).

So, there are two strategies you could take to defuse the problem of overdetermination. You could argue that one of the counterfactuals (O1) or (O2) is either false or vacuous, although for now I will focus on (O2). The strategies differ only in "different substitution instances of (O2)" ((2003) p.489). Therefore, the counterfactual in question when making a vacuity argument is subtly different from the counterfactual evaluated when making a falsity claim.

> "Thus the counterfactual claimed to be false - namely, $(p \ \& \ \neg m) \rightarrow e$ - is not the same as the counterfactual claimed to be vacuous - namely,$(p^* \ \& \ \neg m) \rightarrow e$. *p* and *p** are different events (or properties)." ((2003) p.489)

Event (or property) *p* is the kind of physical event (or property) we ordinarily talk about whereas *p** is a more complex property with "a rather extrinsic essence" ((2003) p.489). *p** includes all the physical circumstances (which *p* lacks) required to necessitate *m*. *p** is thus a `stranger' event or property than *p*.

To clarify, lets examine counterfactual (O2) as understood if you were making the falsity argument; if *p* had happened without *m*, *e* would still have happened $(p \ \& \ \neg m) \ \Box\!\!\rightarrow e$. You're trying to show (O2) is false which means you're using a less strict sense of causal sufficiency when speaking about *p*. This means that

you take *p* to be sufficient to bring about *e* in a looser more everyday sense of the term sufficient whereby certain background conditions are left implicit. A counterfactual evaluated in such terms will be false because it will not be the case that *e* would still have occurred if *p* in this sense had occurred without *m*.

On the other hand, say you want to show (O2) is vacuous. Then you evaluate the counterfactual $(p^* \,\&\, \neg m) \rightarrow e$, in which p^* explicitly factors in the proper background conditions and so on which necessitate *m* and which mean that p^* can bring about *e*. The counterfactual thus understood will become vacuously true rather than false. This is because, if p^* is understood in this way, then it cannot occur without *m* and still be the kind of event or property which can bring about *e*.

The real difference then between the falsity and vacuity strategies, Bennett claims, is the "notion of causal sufficiency on which they rely" ((2003) p.489). The vacuity strategy uses a stricter notion under which events (or properties) can only be causally sufficient when background conditions are included, so they are packed into p^*. The falsity strategy on the other hand uses a looser, more 'everyday' sense of causal sufficiency whereby *p* understood as an event or property in the usual way can be considered causally sufficient in itself. So, Bennett argues, the compatiblist has an answer to the exclusion problem whichever sense of causal sufficiency is used.

7.2 Probabilistic Overdetermination

So far I have been looking at overdetermination in deterministic settings. Does the phenomenon occur in probabilistic settings? On the face of it, it may seem less plausible that overdetermination would commonly occur in a probabilistic setting. This is because, in such settings, the set of an event's causes typically don't determine, let alone overdetermine, that event. However, there may be an analogous phenomenon to the overdetermination of events which occurs in probabilistic settings and which could allow the analogue probabilistic CEA to hold. That is, the overdetermination of an event's probability.

In other words, the causes which fix the probability of another event coming about would overdetermine that fixing. Montero and Papineau make this point when they say "there is no reason to doubt a quantum version of the causal closure thesis, to the effect that the *chances* of those effects are fully fixed by prior physical circumstances" ((2016) p.192, emphasis in original). So, the analogue premise from the probabilistic CEA could rule out widespread, systematic overdetermination of probabilities of events, not events themselves.

7.2.1 Example of Overdetermined Probabilities

Fenton-Glynn (2009) gives the following example of the probability of an event being overdetermined:

> "Barbara and Claire are armed with rifles and have a small spatio-

> temporal window for killing Ernst. *If* they want to kill him, they must shoot at t1 and must shoot through the same small aperture (perhaps a chink in his armour). Both want him dead and at t0 each is disposed to fire when the chance comes. Each is an excellent shot and, if she shoots, has a good chance of accuracy (and this chance is independent of whether the other shoots). If either fires alone at t1 and shoots accurately, then her bullet will travel through the aperture, pierce Ernst`s heart and kill him. If, however, both shoot accurately at t1 there is a high chance of a collision between their bullets that will deflect each other off course. Both in fact shoot at t1 and are on target. Their bullets miss each other by a whisker and pierce Ernst's heart simultaneously. Ernst dies. ((2009) p.284)

In this case, because both women are such good shots and have such a high chance of shooting it could be that the probability of one shooting and thereby deflecting the other is so high as to negate the probability that Ernst will in fact be hit by her bullet. So, whether each shoots alone or whether both shoot together does not affect the probability of Ernst's death. However, if their bullets miss each other and both strike Ernst, both bullets will do so at exactly the same time. It is therefore impossible to say which has the greater claim to have caused Ernst's death. This seems to be a case of overdetermination in a probabilistic setting. So it looks as though there are cases where the probability of an event is overdetermined.

Is it that problematic that examples of probabilities of events being overdeter-

mined can be found though? As long as the overdetermination is not systematic it may not be problematic, some coincidences do happen. Questions could (hopefully) be asked about how realistic such a scenario is in real life, however, the example goes to show that overdetermination of probabilities is at least on the face of it possible in a probabilistic world. To recap the probabilistic analogue of the CEA is as follows;

(P1') Probabilistic Causal Closure of Physics
Every physical event has a physical cause which is sufficient to fix its probability.

(P2') No Systematic Overdetermination of Probabilities
It is not usually the case that there are multiple sets of events that are minimally sufficient to fix the probability of a further event which exist simultaneously.

(P3) Non-Identity of the Physical and the Mental
There are physical states and mental states and these are not identical to each other.

(C) There is no Mental Causation
Mental states cannot be causes of physical effects.

If non-identity and causal closure both hold in probabilistic worlds, then (assuming the analogue CEA is valid) it looks as though arguments against the existence or problematic nature of systematic overdetermination in the probabilistic world are required. This is because, if causal closure holds, then every event will have a physical cause sufficient to fix its probability. Therefore, any mental cause which fixes a probability will be an overdetermining one. There-

fore, assuming for now that causal closure does hold (although I will question this assumption in the next chapter), an argument that overdetermination is not problematic in the probabilistic case is required in order to defuse the Analogue CEA. I will now go through the arguments concerning the deterministic setting given above to see if they can also apply to a probabilistic setting.

7.3 Overdetermination in a Probabilistic Setting

Do the defences of overdetermination as being unproblematic given above still hold good of the probabilistic analogue (P2') of the No Overdetermination premise? I'll now reconsider Sider and Bennett's arguments and see if they still hold good. I'll argue that they do and therefore that questioning the No Overdetermination premise of the CEA is one viable way of countering its argument against mental causation with the benefit of applying in either deterministic or probabilistic worlds. This provides a potential 'back up' argument to the other arguments I will provide in the following chapters.

7.3.1 Sider Revisited

To recap, Sider's argument is that overdetermination is not problematic. He puts forward three different possible objections to overdetermination; (i) metaphysical incoherence, (ii) coincidence and (iii)) epistemic doubts.

Given that Sider explicitly mentions probabilistic theories of causation in his metaphysical discussion, that line of argument seems to straightforwardly apply

in probabilistic settings. So what he said about theories of causation not explicitly ruling out overdetermination should hold true whether the world is deterministic or probabilistic. Indeed many sophisticated probabilistic theories of causation (whether in terms of the conditional probability approach or the counterfactual approach to probability raising[15]) allow for overdeterminers to be causes.

Hitchcock, whose Structural Equation Analysis I outlined above (see Section 7.1.1) explicitly states that his approach is limited to the deterministic case ((2007) p.498). This is not a problem though as there are probabilistic Structural Equation Models available, for example that of Luke Fenton-Glynn.[16] It should be noted though that, like Lewis, Fenton-Glynn is supportive of the intuition that overdeterminers individually are not causes, rather they form a cause together ((M.S.*b*) p.23).

I mentioned above that deterministic Structural Equation Analyses can handle overdetermination by allowing certain variables to be held fixed at non-actual values. I will now illustrate how this could work in a probabilistic setting using Fenton-Glynn's overdetermination of probabilities example given in Section 7.2.1. In this case both Barbara and Claire shoot (represented by variables B and C respectively). If either bullet hits Ernst it will kill him (represented by variable E). In order to capture the probabilistic nature of the case we'll need to change the way the variables are written. Fenton-Glynn makes use of notation

[15]I discussed these approaches in Chapter 3.

[16]See Fenton-Glynn (2017). Halpern and Pearl (2011) note that the structural equations in their models are deterministic but that a simple adjustment to their account can allow it to be extended to the probabilistic case.

created by Goldszmidt and Pearl ((1992) p.669-670) to show when a given variable takes a given value due to an intervention, as opposed to another way, and raises the probability of an effect coming about. The example case can therefore be represented as:

B = 1

C = 1

P(E = 1 | *do*(B = 1 and C = 0)) > P(E = 1 | *do*(B = 0 and C = 0))

P(E = 1 | *do*(B = 0 and C = 1)) > P(E = 1 | *do*(B = 0 and C = 0))

The first of the inequalities states that the probability that E takes the value 1 is greater given interventions to set the values of B and C to 1 and 0 respectively than if they were intervened upon to both have a value of 0. The second states that the probability that E takes the value 1 is greater given that B is intervened upon to set its value to 0 and C is intervened upon to set its value to 1 than if both variables B and C have their values set to 0.

It can be seen that if neither had shot then the probability of Ernst's dying would be lower than if one or both of them had shot. Alternatively, Ernst has a higher probability of dying given that at least one shooting occurs. Now, holding the value of B fixed at 0 (that is, not at its actual value) it can been seen that the probability of Ernst's dying has been raised. Thus both Billy's shooting and Claire's shooting can be considered a cause of Ernst's death. This is again an overly simple example about which much more could be said. However, for

space reasons I will leave this topic here.

So it can be shown that plausible theories of probabilistic causation are compatible with overdeterminers counting as causes in cases of overdetermination of probabilities. It also seems right to me that we shouldn't place a restriction on theories of causation that they *must* rule out overdetermination, especially without a compelling independent reason to do so. This compelling independent reason isn't provided by the metaphysical incoherence argument itself, though it may yet be provided by the argument from coincidence or from epistemic doubts.

It seems to me that (ii), the argument from coincidence, can easily be applied to the probabilistic setting. To recap, the argument from coincidence claims that (to apply it to the mental causation case) it is too much of a coincidence that a physical cause and a mental cause systematically occur at the same time. Sider's reply to this objection was that it's no coincidence that they occur together when you consider the tight relation between the mental and the physical. In my opinion Sider's reply to this argument is strong in either a deterministic or probabilistic world. This is because the systematicity which is required in order for overdetermination to be problematic under the CEA will never be un-law-governed whether we're considering overdetermining events or overdetermining the fixing of probabilities. An argument like the argument from coincidence may have force in an utterly random world, if, in such world, there's not even a nomic connection between the mental and the physical. In that case systematic overdetermination would be a miraculous coincidence.

However, I stress again, that my focus is more narrow, and that on a probabilistic picture, there is no reason to think the connection between the mental and the physical cannot be robust enough to counter coincidence.

Lastly, turning to (iii), the epistemic argument, the idea is that we can have epistemic doubt over the existence of some of the entities in question. That is, if both the macro-object and its constituent parts both cause a window smashing, then we may have reasons to doubt the existence of the macro-object. There are parsimony reasons to doubt their existence; they're epistemically redundant and by extension, metaphysically redundant. This is the 'mereological version' of the CEA as applied to macro-objects.

In the probabilistic case though we are talking about a macro-object and its atoms both fixing the probability of an events occurring (other than to probability 1 and 0 as in the deterministic case). It looks as though Sider's arguments still hold. The macro-object and its atoms both fix the probability of the window smashing and Sider's arguments as to why this is unobjectionable still apply. That is, we have independent reasons to think that macro-objects exist.

Likewise, the physical state and the mental state can both fix the probability of an events coming about. For example, a certain set of neuronal firings *and* my thirst both fix the probability of my reaching for a drink. It's true though that we still require some independent reason for positing mental causation. Do we have those in the probabilistic setting? I would argue that we do and those reasons are the same as I cited above that I give in Chapter 5. Briefly, these are

the argument from the MMI (see Section 5.1), the argument from evolution (see Section 5.2) and the argument from the inference to the best explanation (see Section 5.3). What's more, these arguments do not depend on whether the world is probabilistic or deterministic.

7.3.2 Bennett Revisited

Do Bennett's arguments also stack up in a probabilistic world? I think that they do and I will argue why in the final section of this chapter. Recall that Bennett claimed the compatiblist has two options to stave off the threat of (bad) overdetermination. She can either claim that one of the counterfactuals in her test for overdetermination are false or that one is vacuous. To recap, her counterfactual test was:

> "(O1) if m had happened without p, e would still have happened,
> $(m \,\&\, \neg p) \,\Box\!\!\rightarrow e$, and
> (O2) if p had happened without m, e would still have happened,
> $(p \,\&\, \neg m) \,\Box\!\!\rightarrow e$." ((2003) p.480)

If both (O1) and (O2) are non-vacuously true of an event, then that event is overdetermined. For the probabilistic case though we must talk about the probabilities of events being fixed, not the event itself occurring. I have been discussing two ways of interpreting the concept of probability raising; the conditional probability approach and the counterfactual approach. Take first the conditional probability approach? The analogue for Bennett's overdetermina-

tion test would run as follows:

> (O1*) the probability of *e* occurring given *m* had happened without *p* is the same as the probability of *e* given that *m* and *p* had happened, $P(e \mid (m \,\&\, \neg\, p)) = P(e \mid (p \,\&\, m))$, and
> (O2*) the probability of *e* occurring given *p* had happened without *m* is the same as the probability of *e* given that *m* and *p* had happened, $P(e \mid (p \,\&\, \neg\, m)) = P(e \mid (p \,\&\, m))$

Would $P(e \mid (p \,\&\, \neg\, m)) = P(e \mid (p \,\&\, m))$? Would the probability of *e* occurring given that the physical and mental state occur be equal to the probability given the physical state had occurred without the mental state? Given the tight relation between *m* and *p*, which is not effected by being in a probabilistic setting (whichever interpretation of probability is used) Bennett's argument still holds, indeed the deterministic scenario would be a special case. I will briefly run through the falsity and vacuity strategies put forward by Bennett to clarify my point.

If you take the falsity line then the physical state *p* can't occur without *m* (and still be a *p* which could raise the probability of *e*). So, the probability of *e* occurring if physical state *p* occurred without mental state *m* is lower than if they both occurred and thus the test comes out false and there is no (bad) overdetermination. Whereas, if you take the vacuity line, then the physical state in question *p** (which packs in more extrinsic information than physical state *p*) will just be a state which necessitates *m*. Therefore, physical state *p**

cannot occur without *m*. So, since conditioning upon an event with probability 0 is undefined, the conditional probability of P(*e* | *p* & ¬*m*) is undefined. Once again then, the test fails and there is no (bad) overdetermination.

What about the counterfactual approach to probability raising? So, the test in a probabilistic setting should be (where the probability of *e* occurring if *m* happens (M = 1) and *p* happens (P = 1) is P(E = 1)):

> (O1') if *m* had happened without *p*, P(E) would still have been the same,
> P(E = 1 | *do*(M = 1 and P = 0)) = P(E = 1 | *do*(M = 1 and P = 1)) and
> (O2') if *p* had happened without *m*, P(E) would still have been the same,
> P(E = 1 | *do*(M = 0 and P = 1)) = P(E = 1 | *do*(M = 1 and P = 1))

It can be seen that Bennett's arguments for the vacuity or falsity of the original test apply also to this adapted test. (O2') can still be viewed as either false or vacuous. Given that the relation between the mental and the physical remains sufficiently tight in a probabilistic setting, Bennett's argument should still go through. Parallel reasoning to that used in the conditional probability approach applies here. The tight relationship between the mental and the physical still means that if *p* were to occur without *m* then it would not be a *p* which could fix the probability of *e* to be the same as it would have been if *m* had also occurred.

Therefore, I think Bennett's arguments against (bad) overdetermination still apply

in the probabilistic setting. Furthermore, if you're not convinced by either of my following arguments against the truth of causal closure, then this avenue to arguing against the CEA is still open. I will now move onto the first premise of the CEA, the causal closure of physics, and put forward my argument for how to dissolve the CEA.

8

Causal Closure of Physics and the CEA

"I suppose I'll have to add the force of gravity to my list of enemies"

The Penultimate Peril - Daniel Handler (2005)

I will now examine the premise of the CEA which I believe is the key to dissolving the CEA. If what I have to say about the Probabilistic Causal Closure premise is compelling then, no matter how convincing you find arguments against No Overdetermination of Probabilities premise, the Probabilistic CEA analogue is

unsound. Therefore, the door is opened to mental causation of physical effects.

8.1 Causal Closure of Physics

I have argued in Chapter 6 that the mental and the physical are not identical. And, although I have put forward some arguments as to why overdetermination is not problematic in Chapter 7, I will assume for now that there should not be any systematic overdetermination of probabilities. Even though I will grant the no overdetermination premise for now, I think we still have reasons to doubt the causal closure of physics and therefore the soundness of the CEA.

If physics were not closed (and the non-identity of the mental and the physical holds) then it would be uncontroversial that non-physical causes could bring about physical effects without causing widespread and systematic overdetermination. There needn't be any overdetermination to object to and the CEA would be unsound. So the question boils down to, does causal closure hold in a probabilistic world?

To recap, the original premise is (P1) Causal closure of Physics which states that every physical event has a sufficient physical cause. The Probabilistic analogue (P1') Probabilistic Causal Closure of the Physics states that every physical event has a physical cause which is sufficient to fix its probability. Does combining this with the Non-Identity premise and the No Overdetermination of Probabilities premise (it is not systematically the case that there are multiple sets of events that are minimally sufficient to fix the probability of a further event which exist

simultaneously) mean there is no mental causation even in probabilistic worlds?

8.1.1 What is Causal Closure?

In a sentence, the causal closure of physics (or of the physical) states that any physical event which has a cause, has a sufficient physical cause. The probabilistic version states that every physical event which has a cause has a physical cause which is sufficient to fix its probability. However perhaps more than any other premise of the CEA the devil is in the detail, particularly when it comes to what you mean by 'physics' and 'physical'. Furthermore, care must be taken over exactly how to interpret this sentence as it can be read more strongly or more weakly (I will discuss this further in the next section).

To be as charitable as possible to the defender of the CEA, I will take 'physics' to be anything under the purview of the physical sciences in the broadest possible sense. In other words I will take to be 'physical' not only all those things such as quarks, forces and fields which are posited by our best current theories of quantum mechanics and general relativity but also all those things posited by our other best current theories of other natural sciences such as those of biology and chemistry.

Of course, these theories are always open to revision, indeed we expect that our theories will change over time. So our ontologies will have to be updated with the introduction of new and improved theories. For reductionists, the special sciences ultimately reduce to physics and would thus be included that way.

However, I don't need to take a stance on reductionism since including the higher level sciences won't trivialise the principle.

That said, including psychology into the family of higher-level sciences would trivialise the principle. Therefore, I won't be including psychological sciences within the purview of 'physics'. By extension, I will also not be including any sciences which include psychological concepts such as sociology or economics which makes recourse to the ideal rational agent.

Why is it so vital to pin down exactly what is meant by 'physics'? The reason is to avoid triviality when applying causal closure to the CEA. If 'physics' were to be interpreted too broadly then potentially any phenomenon could count as falling under it and therefore could count as a cause. Causal closure would become trivial under these circumstances. In that case, it would be a simple matter to stipulate a physicalist world view whilst also maintaining the existence of mental causation. The premise would not allow the CEA to go through as the mental's not being an overdetermining cause would be compatible with completeness. Therefore, to stay true to the spirit of the world view the principle is supposed to represent we must restrict what we mean by physical.

This restriction cuts both ways however. It is important that, though we restrict what we mean by physics that this limits the phenomenon which causal closure covers, namely physical effects. If, on the other hand the phrasing was "any effect has a sufficient physical cause" this would be an overly stringent principle. Firstly, depending on how you understand the term "physical", this wording may

rule out causation in higher-level sciences. This would go beyond the available evidence precisely because much of the evidence we have for the causal closure principle comes from generalisations from the special sciences. That being said, if you have a less narrow understanding of the term "physics", such as the one I will be using, then this is less problematic.

Secondly, we need to word the causal closure principle carefully because the CEA is compatible with mental causation even if only in the case of a mental cause bringing about a mental effect.[1] This is important because it means mental causation is not completely written off.

Before moving on to give arguments for and against holding causal closure, I need to take a brief sidestep to discuss the related notion of the completeness of physics.

8.1.2 Causal Closure and Completeness of Physics

What is the relationship between the principle of causal closure and the thesis of the completeness of physics? The two terms are sometimes used interchangeably. Others consider causal closure to be the stronger formulation of completeness, see for example Marcus ((2005) pp.28-29) or Montero (2003). In such cases the completeness thesis states that we don't need to look beyond physics in order to find sufficient causes for physical effects. I will call this *weak*

[1] It also remains silent on the issue of a physical cause bringing about a mental effect. Supervenience is not causation so that's not problematic, but we do think that physical causes can have mental effects.

closure;

> *Weak Closure*. Any physical event has a sufficient physical cause, although this does not rule out the possibility of it also having a non-physical cause.

Contrast this formulation to causal closure in the stronger sense which claims that there can be no sufficient non-physical causes to physical effects. Kim (2008) also mentions this issue. He considers adopting the principle he calls;

> "*Strong Closure*. Any cause of a physical event is itself a physical event - that is, no nonphysical event can be a cause of a physical event" ((2008) p.50)

Kim rejects the use of the stronger principle in favour of the weaker principle that doesn't explicitly rule out the possibility of non-physical causation of physical events. This is because he doesn't want to beg the question against the possibility of mental causation ((2008) pp.51-52).

So, the weaker version of causal closure, which is equivalent to the completeness of physics, is used in the CEA. That is, the causal closure premise used in the CEA states that we don't need to look beyond physics in order to give a full causal history of the world. To use the stronger formulation would be to beg the question of the CEA. This is therefore the sense I will use from now on and I will be using 'causal closure' and 'completeness' interchangeably to describe this weaker principle. Furthermore, I will be using physics in the broad sense of the term discussed above.

8.1.3 The History of the Thesis of the Completeness of Physics

In the appendix to *Thinking About Consciousness* ((2002) pp.232-256) Papineau gives a brief history of the thesis of completeness of physics which I shall now even more briefly summarise. I will refer back to this in Section 8.1.5 when I discuss physicalist bias.

Papineau gives Leibniz as the first example of someone who's conservation laws were such as to give completeness of physics. This is because his conservation of linear momentum and kinetic energy together (plus the assumption of no action at a distance) are enough to close physics to any kind of mental `interference'. Newtonian physics took a different tack to Leibnizian physics in taking neither contact nor impact as his basic notion, but rather `impressed force'. Such impressed forces are much more permissive in their origins, thus opening up the possibility that mental forces could be among them. This is because, at least initially, while Newtonian physics, like Leibnizian physics conserved momentum, it did not conserve energy. Physics does not look so complete or closed anymore. This is because it allows energy to enter into the physical world perhaps due to mental forces.

Latterly however, the conservation of energy did come to be considered a basic physical tenet. Experiments done by scientists such as James Joule led people to think that something was in fact conserved in some physical processes. For example, Joule's experiments dealt with heat and mechanical energy, which he found to be equivalent. Such work in fact led to the creation of the universal

theory of the conservation of energy. Herman von Helmholtz was the one to bring all the loose ends together. Luckily for history (if you will) Helmholtz had a reductionist project of his own, attempting to reduce biological phenomena to underlying non-biological laws. The pursuit of this project led him to make the assertion that energy must be conserved by all forces, even those, such as friction, which had traditionally not been considered conserved.

From this point on in history conservation was taken as given. One of the questions this raised was what implications this had for the completeness of physics. Papineau cites what he calls the "argument from fundamental forces" which is the argument that "all apparently special forces characteristically *reduce* to a small stock of basic physical forces which conserve energy" ((2002) p.250, emphasis in original). He credits this line of thought with leading scientists such as Helmholz to hold their view that there were no animate forces meaning that conservation applied to only physical forces. Advances in the 1950s into biochemical and neurophysiological forces made it more and more difficult to argue for extra-physical, animate forces. This addition of empirical evidence to the position of the conservation of energy left little room for those who did not want to hold the completeness of physics.

8.1.4 Arguments For Causal Closure

There is no `knock-down' argument in favour of holding causal closure. Advances such as those described in the previous section can be used as evidence that it holds. However, perhaps the strongest argument which can be made for

causal closure comes from physicalism and our general current scientific world view.

Physicalism

Perhaps the most obvious and compelling reason to think the causal closure must hold is that it so naturally fits with our current scientific and wider philosophical world view; namely physicalism and adherence to scientific practice. Science, and in particular, the natural sciences, has had a hugely successful track record. Take for example the massive advances in medical sciences[2] or the achievement that was unifying electromagnetic theory.[3] The progress of science is unparalleled which can lead thinkers to place all their eggs in its basket.

The argument goes, because physics (in the broad sense of the term) has operated well without reference to non-physical causation, that we should extrapolate from past experience to the logical conclusion that there is no non-physical causation. It is essentially an inductive argument from our best scientific experience. How strong is this as a defence of causal closure though?

One note should be made here however. Usually, rather than physicalism being used as a reason for holding causal closure, the opposite argument is made. That is, that causal closure and the success of physics are used as reasons to be

[2]Perhaps interestingly these advances, in my own anecdotal experience, seem to have occurred more in physical heath than mental health.

[3]See for example Maxwell (1863) as one step in this journey.

physicalist.[4] However, I believe as far as both physicalism and causal closure rest on potentially biased foundations, they stand or fall together. Therefore, I argue both are vulnerable in the same way in that both views could potentially stand a little scrutiny as I shall go on to argue in the next section.

8.1.5 Arguments Against Causal Closure

As I've discussed there is not any fully convincing argument in favour of physicalism and causal closure. Now I will discuss two reasons for thinking causal closure may not hold: physicalist bias and causal closure as a mere typicality condition.

Physicalist Bias

Jones (2008) refers to causal closure as a "sort of `philosophical glue' that binds a theory together" ((2008) p.181) rather than a straightforward summary of physical, scientific observations. By this he means that philosophers and scientists use causal closure as a kind of heuristic (although this is not a word he uses) with which to build their theories. There is no *direct* observation of causal closure, rather it is an inductive conclusion we have come to from our physical observations to date. Additionally, an inference to the best explanation can be made; treating the world as though it is causally closed has yielded promising results so we should continue to do so. Furthermore Vicente (2006)[5] says:

[4]See for example Lewis who calls causal closure "the empirical foundation on which materialism builds" ((1966) p.23). See also Papineau (2002) among others for another example.

[5]Vicente's paper is a defence of causal closure and therefore should put forward the best case for it holding.

> "However, it (causal closure) is not a law that appears in physics textbooks. Where does it come from? Two answers spring to mind. First is can be said that it is not a physical law, but rather a methodological norm or principle that guides physicists in their research. Moreover, it can be defended that it is a norm well supported by inductive evidence... Second it may be said, although causal closure P (causal closure of physics) is not strictly a truth of physics, it is supported by, or depends on, actual laws of physics." ((2006) pp.150-151)

If the best that can be said for causal closure is that it coheres with a wider world view, or is a useful heuristic norm, then it might not be on the sturdiest ground. As genealogist theorists such as Michael Foucault[6] have argued for decades, if you can track the history of an idea and find it originates in bias then that theory, at best, should be thoroughly examined. While I would not go so far as some genealogist thinkers in saying because we can trace the origin of the idea to human bias we should abandon the theory altogether, I think it definitely shows that more argumentation needs to be put forward as to why we should hold to this principle.

As I have shown through my discussion in section 8.1.3 of Papineau's appendix which traces the history of the completeness of physics, this principle can indeed be traced back to biases, contingencies and what may be colloquially termed as `physics envy' construed in the broad sense of physics. Take Helmholtz as an

[6]See *The History of Madness* (Foucault & Khalfa (2006)), *The Birth of the Clinic* (Foucault (2010)) and *Discipline and Punish* (Foucault (1991)) to name but a few examples.

example. Papineau notes that Helmholz's "physiological context undoubtedly played a fundamental role in Helmholz's articulation of a universal principle of the conservation of energy" ((2002) p.246). Further Papineau says it's "likely that it was Helmholtz's specific combination of physiological interests and sophisticated physical understanding that precipitated his crucial synthesis of the different strands of research feeding into the conservation of energy" ((2002) p.247). All this is to make just one example of how historical coincidence can lead the course of intellectual history in a particular direction. This case may be considered more of a contingency than a bias, given it was Helmholtz's time and place which led to his interests, but nevertheless this could be considered troubling if you want to examine the reasons for holding a principle. Had Helmholtz had a different particular history then the course of the thesis of causal closure may have been derailed and may not have taken its predominant spot in our philosophical world view. When it comes to biases, the case is all the worse. Of course, this line of reasoning is speculative, but it does lead me to question the extent to which contingencies lead to the philosophical positions, causal closure in particular, that we hold dear.

At the very least, raising awareness of these contingencies leads me to think we must be very careful in examining why we hold the views that we do to make sure we don't place more faith in them than the evidence would allow. In Helmholtz's case Papineau himself asks "how far was this almost immediate agreement on the conservation of energy dictated by the strength of evidence rather than by intellectual fashion" ((2002) p.250). In Helmholz's case the ev-

idence was strong, is this the case for causal closure also? Ultimately every philosophical position relies on intuition and assumptions. Uncovering these biases to see why and how they could be affecting our views leading us to accept some conclusions over others can surely never be wasted work.

So far the argument has been all negative, reasons to knock our belief in causal closure. Now I will offer a positive argument for thinking causal closure doesn't hold in the form of Bishop's argument from Causal Closure as a Typicality Condition.

Causal Closure as a mere Typicality Condition

Bishop (2006) argues that causal closure can at most be considered a typicality condition. By this he means "that in the absence of *non*-physical influences, physical events will proceed typically" ((2006) p.46) by following fundamental physical laws. In order to make his point, Bishop gives the example of Newton's first and second laws. Newton's first law of motion is itself a typicality condition in that it states how a free body will behave as long as no external force acts upon it. But Newton's first law can't specify what kind of forces exist. As for Newton's second law, $F = ma$, a force F is sufficient to cause an apple to fall from a tree, at least until someone sticks their hand out and catches the apple. The behaviour of the apple has been changed without violating a law or any overdetermination occurring. You could try to account for this by expanding the forces encompassed by your calculations but Bishop points out this won't help. This is because nothing in the second law itself can tell you which forces

do and don't exist and therefore if you have included them all. An extra premise is required stating that these are all and the only forces that exist. He considers that all physical and other special science laws hold in a qualified and idealised manner. Although causal closure is a metaphysical principle rather than a law of physics, Bishop argues that likewise we shouldn't expect causal closure to strictly hold. Rather it holds in usual conditions but it is beyond the scope of causal closure to specify exactly what these conditions are. In other words, the principle of causal closure should be considered as a typicality or ceteris paribus generalisation.

If causal closure is merely a typicality condition, then both the deterministic version of the CEA and the probabilistic analogue CEA are liable to be unsound. This is because (P1') the Probabilistic Causal Closure premise of the CEA is not stated as a typicality condition. When a mental event influences causation (or the fixing of probabilities) then it is plausible that typical circumstances don't hold and therefore causal closure doesn't apply. And if causal closure doesn't hold then the CEA isn't sound. This line of argument holds whether or not the world (and therefore the CEA) is deterministic or probabilistic.[7]

To summarise, there need not be any overdetermination by mental causes as these can "modify or co-opt" (Bishop (2006) p.48) the typical conditions. If Bishop is right and causal closure is at most a typicality condition then the CEA would be unsound unless it was updated to make the causal closure premise

[7] Interestingly Bishop also claims his argument holds independently of how broadly or narrowly you define physics (Bishop (2006) p.45).

hold ceteris paribus. But in that case the CEA would be invalid as it would no longer rule out mental causation of physical effects. To make it valid again the hidden premise that *only* physical events can bring about physical events would have to be added. But that renders the whole CEA question begging.

Now, I will move on to evaluating the probabilistic version of the CEA in the context of our assumed probabilistic setting.

8.2 Evaluating the Probabilistic CEA

To recap the probabilistic analogue of the CEA I have set it out again here:

(P1') Probabilistic Causal Closure of Physics
Every physical event has a physical cause which is sufficient to fix its probability.

(P2') No Systematic Overdetermination of Probabilities
It is not systematically the case that there are multiple sets of events that are minimally sufficient to fix the probability of a further event which exist simultaneously.

(P3) Non-Identity of the Physical and the Mental
There are physical states and mental states and these are not identical to each other.

(C) There is no Mental Causation
Mental states cannot be causes of physical effects.

The main question for the remainder of this chapter will be if this probabilistic

analogue holds in our probabilistic world. If the Probabilistic Causal Closure premise doesn't hold then the CEA will be unsound. In the previous sections of this chapter I have already put forward some reasons to think that causal closure may not hold independently of whether the world is deterministic or probabilistic. However, my argument will be that we have more reason to think that Probabilistic Causal Closure doesn't hold *because* the world is probabilistic than we would have in a deterministic world.

8.2.1 Causal Closure and Indeterminism

The original version of the CEA may still hold in deterministic worlds. However, whether or not the original Causal Closure premise holds in deterministic worlds, we've got good reasons for thinking that the premises of the original version of the CEA are not true in our probabilistic world. This is specifically because the original Causal Closure premise doesn't hold in probabilistic worlds. Therefore, I turned to analysing the probabilistic analogue version of the CEA, whose premises we have better reason to think are true in our world.

I think my argument against causal closure is much stronger in probabilistic worlds. If I am right and even the analogue probabilistic Causal Closure premise is false in probabilistic worlds like ours, then the analogue CEA will be unsound. This is because if the thesis of causal closure if false, then physical causes do not always guarantee their effects, nor fix their probabilities and there may be room for the mental to be doing some work.

However, it is possible that even in probabilistic worlds the mental doesn't do any causal work. It could be the case that physical and *only* physical causes serve to fix the probability distributions for further physical effects. Any fixing done by mental causes could be viewed as overdetermining in such a scenario. But the picture is complicated. There are different places along the causal chain where mental causation could enter. Examine Figure 8.1.

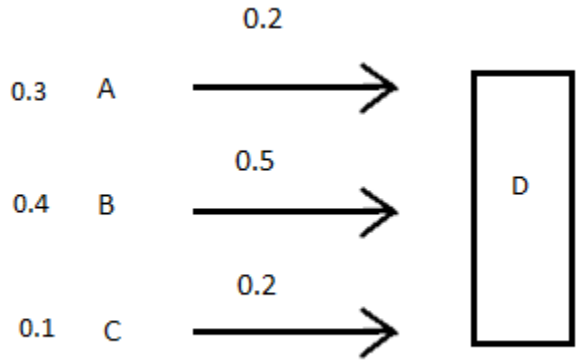

Figure 8.1

Here there are three potential causes of event D[8] represented by the rectangle labelled D which is can be brought about by either A, B or C. It is not certain that if any of the physical causes were to come about that they would cause D. A has a 0.3 chance of occurring and a there is a 0.2 chance of D occurring given A (P(A) = 0.3 and P(D | A)=0.2). B has a 0.4 chance of occurring and D has a 0.5 chance of occurring given B (P(B) = 0.4 and P(D | B)=0.5). C has only a small 0.1 chance of occurring and D has a 0.2 chance of occurring given C

[8]Or rather the causes of a kind of event represented by D for simplicity.

(P(C) = 0.1 and P(D | C)=0.2).

In Figure 8.1 there is not sufficiency of the physical cause. If the world is not running strictly according to deterministic rules then it looks as though the fixing of the probabilities of events could potentially be underdetermined by their physical causes; in other words, causal closure does not hold. It is not the case that any cause A, B or C will definitely cause D. This is an importantly different situation than in the deterministic case. In the deterministic case, once the probabilities are fixed by the physical causes, there is no non-overdetermining work left for the mental to do. If the probability of an event, given a physical cause is 1 then adding mental causes to the story cannot raise the probability of the event any further.

In the probabilistic world though, this is not the case. The door is ajar to mental causes 'topping up' the probabilities fixed by the physical causes alone. In the deterministic case there seems to be no wiggle room, whereas in the probabilistic case there is a route by which the mental can be doing non-overdetermining work.

Take event A. There's only a 0.3 chance that physical cause A will occur and, given A occurs, D then has a 0.2 chance of occurring (P(D | A)=0.2). Say event A does come to pass, then it seems as though there is room for mental state *e* to contribute to raising the probability. In other words, it could be that some physical cause confers a probability onto an effect but either also requires a mental cause to fully fix the probability, or at least allows room for mental

states to effect the probability of the event occurring. This would mean that the probability of D occurring is not actually fixed purely by the physical cause A which means that (P1') the Probabilistic Causal Closure premise does not hold. It is also the case that the mental state is not an overdetermining cause nor an overdeterminer of the probability of the physical effect. Contrast this with the deterministic case in which all physical events have probability 1 conditional on their physical causes so there is no room for mental states to 'top up' the probability.

Alternatively, it is also possible that a mental event could lower the probability of a physical event. Whether or not probability lowering counts as "causation" per se, it is clear that it has causal influence in that it can prevent an event occurring or change the manner in which the event occurs.[9]

However you philosophically frame physical probability lowering "causes"[10], an analogous story can be told for mental events. John Dupré gives an example of a hypothetical gene which brings about a disposition both to smoke and to exercise regularly ((1984) p.170). Or, say you don't have the hypothetical gene, perhaps the desire to lose weight could function as a mental event which could give rise to both these behaviours. Roughly speaking, smoking raises your chance of a heart attack while exercising lowers it. The overall effect of the hypothetical gene is therefore unclear. If the smoking and exercising behaviours

[9]Though, depending on how finely you individuate events, changing the manner in which the event occurs may actually amount to bringing about a different event.

[10]See Salmon (1998) for an example of someone who discusses this issue. For more on this topic see Chapter 3 where I discuss probability raising theories of causation.

were instead motivated by a desire to lose weight, then it is possible that this mental event could raise or lower the probability of a heart attack event.

Furthermore, a mental event might change the manner in which a physical event occurs. Take exercising as an example. Say I have a dislike of exercise which is not quite high enough to prevent me from doing it altogether. The way I go about exercising on a typical day may differ from how I would exercise on a day where I was feeling uncharacteristically motivated and fully wanted to. For example, on a usual day I may put in less effort thereby not raising my heart rate as much or I may leave the gym sooner. Of course, the converse is also true for those days where I am feeling more motivated.

8.2.2 Disanalogy between the Deterministic and Probabilistic CEA

If mental states help to fix the probability of the physical event occurring then are they in fact causally efficacious. For example, the mental state of desiring a coffee may raise the probability of my getting a coffee over and above the mere physical neural firings. Or, it may help shape the exact manner in which I go about getting the coffee, for example by raising the probability that I rush to the coffee shop as opposed to walking slowly. Actual me may have a different probability distribution over collecting coffee in any particular way to 'zombie me', as the latter has only physical causes to fix the probabilities.

Another way of understanding the question here is do we have more reason to think that (P1') the Probabilistic Causal Closure premise doesn't hold in a

probabilistic world than to think the deterministic version (P1) doesn't hold in a deterministic world? Is there any disanalogy between the deterministic and probabilistic worlds which would allow us to rule out this possibility?

Take again the suggestion that mental states could influence the manner in which an event is carried out. So, if I'm thirsty, this will cause me to rush to the coffee shop faster than I would have otherwise done. This may be the case in the deterministic world because the nearest world in which I'm not thirsty is one in which I do not have the corresponding `thirsty' brain state.

However, I claim that in the deterministic world the mental state itself may not be making any overdetermining causal difference because my coffee shop visit is already sufficiently caused by the physical causes. This is because if I did have the thirsty underlying brain state then it doesn't matter if I feel the thirst or if I'm a philosophical zombie, it won't affect the way I go to the coffee shop as the event is determined by the physical alone. To put this into terms of inequalities, say T represents having a `thirsty' mental state, B represents having a corresponding `thirsty' brain state and C represents going to the coffee shop. In the deterministic world $P(C \mid B) = 1$. There's just no room for anything to increase the probability above 1. So even though $P(C \mid B \& T) = 1$ it seems that T is making no causal difference. And this will be true for any other mental state. This argument can't be made in probabilistic worlds where $P(C \mid B) < 1$.

Before moving onto the topic of what models of mental causation could look like in a probabilistic setting, I want to press the point that we don't always have

to favour physical causes over mental ones even when they are not acting as joint causes.[11] Just as it is possible for the physical to screen off mental causes, so it is possible for the mental to potentially screen off physical ones. For example, my desire for coffee could screen off the underlying brain state associated with it. The brain state is screened off from causing my getting coffee once the desire fixed the probability of my acting. Once the mental cause has fixed the probability of the event, the physical cause cannot change it. In that case there would be no room for the physical to 'top up' the probability. To put this in formal terms, just as it's possible that P(E | P & M) = P(E | P) can hold, so P(E | P & M) = P(E | M) can hold. That is, the first equality says that the probability of an event given a physical and a mental event holding can be equal to the probability of that event given just the physical cause. Likewise, it is possible that the probability of any given event occurring given a physical and mental event occurring can equal the probability of that event given just the mental cause. In summary, there's no reason why mental and physical causes shouldn't have this symmetrical relationship to causing. It is just our biases, particularly that of the principle of causal closure which leads us to favour physical causes to mental ones.

To reiterate; typically in probabilistic worlds, the probability of a physical effect given its physical cause will be less than 1. This leaves room for the mental to

[11] For example, Yablo (1992) puts forward a similar type of argument in "Mental Causation". He claims that mental causes may be more proportionate than their corresponding physical brain states and when this is the case we should treat the mental rather than the physical as the cause. I discuss Yablo's arguments in detail in section 9.2.3. I also make a similar style argument based on the naturalness of mental and physical states in Chapter 9.

'top up' the probability (or indeed lower it). It is possible that an event may have a higher probability of occurring given the mental state and physical state than given just the physical state alone. However, in deterministic worlds, the probability of a physical effect given its putative physical cause will typically be 1. There is now no room for the mental to do any non-overdetermining work because the probability cannot be greater than 1.

8.2.3 How Models of Mental Causation can Work in Probabilistic Settings

So, if it is the case that mental states can contribute to fixing the probability of a physical state (over and above the contribution made by the brain states) thereby causing a physical effect, then it looks as though we have a way in which mental causation can exist contra the causal exclusion argument. This raises the question of how models of mental causation would work in a probabilistic setting? Let's examine Figure 8.1 again:

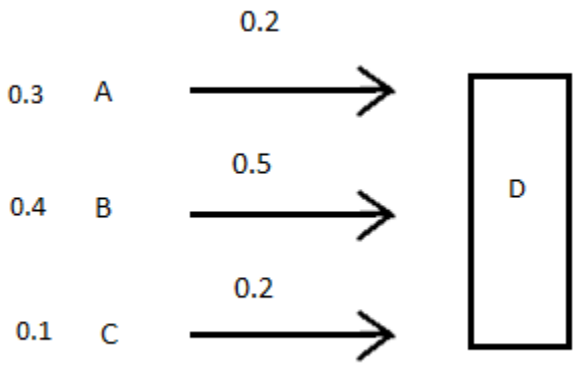

Figure 8.1

The diagram indicates that there is no one sufficient physical cause of the event D's occurring. Let's say that the event occurs and is caused by physical event A. If causation was deterministic then the probability of D occurring given A would be 1. But in this probabilistic setting A has a 0.3 chance of occurring and, given it occurs, D now has a 0.2 chance of occurring. Perhaps if we add the mental into the picture we can see how it can be causally efficacious by raising the probability that the event will occur. I will now present my argument in the form of inequalities to make my point clearer. Call the event E and the physical cause A. We can demonstrate that physical cause A is a cause by showing that the following holds[12]:

P(E | A) > P(E | ¬A)

All this inequality says is that the probability of the event occurring is higher given the physical cause than the absence of the physical cause. If this inequality holds then we have reason to think that A is what Suppes' called a "prima facie cause" ((1970) p.12) of B.[13]

[12]It's important to note here that I will use a simplified version of probability raising accounts of causation. This is just to make a demonstration of how my point can be made and I hope that the key points will also apply to more sophisticated and nuanced probability raising accounts.

[13]A full discussion of Suppes' definition would take us too far afield but briefly it is as follows:

> "*The event $B_{t'}$ is a prima facie cause of the event A_t if and only if:*
> *(i) t' < t,*
> *(ii) P($B_{t'}$) > 0,*
> *(iii) P(A_t | $B_{t'}$) > P(A_t).*" ((1970) p.12)

That is event $B_{t'}$ is a prima facie cause of event A_t if and only if three conditions are met. The first is that $B_{t'}$ occurs before A_t. The second is $B_{t'}$ has a non-zero probability of occurring. And the third is that the probability of event A_t occurring is higher given that $B_{t'}$ occurs than the

If you prefer the counterfactual approach to probability raising then the situation can be represented in Goldszmidt and Pearl's (1992) notation:

P(E = 1 | *do*(A = 1)) > P(E = 1 | *do*(A = 0))

This inequality states that the probability that E would have happened if A had happened is higher than the probability that E would have happened if A hadn't happened due to interventions. Now call the mental state M. If the following inequality holds then this would be a sign of the causal efficacy of mental state M. If the inequality holds when we hold A fixed, then this implies that it is mental state M that is making the difference.

P(E | A & M) > P(E | A & ¬M)

In prose, the probability of the event occurring given the presence of the physical and mental state is higher than the probability of the event occurring given only the physical state.

This is a stronger test than the physical case mentioned above which only gave us reason to think that A was a prima facie cause of E. However, to allude to Suppes' idea of "spurious cause"[14] ((1970) pp.21-28), if it transpired that P(E | A &

unconditional probability of A_t.

[14]Suppes gives the following as preliminary definition of a spurious cause:

> "*Let $B_{t'}$ be a prima facie cause of A_t. Then $B_{t'}$ is a spurious cause of A_t if and only if there is a t" < t' and an event $C_{t''}$ such that P($B_{t'}C_{t''}$) > 0 and P(A_t | $B_{t'}C_{t''}$) = P(A_t | $C_{t''}$)*" ((1970) p.21)

The idea is that although $B_{t'}$ may be a prima facie cause of A_t, there may be an earlier event $C_{t''}$ which can show $B_{t'}$ to not actually be playing a causal role, thus making it a spurious cause. It is important to note that this is just Suppes' preliminary definition to capture the intuitive idea. However, this is sufficient to make my point so for space reasons I will not discuss Suppes' full definition.

M) = P(E | ¬A & M) then we would have reason to think that A is not a cause of E. That is, if the probability of event E occurring given that physical event A and mental event M both occur is the same as the probability of event E occurring if only mental event M occurred, we would have reason to think that it was M and not A that was making the causal difference.[15] The same reasoning can be applied if a mental state is a prima facie cause.

Alternatively, on the Golszmidt and Pearl notation the situation can be represented as:

P(E = 1 | *do*(A = 1 & M = 1)) > P(E = 1 | *do*(A = 1 & M = 0))

This inequality states that the probability that E would have happened if physical event A and mental event M had both happened is greater than the probability that E would have occurred if A had happened without M due to interventions.

I believe the same point could be made of other special science properties which would mean that special science higher level properties would count as causes of physical events.[16]

[15]This is merely an allusion to Suppe's idea because the relation between the mental and physical in my example is not the same relation as between $B_{t'}$and $C_{t''}$ in Suppes'. For one thing, $C_{t''}$ is explicitly an earlier event than $B_{t'}$ whereas this will not be true of a mental state supervening on a physical state.

[16]I will discuss this idea further in Section 10.1.

8.3 The Physical Without the Mental?

There is a question that needs to be answered now however. Does the physical state necessitate the mental? That is, can physical event A occur without its associated mental event M? If it is the case that the physical event cannot occur without its associated mental one (because the physical necessitates the mental in that case) then the conditional probability P(E | A & ¬M) is not defined. This is because conditioning upon an event with probability 0 is undefined. This is problematic for me because I cannot argue that the inequality P(E | A & M) > P(E | A & ¬M) holds if the latter half of the inequality is undefined.

The counterfactual approach to probability raising suffers similarly because the counterfactual "if *p* had happened without *m*, *e* would still have happened" ((*p* & ¬*m*) $\Box \rightarrow$ *e*) becomes vacuous. This is because it would not be possible for the physical state to occur without the associated mental state. For the purposes of the rest of this section I will take the counterfactual approach to probability raising.[17]

I discussed this issue in the previous chapter (Section 7.1.2) when I discussed Bennett's (2003) solution to the exclusion problem. Bennett argues that the physical event couldn't have occurred without its associated mental event and still been able to cause its effect. Recall this renders one of her test counterfactuals[18], which states that if the physical had occurred without the

[17] Although I believe my point could also be expressed in terms of the conditional probability approach.

[18] To quickly recap those counterfactuals, they are:

mental then the effect would still have happened, either false or vacuous depending on how exactly you read the counterfactual.

I want to argue that the necessitation of the mental by the physical is not a problem for my argument. To that end I will now discuss Woodward's account of causation and explain how this can allow for the mental to be a cause despite being necessitated by a physical state. I also discuss a response to Woodward put forward by Baumgartner.

8.3.1 Woodward's Interventionist Account

According to Woodward, his (2003) interventionist account can allow for a mental state to be the cause of a physical state (or further mental state[19]) even if the mental state is necessitated by another physical state because of a supervenience relation (see Woodward (2015),(2017)).

Baumgartner ((2009), (2010)) argues however, that it is not possible on Woodward's interventionist account, at least not without weakening it to the point of being unsuitable to do the work non-reductive physicalists want it to. To see why let me introduce a causal graph taken from Woodward (2015).

> "(O1) if m had happened without p, e would still have happened (m & $\neg p$) $\Box \rightarrow e$, and
> (O2) if p had happened without m, e would still have happened (p & $\neg m$) $\Box \rightarrow e$."
> (Bennett (2003) p.480)

[19]The debate between Woodward and Baumgartner is framed in terms of a mental event (or property) causing (an instantiation of) a further mental event (or property). Therefore, for the purposes of this section, and the causal graphs therein, I will do the same. But, the same points will apply to cases of mental to physical causation.

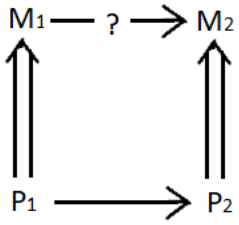

Figure 8.2: A Graph in which M_1 and M_2 supervene on P_1 and P_2 respectively and P_1 causes P_2

Figure 8.2 represents a case where a mental state M_1 supervenes on physical state P_1 shown by the double-tailed arrow. Likewise, mental state M_2 supervenes on P_2. P_1 causes P_2 as reflected by the single-tailed arrow. The question at hand is, does M_1 cause M_2?

Baumgartner argues that the interventionist encounters their own exclusion argument involving the following three premises:

> "(1) causation is to be spelled out in terms of (Woodward's interventionist account), (2) a macro property *X* supervenes on a physical micro supervenience base MSB(*X*) such that $X \neq$ MSB(*X*), and (3) MSB(*X*) is causally relevant to a micro effect *Y*" ((2009) p.169)

It's impossible to intervene on M_1 with respect to M_2 or P_2 because of the need to hold P_1 fixed. This is because, on Woodward's view, interventions on M_1 need to be independent of all other off path variables which can also change M_2. As P_1 is not an intermediate on the path from M_1 to M_2, it will need to be held fixed. The reasoning for holding off path variables fixed is to prevent those other

variables from potentially confounding causal results. Woodward introduces an example involving smoking (S), having yellow fingers (Y) and having lung disease (D) ((2015) pp.310-311). He uses this example to illustrate how interventions must be done carefully to avoid confounding ((2015) p.314). Say an experimenter manipulated Y by intervening on S. Because smoking causes both yellow fingers and lung disease intervening on S also leads to variation in D. Y and D will both correlate under interventions on S but this correlation does not reflect any direct causal relation between Y and D. Such a correlation would be "spurious" ((2015) p.338) in that "manipulating (Y) is not a way of manipulating (D)" ((2015) p.338). Rather, if an experimenter wanted to investigate if Y does cause D, they would have to hold other causes of D (such as S) fixed. Indeed, once S is held fixed, there will be no correlation between Y and D.

To return to the mental case, Baumgartner argues that it is precisely because there can be no change in a mental state without a change in its underlying physical base that means it will be impossible for M_1 to cause M_2 or P_2. This is because once you hold the supervenience base for M_1 (that is P_1) fixed, there is no change in M_2 when you intervene on M_1. According to Baumgartner, as this will always be the case with supervenient properties, they are always causally inert. And as both Woodward and Baumgartner note, this point will generalise from the mental case to the special sciences.

Woodward believes that Baumgartner is mistaken in his assessment of his interventionist account to the point of begging the question of the exclusion argument. In a sentence, Woodward argues that Baumgartner is wrong to

control for, or hold fixed the supervenience base of the mental state. Woodward claims that Baumgartner is misapplying the use of causal graphs. He is attempting to read a "mixed" graph which includes the non-causal dependency relation of supervenience as a purely causal graph which includes only potential causal relations. Figure 8.3 below is a graph depicting only causal relations such that Baumgartener's argument holds on Woodward's reading of his interventionist account.

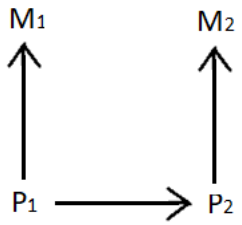

Figure 8.3: A Causal Graph showing only causal dependency relations in which P_1 causes M_1 and P_2 and P_2 causes M_2

Here if we hold P_1 fixed and intervene on M_1 it will be the case that there is no change in M_2 and therefore M_1 does not cause M_2. But, Woodward ((2015) p.308) and Baumgartner ((2009) p.170, (2010) p.388) both agree that however the supervenience relation is to be understood exactly, it is widely considered not to be a causal relation. So to represent mental causation in graphs such as Figure 8.3 is to misrepresent the situation. Rather, the mixed graph in Figure 8.2 is appropriate but cannot be straightforwardly read in the same way as a purely causal graph.

So what is the correct way to read graphs such as the one in Figure 8.2? Which

variables is it correct to hold fixed? Or, more specifically, is it appropriate or inappropriate to hold supervenience bases of variables fixed as if they were off path variables? Woodward says no, it is inappropriate to do so precisely because it is not possible to do so. Any intervention on M_1 must be accompanied by a change in P_1 because of the nature of supervenience. Therefore, any intervention on M_1 should be considered the same as an intervention on P_1 as depicted in Figure 8.4 taken from Woodward ((2015) p.331).

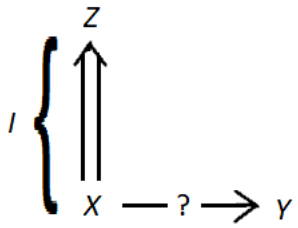

Figure 8.4: A Single Intervention *I* Operates on Both *X* and *Z*. From Woodward ((2015) p.331)

But why does Woodward think this is the correct way to treat supervenience bases? He makes an analogy to another category of non-causal dependency relation between variables. "Variables that bear definitional relations to *X* and *Y* should not be thought of as potential `confounders' that need to be controlled for in the way that variables that bear causal or correlational relations to *X* and *Y* may be confounders requiring control" ((2015) p.336). He gives an uncontroversial example from Spirtes and Scheines ((2004) p.836) involving cholesterol and graphed in Figure 8.5.

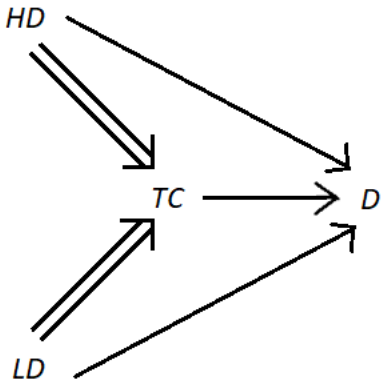

Figure 8.5: A Mixed Graph Showing Causal (Single-Tailed Arrows) and Definitional Dependency Relations (Double-Tailed Arrows). From Woodward ((2017) pp.260-261)

As a matter of definition total cholesterol (TC) is made up of high-density cholesterol (HD) and low-density cholesterol (LD) such that TC = HD + LD. The double-tailed arrows in Figure 8.5 therefore represent a definition al as opposed to a supervenience (non-causal) relation. The single-tailed arrow still represents a causal relation. In this case HD, LD and TC all cause heart disease (D). However, it is impossible to intervene on HD while holding both LD and TC fixed and likewise, impossible to intervene on LD while holding both HD and TC fixed.[20] So according to Baumgartner, HD is not a cause of D. This conclusion seems wrong though. Better, Woodward says, to understand the graph differently. Rather, an

[20]TC would be considered an off path variable given the direct causal relationship between HD and D if all the arrows in the diagram were single-tailed, or if all arrows were interpreted as if they were representing causal relations. Given that, according to Woodward, this is the wrong way to interpret this graph, TC can not really be considered an off-path variable and so, Woodward says, speaking of holding it fixed while intervening on HD does not make sense.

intervention on either HD or LD is also an intervention on TC such that if we were to intervene on HD we would also track the change in TC. Holding LD fixed, and intervening on HD there would be a change in D so we can say HD is a cause of D. Of course, it is important not to 'double count' the causal influence of variables related in non-casual ways. For example, it is important not to double count the effect intervening on HD has on D by interpreting this effect as being something over and above the change produced by TC. It is the same change. This is why it is much less straightforward to work out the results of interventions on mixed graphs than purely causal graphs.

Woodward argues that his treatment of this case is uncontroversial. As the supervenience relation is a non-causal dependency (as the definitional relation is) then we should apply the same treatment to supervenience cases as we do to definitional cases. So if we were to intervene on M_1 rather than hold P_1 fixed, we should act as though it is also intervened on and see then if there is a change in M_2. As it is possible that M_2 could change when such an intervention is made then, contra Baumgartner, it is possible for M_1 to cause M_2 or indeed P_2. Therefore, although intervening on a physical or mental state may always be accompanied by a corresponding change in its supervening property or supervenience base, this doesn't mean the associated counterfactual is vacuous.

8.3.2 A Two Strand Solution

Perhaps a two strand approach can be taken here inspired by Bennett's (2003) approach to the overdetermination problem.[21] To recap, the question in hand is can the physical state occur without the mental state? In other words can (P & ¬M) obtain or not and thus how do we understand the counterfactual (*P* & *¬M*) $\Box \rightarrow$ *E*? There are two possible routes. Either (P & ¬M) can obtain or it cannot. Correspondingly the counterfactual (*P* & *¬M*) $\Box \rightarrow$ *E* can either be vacuous or not.

Let's say first that (P & ¬M) can obtain. Then the counterfactual it is the antecedent of is not vacuous. Woodward's arguments from the previous section can be put forward (among others) to argue that is can obtain. Inequalities such as P(E | A & M) > P(E | A & ¬M) can therefore hold and mental states can raise probabilities or otherwise fix them over and above the contribution made by the physical.

On the other hand, the more problematic case for me is if (P & ¬M) cannot obtain (that is you think that the physical does necessitate its associated mental state) then the counterfactual it's the antecedent of is true but vacuously so. This is potentially problematic for me in that I want to argue that inequalities such as P(E | A & M) > P(E | A & ¬M) holding show that mental states can raise (or lower) probabilities over and above the physical states contribution. There are possible solutions in such cases however.

[21] I discussed her arguments in detail in sections 7.1.2 and 7.3.2.

One solution would be to argue as Bennett does that is the counterfactual is vacuous then there is no (bad) overdetermination in play. Therefore, according to Bennett, the counterfactual we must assess is (P^* & $\neg M$) $\Box \rightarrow E$ where P^* contains extrinsic elements not contained in P. These extra physical circumstances are what means that P^* necessitates the occurrence of its associated mental state, that is M must obtain. So while P^* can fix the probability distribution of E it cannot do so without M also obtaining.

Alternatively, in such a case, I argue you could take the approach that Woodward takes in thinking that "it is not legitimate to use such counterfactuals in assessing causal efficacy" ((2015) p.335). However you view possibility of the physical state occurring without its corresponding mental state then, I argue there is a route you can take to accommodate mental causation.

8.3.3 Why is my Account Preferable?

The twin aims of my thesis have been to show that the CEA doesn't hold in probabilistic worlds (like ours) and to put forward a world view in which mental causation exists. To that end I have tried to keep as much open as I could. For example, I have put forward a world view based on a probabilistic counterfactual theory of causation, but I think that my view could be amended to accommodate other theories of causation.

But why should anyone prefer my account? A sceptic could always ultimately argue that the CEA holds as much in its probabilistic version as it does in its

deterministic version. They could argue that the burden of proof lies on my shoulders given that I make the comparatively un-parsimonious claim this class of non-physical causally efficacious events exist.

I argue however that the reverse is true. The burden of proof lies on the opponent of mental causation. This is because of arguments I have made elsewhere in this thesis such as the argument from the MMI, the argument from evolution and the argument I will go onto make in the next chapter based on natural kinds. Although I have mentioned the principle of the MMI several times throughout this thesis, I believe it bears repeating one last time:

> The Mental Manifest Image: Things are how they appear to us in our introspections and mental phenomenology and we should try to accommodate this in our philosophical theories in so far as is possible given our best current scientific theories.

If we have evidence of anything it's of our introspections and mental phenomenology. And they present the world to us as though mental causation exists. This is why I think my account is preferable to accepting the conclusion of the CEA.

9

NATURAL KINDS

"The natural phenomena that take place every day before our eyes did not escape my examinations"

Frankenstein - Mary Shelley (1818)

In this chapter I will make my last positive argument in favour of mental causation centred on the concept of natural kinds. In a sentence, if mental states can form natural kinds, and more perfectly natural kinds than their corresponding physical states can, then they can play a role in laws of nature. If they play a role in laws of nature, then they can be causally efficacious.

9.1 Natural Kinds and Scientific Laws: A Case for Mental Causation?

The aim of this chapter is to motivate one argument for mental causation which at the least places the burden of proof back onto the sceptic. I have put forward a world view along with arguments as to why mental states can be causally efficacious. I argue that my view is at least as plausible as alternative views which rule out the existence of mental causation and that therefore it is for the sceptic to argue why we should continue to rule out the possibility of mental causation. Further, by the principle of the Mental Manifest Image[1] if two theories are equally explanatory but only one allows for mental causation then that theory is to be preferred. In this case if I can argue that mental states are natural kinds and therefore can participate in scientific laws then I hope to place a burden of proof onto the sceptic to explain why mental causation cannot exist.

A final brief recap of the causal exclusion argument would be helpful here to place my argument in context. I present the probabilistic analogue of the CEA here but I believe what I have to say in this chapter could apply whether the world was probabilistic or deterministic. After all, as I briefly mentioned in the opening to Chapter 6, the two versions of the CEA are the same if the

[1] As a reminder, the Mental Manifest Image states that things are how they appear to us in our introspections and mental phenomenology and we should try to accommodate this in our philosophical theories in so far as is possible given our best current scientific theories. See Section 5.1.

only probabilities which can be assigned are 1 or 0 as would be the case in the deterministic setting. The original CEA could therefore be thought of as a special case of the Probabilistic Analogue CEA.

(P1') Probabilistic Causal Closure of Physics
Every physical event has a physical cause which is sufficient to fix its probability.

(P2') No Systematic Overdetermination of Probabilities
It is not usually the case that there are multiple sets of events that are minimally sufficient to fix the probability of a further event which exist simultaneously.

(P3) Non-Identity of the Physical and the Mental
There are physical states and mental states and these are not identical to each other.

(C) There is no Mental Causation
Mental states cannot be causes of physical effects.

If mental and physical states are not identical and if physics is causally closed, then there is no non-overdetermining work left for mental states to do. Therefore, there cannot be any mental to physical causation. For the purpose of this chapter, lets say that (P2') No Systematic Overdetermination of Probabilities is true. So, lets say that there cannot simultaneously be a mental and a physical overdeterminer of an event's probability (whether this is fixed to 1 or less). It could still be wrong to conclude that mental states can *never* be a cause. It may be that in some cases we should rather dismiss the corresponding physical state as a cause. In sum, this chapter I want to argue that there may actually be

work that mental states are better suited to do than their corresponding physical brain states and that they therefore have a non-overdetermining causal role.

In a sentence, the idea is that if mental states can be a better candidate than brain states for the natural kind role in scientific laws, then this is evidence that it is the mental rather than the brain states which are doing the causing. This is because laws of nature underwrite causal relations on many theories of causation.

This chapter is therefore dedicated to discussing what makes something a natural kind, what makes something a scientific law, and why it is the case that to feature in the latter, you must be the former. This is important for my view because if mental states can qualify as natural kinds which feature in scientific laws then this can add strength to the argument that they are causally efficacious. I will begin by arguing that we have prima facie evidence that mental states can figure in laws. I will then discuss if mental states can be considered as genuine natural kinds (albeit imperfect ones). To do this I will consider various criteria for natural kindhood which have been posited. I will argue that, bar one (which may be too strong a criterion), none rule out that mental states could constitute natural kinds. I will discuss Yablo's (1992) paper "Mental Causation" and argue that his argument is incorrectly premised on multiple realizability. Therefore, while Yablo and I end up agreeing that mental causation exists, we get to that conclusion by differing paths. I will then argue that mental states can plausibly be thought of as more (albeit imperfectly) natural than the physical states which are associated with them. Therefore, they

are more suited to featuring in laws and playing a causal role, at least in some circumstances. I then conclude the chapter by tying the overall discussion back to the CEA.

9.1.1 Scientific Laws

What is a law and what demarcates a scientific law from, for example, pseudoscientific laws? This will be important as non-lawful generalizations cannot underwrite any causal relations whereas scientific laws can. In particular I will be looking at 'laws' from the field of psychology because if mental states can play a role in laws then these seem to be a good candidate for the type of law they would appear in.

Psychology is a special science which studies an area (the human mind and behaviour) which has sometimes been considered resistant to being analysed in a nomological way. For example, Davidson (2001*b*) states "mental events such as perceivings, rememberings, decisions, and actions resist capture in the nomological net of physical theory" ((2001*b*) p.207). Indeed some people, including Davidson, have questioned psychology's status as a science due to the fact that they doubt it's possible to formulate psychological laws. In "Psychology as Philosophy" Davidson says that psychological phenomenon are not "even in theory, amenable to precise prediction or subsumption under deterministic laws" ((2001*c*) p.239). As a result "psychology is set off from the other sciences in an important and interesting way" ((2001*c*) p.241). So, as I am going to argue that there can be such things as psychological (and other

special science) laws, I have to make sure that I have a conception of laws which is up to the job I have assigned them.

What Makes a Law a Law?

There is usually held to be an intimate connection between causation, laws and kinds. The precise nature of this connection will differ depending on your exact philosophical outlook. Laws are generalisations. They can either be exceptionless[2] (as the more fundamental laws arguably are) or ceteris paribus (as many if not all special sciences laws are).

It's also widely held that laws support causal relations, indeed many of the leading theories of causation invoke laws in some way or another. For example, Fodor argues that "intentional properties are causally responsible in case there are intentional causal laws" ((1989) p.65). Davidson states "events related as cause and effect fall under strict deterministic laws" ((2001*b*) p.208), indeed he calls this the "Principle of the Nomological Character of Causality" ((2001*b*) p.208). For Davidson, there are no such things as strict psychophysical laws.[3] More generally, regularity theorists would hold the connection between laws and causation to be direct. Laws often feature in the counterfactual semantics of counterfactual accounts of causation (see for example Chisholm (1954)). Likewise, laws feature in probabilistic accounts of causation, as relevant proba-

[2]For example, Hempel (1965) believed they were exceptionless generalisations which were more than accidentally true so that they could feature as premises in deductive-nomological models.

[3]For more on Davidson's views please see Section 2.3.1.

bilities must be lawfully-derived objective chances such as those from a Lewisian Best System (1994). The Woodwardian interventionist approach relies on invariant generalisations (2003). While not equivalent to laws of nature as typically understood they do share some common features.[4] One such feature is that, like laws, invariant generalisations can support interventionist counterfactuals in a way accidental generalisations can not ((2003) p.279). Lastly, process theorists (such as Salmon (1997) and Dowe (1995), (2000)) will decide which quantities are conserved by looking at the laws of nature.

Scientific laws from physics are the laws 'par excellence'. They are supposed to be universal and exceptionless. An example of such a law is the Schrödinger equation (see Schrödinger (1926)). Biology and chemistry have their own respective laws (leave aside for now the question of how these sets of laws relate). For example, the law of the conservation of mass in chemical reactions is fundamental to chemistry and cannot be theoretically broken. Laws in biology on the other had are less hard and fast and more generalised rules. Take for example Mendel's First Law. Godfrey-Smith notes that exceptions to the law include "cases of Down Syndrome in humans, and cases where particular genes have evolved the capacity to make their way into more than their fair share of sex cells" ((2014) p.12). Often they are said to hold ceteris paribus.

We have prima facie reasons for thinking that mental states can figure in laws. Many thinkers hold that there can be psychological laws. For example, Lep-

[4]Woodward even states "laws are simply generalisations which are invariant under a particularly wide range of changes and interventions" ((2003) p.240).

ore and Loewer argue in "Mind Matters" (1987) that there can be, contrary to Davidsons' view, laws involving psychological properties which can support counterfactuals even if those laws are non-exceptionless. Fodor also argues that there can be psychological laws in "Making Mind Matter More" (1989), although for different reasons. He claims that a property can be causally efficacious if it can support nomic sufficiency for its effect. Further, he claims that psychological properties are among those properties which can support nomologically sufficient relations. In a sentence; "mental properties are causally responsible because there are intentional generalizations which specify nomologically sufficient conditions for behavioural outcomes" ((1989) pp.69-70).

So what kinds of psychological generalisations could be put forward as laws? In conversation with Dr Jianan Bao, a practicing psychiatry trainee working for the NHS we discussed a few examples. They ranged from the neurobiologically based to the more purely pyschologically based. For a more biologically based example she gave the progression of Alzheimer's disease. Alzheimer's disease attacks the brain chipping gradually away at the memory and personality of the sufferer (see *Alzheimer's disease - Symptoms* (2018)). The mechanisms by which Alzheimer's disease effects the brain are still not fully understood. But the effects of the disease are sadly familiar to many. There seems to be a pattern to the memory degeneration that occurs in that patients with this disease tend to hold on to longer term memory while losing their ability to make and retain recent memories. As the disease progresses they lose more and more function. While there are many exceptions this pattern informs how patients are treated

and can be used to make predictions about behaviour and prognosis. Overall, generalisations can be made with some level of certainty about the progression of the disease and the patients life.

This falls more heavily on the neurobiological end of the examples as the mental degeneration in Alzheimer's is a direct result of damage to the brain. Dr Bao also gave another example which was of patients with temporal lobe damage which reliably leads to greater levels of inhibition. Phineas Gage is the textbook example of physical brain damage resulting in behaviour change. Briefly, in 1848, Gage was working on constructing a railway. An accidental explosion resulted in a tamping iron (a metal rod) being blasted through Gage's brain. Amazingly, he survived the accident but suffered psychological and personality changes (see Tobia (2017) for an overview of the case).

In terms of the more psychological 'laws' Dr Bao suggested generalisations such as peoples' response to death or trauma. Human psychology is such that we tend to make attachments with other human beings. When those other human beings are hurt or die, we tend to have adverse psychological outcomes (such as sadness, anger and stress[5]). Psychological stress is a common response to experiencing trauma. And experiencing higher levels of stress is a reliable sign that you are more predisposed to mental health problems. However, these are all examples of physical to mental causation. What I need to provide are mental to physical examples.

[5]An important note about how granular to be when discussing mental states will be made in Section 9.2.4.

Perhaps placebo laws which I discussed in Section 5.3 could also be cited as examples. I further wonder if standardised psychological treatments (such as SSRIs, Cognitive Behavioural Therapy and psychotherapy to name a few examples) can work in lawlike ways or be effective because they tap into underlying laws. Certain conditions respond to certain treatments but not others. For example, SSRIs treat depression and anxiety (see the NHS website (2017)) but not schizophrenia for which another class of medication, antipsychotics (see again the NHS website (*Schizophrenia - Treatment* (2017)).[6] Analogously there are different talking therapies tailored to different conditions. For example the NHS website says "for some problems and conditions, one type of talking therapy may be better than another" (2019). This is a case of mental interventions producing (among other things) physical behaviour changes in a generalisably reliable way.

All of these generalisations are just that; very general and imprecise. They do not always hold and cannot always be used to make *precise* predictions, explanations or inferences. But, generalisations such as these do have explanatory, predictive and inferential value. For instance, if as a doctor, you have to give bad news to a patient, there are certain procedures to follow. There are certain reactions you are taught to expect. And there are certain actions you must

[6]Examples involving medicating mental heath conditions are also examples of physical to mental causation so in that sense are not pertinent to the discussion of the CEA specifically. However, they are pertinent in so far as they can demonstrate the possibility of laws of psychology. Perhaps, as I go on to suggest, similar laws could be found linking different kinds of talking therapy to different conditions which could act as psychological laws involving mental to physical causation (on the plausible assumption that the talking therapy can impact on behaviour by way of intervening on mental states).

carry out as a result, offering grief counselling for example. Furthermore, they hold to the extent that people who fail to follow these 'rules' or break these expectations are sometimes deemed to be pathological. For instance people who can't form attachments to others can be diagnosed with a range of mental health problems such as anti social personality disorder. From discussion with Dr Bao, it seems as though these generalisations, and ones like them, do inform her work and therefore are of clinical and scientific value.[7]

The psychological laws mentioned above don't hold to the same exceptionless standard of some physical laws. But this is not problematic, indeed it is to be expected. Mental states may not be as natural as the most perfectly natural physical kinds (although it is perhaps not possible to immediately rule this out). But, importantly, they are arguably at least imperfectly natural which is enough to support a nomological generalisation.[8] Special science properties are imperfectly natural and it's plausible that there are special science laws. So it's possible that psycho-physical laws could work the same way as other special sciences generalisations. I agree with Fodor ((1974), (1991), (1997)) that imperfectly natural kinds will feature in non-exceptionless or ceteris paribus laws. This is because as they are imperfectly natural there maybe more variation within the kind which translates into less generalisability. To conclude this section, it seems we have at least prima facie evidence that mental states can figure

[7]Harré offers some more examples of psychological laws such as "Information is first retained in the short-term memory store" ((2002) p.70).

[8]There may be perfectly natural mental states, the possibility of which I will discuss below (see section 9.3.2) to which this point would not apply. However, even if you hold there can be perfectly natural mental states, it doesn't follow that all mental states are perfectly natural.

in laws. This still leaves open the question as to whether mental states can constitute genuine natural kinds whether perfectly or imperfectly. To this end I will now discuss a variety of criteria for natural kindhood which have been suggested and argue that no compelling criteria gives us reason to think that mental states cannot be considered (imperfectly) natural kinds.

9.1.2 Natural Kinds

What is a natural kind? Can mental states be classified as natural kinds? And are the brain states underlying those mental states better suited to play the natural kind role? These are the three key questions for this subsection. I will begin by outlining potential criteria for natural kindhood.[9] I will argue that none of the criteria suggested rule out that mental states can constitute natural kinds, except Ellis'. I will argue that this is not problematic for me, as Ellis' criteria may be too strong given that it would also rule out special science kinds as being considered imperfectly natural.

Briefly, Lewis viewed natural kinds as either perfectly natural or imperfectly natural. On Lewis' view, perfectly natural kinds correspond to universals while imperfectly natural kinds correspond to a close-knit family of univerals (see for example Lewis (1983)). Imperfectly natural kinds are not gerrymanders though and are still 'natural enough' to feature in laws. I should note that it's not necessary to hold a Lewisian view of natural kinds for my argument to work

[9]It should be noted that I will take a naturalist stance towards kinds and therefore put constructivist theories and issues to one side. See Hacking ((1983), (1999)), Armstrong (1997) and Dupré (1993) for examples of naturalist philosophy.

although I will be making use of his concepts of imperfectly and perfectly natural kinds.

Natural kinds are types or groupings of things specified particularly by the ontology of a scientific theory. There are various categories of kinds besides natural, for example, social and gerrymandered[10] kinds. While these categories are not always mutually exclusive, I will focus on discussing only natural kinds.[11] A natural kind is a type of thing which is found to be grouped together by objective resemblance (Lewis (1983)) in the world rather than by a categorization we impose on the world. Importantly, naturalness can admit of degrees.[12] As each science has its own taxonomy of kinds, the natural kinds can be subdivided into physical kinds, chemical kinds, biological kinds and so on. Examples include 'electrons', 'gold' and 'FN1 gene[13]' respectively.

[10]A gerrymandered kind is a kind which contains members which lack an appropriate level of objective resemblance. Thus gerrymandered kinds do not qualify as natural kinds. For example, take a disjunction with disparate disjuncts. They don't have to be entirely arbitrary from certain points of view. For example, I could claim that all the objects in my living room form a kind. However, this kind would lack objective resemblance and would not appear in scientific theorising. A social kind on the other hand, is a kind which is socially constructed. For example, 'gender' as a societal role is therefore a social kind. Of course the debate around the precise nature of 'gender' is complex, nuanced and still ongoing. See for example Butler ((1993), (1999)), Haslanger ((2012*a*), (2012*b*) and (2013)), Carlson (2010) and Bach (2012).

[11]So while a gerrymandered kind could never be natural, I do not want to rule out that a social kind may fall within the natural kind spectrum. But as stated I will put social kinds aside now.

[12]A terminological note must be made here. When I use the term 'natural kind' I actually mean a kind which falls somewhere on the perfectly to imperfectly natural spectrum whether or not I explicitly state this. Another terminological note I should make is that I will speak of mental states and mental properties as though they are interchangeable. More properly when I speak of mental states I mean instantiations of mental properties, but for ease of writing I don't always state this.

[13]This is the gene that codes for the protein fibronectin (*FN1 Gene - GeneCards | FINC Protein | FINC Antibody* (n.d.)).

As noted not all kinds can qualify as natural. Gerrymandered kinds count as unnatural and are therefore not suitable for inclusion in scientific laws. This is because the members of a gerrymandered kind don't resemble each other in any relevant way. Take Goodman's 'grue'; a property which "applies to all things examined before *t* just in case they are green but to other things just in case they are blue" ((1983) p.74). Grue is not taken to be a natural kind, or, in other words, is considered unprojectible, and therefore unfit to feature in laws. If mental states did count as gerrymandered in some way then they would therefore not be able to feature in scientific laws and my argument would not go through.

One last clarification should be made here regarding sortal kinds. Sortal kinds are the kinds that can be used to identify and count things. They define the properties of which are essential to be a thing of that kind.[14] If an entity has these properties then it is necessarily a thing of that kind. Furthermore, they must have these properties, if it lacks one or more of them then, again necessarily, it cannot be a thing of that kind. Take for example an electron. An entity can't have the 'electron' set of properties (for example, being negatively charged, having a certain mass and spin and so on) and not be an electron. Likewise if an entity lacks these properties then it is not an electron. In other words, this set of properties is essential to what it is to being an electron and it's sufficient to have these properties in order to be an electron. Knowing this allows us to identify and count those entities which are electrons.

[14]See Locke (1970) and Strawson (2006).

On the other hand there are natural properties which are not sortal[15]. Take the property of being negatively charged. In and of itself this is not a sortal kind though it is a natural kind. We can see this when considering that different entities belonging to different sortal kinds can both have the property of being negatively charged. For example, both electrons and muons are negatively charged, in fact they have the same level of negative charge. If they both share this property, but belong to different sortal kinds, then the property of being negatively charged can't be essential to any one kind and therefore cannot be a sortal kind in and of itself.

Similarly and importantly mental states probably do not form sortal kinds but plausibly are at least imperfectly natural. Take pain as an example. There is arguably no way of individuating and counting pain states. Compare this to the sortal kind `being an electron'. We can pick out those entities which are electrons and count how many there are in any given place (in principle at least). Importantly though; a kind does not need to be sortal to feature in laws, therefore I will not restrict myself to focusing only on sortal kinds. Rather I will be interested in the broader category of natural kinds (both the perfectly and imperfectly natural kinds). So for my purposes the class of sortal kinds and the class of natural kinds will cross cut but not being a sortal kind is not enough to

[15]Likewise, even within sortal kinds there can be a good degree of variation, so not all sortal kinds will be perfectly natural. Take the example of cats. There is a fair amount of variation between different cats. But there is enough resemblance shared between members of the group to qualify as a kind. So, enough cats share the properties of being a mammal, having whiskers and a tail, being carnivorous and so on. Because of this variation it seems that not all sortal kinds will fall into the perfectly natural category as this allows of much less variation.

disqualify a kind from being natural. Now I will turn to some criteria which have been suggested for natural kindhood.

Criteria for Natural Kindhood

I will now consider four different criteria that have been put forward for understanding natural kinds.[16] First, the suggestion that kindhood membership should allow for inductive inferences and then secondly that kindhood membership should allow for featuring in scientific laws. Third, I will discuss the criterion of shared natural properties and objective resemblance with a focus on Lewis' account of kindhood. Lastly I will briefly mention categorical distinctness.

The first criterion I will discuss is that natural kindhood membership should allow for inductive inferences for members of that kind and vice versa (see Quine (1969)). In other words you should be able to infer things about the whole kind without having to observe every member of that kind. This is because all members of the kind are relevantly similar to each other in a way that members of a gerrymandered kind are not. For example, given that this sample of gold conducts heat, you can infer that all samples of gold (at least gold of the same quality) will also conduct heat.

However, as Bird & Tobin (2018) note, this is at best a necessary but not a sufficient condition. This is because the inferences which can be made about kind groups depend upon the similarity of the members. If the members of the kind are not appropriately similar then inferences about the whole kind can't

[16] I take my exposition from Bird & Tobin (2018).

be made.

This criteria does not seem to rule out the possibility that mental states can be at least imperfectly natural kinds. On a general level, different pains are relevantly similar enough to each other that certain inferences can be made regarding them. Perhaps this is even truer if we talk about specific types of pain, such as headaches or throbbing headaches.

Relatedly, a criterion which has been posited is that natural kinds can play roles in laws of nature, indeed a kind must be natural in order to play such a role. Thus, gerrymandered kinds cannot participate in laws of nature. My argument does not require that kinds must participate in laws of nature in order to be considered a natural kind. Indeed that would be circular as my argument is that mental states can figure in laws and thus play causal roles because they are natural kinds. But if a kind must participate in laws of nature in order to be considered a natural kind then it would be so much the better for my view.

As Bird & Tobin (2018) note, this is a stronger version of the first criterion because it is the inductive inferences you can make between kinds and members of that kind which are closely related to the laws. For example, take copper (I take this example from Goodman ((1983) p.73). It is a law that copper conducts electricity. Therefore, if you know an object is a member of the kind 'copper', you know it will conduct electricity. The resemblance between members of the kind copper are such that they support the lawlike generalisation. Hence the natural kind can act in a law of nature.

Perhaps this becomes clearer at the higher level special sciences which feature less perfectly natural kinds. Given the looser objective resemblance between members within such kinds, it's possible that special science kinds won't feature in any *exceptionless* laws of nature. Cats may have appeared in Schrödinger's (1935) thought experiments but not in his equations. This is not to say however that they cannot feature in any laws. If Fodor ((1974), (1997)) and others are correct then there are special science laws. So, while it would be circular for me to rest on this as a criterion for natural kindhood, it is not inconsistent with my claims and gives us no reason to think mental states can't be considered as natural kinds.

Another criterion which has been suggested and which seems intuitive is the suggestion that the members of the kind must share some 'natural property' in common. For example, as mentioned, all electrons share the natural property of being negatively charged. Mill (1846) suggested that sharing a natural property is at best a necessary but not a sufficient condition for natural kindhood. This is because a group of objects can share a natural property but nevertheless not form a kind or represent a classification in nature. He pointed out that all white objects share the natural property of 'being white'. But, this group will otherwise be very heterogeneous and unsuitable for consideration as a kind in itself.

An argument could also be put forward that the criteria that members share a natural property in common isn't a necessary condition. This argument comes from the philosophy of biology specifically from the debate around whether

species count as natural kinds.[17] The thought is that species are natural kinds. However, because of the nature of evolution, change is ever present both inter and intra-species wise. In other words, a lot of change has to take place within a species before that species can be said to have evolved into a new one. The result is that members of the same species can sometimes have fairly different properties of which they share few. It becomes harder to see how inferences can be made across such a group compared to a group which shares all or most of its properties in common.

However, this may say more about our concept of species (and its need for refinement) than it does about the criteria for natural kindhood. It could be that we have not yet developed sufficiently sophisticated scientific tools and methods with which to properly classify species. Or it could also be the case that species are not in and of themselves natural kinds.

Perhaps there is nothing particularly wrong with our concepts of species or kindhood. Species fall on the more imperfect end of the natural kind spectrum so the lines which differentiate different species look more blurry and vague than the lines between more perfectly natural kinds. So, when making inferences about a less perfectly natural kind the inference will not be as straightforward as it would be in the case of a perfectly natural kind. For example, it may be that there is a "close-knit family" (Lewis (1983) p.347) of properties in which members of the species kind share. Not every species member will have every property but each member will share enough from the pot (or will share in enough

[17] See among others, Kitcher (1984), Elder (2008) and Ereshefsky (2016).

of the most important properties) that they resemble each other enough to qualify for kindhood, at least imperfectly natural kindhood. As long as there are enough shared properties from a core set, in a Wittgensteinian (2010) `family resemblance' sense, then perhaps this is good enough to qualify as membership in a kind and will allow for some inferences to be made.

Lewis' view is that a perfectly natural kind is one whose members "are all and only those things that share some one universal" ((1983) p.347) where universals are "repeatable entities, wholly present wherever a particular instantiates them" ((1983) footnote 2, p.343). A property is a class such that class membership means a particular has that property. An imperfectly natural kind is so "in virtue of a close-knit family of genuine universals one or another of which is instantiated" by its members ((1983) p.347). The kinds specified by physics such as being an electron would lie on the perfectly natural side of the spectrum while being a cat (and quite possibly mental states) would fall on the imperfectly natural side. In order for two objects to be within the same kind, then those two objects must objectively resemble each other. Resemblance in this case will be cashed out in terms of these shared natural properties or in the case of imperfectly natural kinds, sharing from the `close-knit family' pot.

Lewis gives the example of "being metallic" ((1983) p.347) as an example of an imperfectly natural property. He states that even if there were no such universal as `metallic' there is a "close knit family of genuine universals one or another of which is instantiated by any metallic thing" ((1983) p.347). He also specifies that naturalness comes in degrees. So, other special science kinds will fall more on

the imperfectly natural end of the spectrum. This is because they are closely related to the fundamental kinds which are perfectly natural[18], but are not in themselves fundamental and therefore are only imperfectly natural. Indeed, this suggests there can be degrees of `imperfect naturalness' just as there are degrees of fundamentality.

However, the more imperfectly natural kinds are amenable to multiple realization.[19] For example, the property "being metallic" can be instantiated in numerous different ways, by an iron object, a steel object, a copper object and so on. It still seems plausible to say "being metallic" is made imperfectly natural by its relation to the more fundamental kinds it shares in. In these cases the kind is less perfectly natural as its instances less perfectly resemble each other.

Fodor argues that functional kinds can be perfectly natural even if the physcial bases underlying the functional kind are not. He says, "it is unlikely that every natural kind corresponds to a physical natural kind" ((1974) p.103) but there are higher level functional kinds which can also be considered quite natural. Fodor gives the example of "monetary exchanges" ((1974) p.103) such as using different forms of cash and using cheques. Nevertheless helpful generalisations, Fodor mentions Gresham's law ((1974) p.103), can be made regarding members of the kind in question. Fodor reasons that even when two entities differ in their

[18]Indeed they are the fundamental kinds because they are the perfectly natural ones. And because the kinds of fundamental physics are the perfectly natural ones, they will feature in the fundamental kinds. This raises the question of whether only the physical kinds can be the fundamental or perfectly natural ones? If this was the case it wouldn't disprove my argument though as it's plausible that imperfectly natural kinds can feature in laws. I discuss this issue more in Section 9.3.1 on mental kinds below.

[19]For a more thorough discussion on multiple realization see my Section 9.2 below.

physical implementation they "must nevertheless converge in indefinitely many of their properties" among which some may have "lawful inter-relations (which) support the generalisations of the special sciences" ((1974) pp.113-114).

However, the question of whether functional kinds can be considered natural, whether perfectly or imperfectly so, is slightly tangential to my arguments. This is because I don't hold mental properties, at least qualia, to be functional kinds. So, the important question for me is whether mental states can plausibly be held to be natural kinds and can therefore feature in laws of nature.

This raises the question of how perfectly or imperfectly natural mental states are? Can there be mental states which consist in instantiating a universal and thus are perfectly natural? If this is the case then it would be all the better for my claims (though not necessary). I consider these questions below in Section 9.3.1 where I discuss mental kinds and universals. Once again though, this criteria gives us no reason to rule out mental states as candidates for imperfect natural kindhood.

I will lastly briefly mention an idea supported for example by Ellis (2001) that natural kinds must be categorically distinct. The idea is that there are no 'fuzzy' lines between natural kinds, rather they fall into distinct categories. The exemplar used to illustrate this is the chemical kinds. There cannot be a chemical element with 'many' protons, the number must be determinate. This shows that these lines are drawn by nature and not by us. Vaguer or more ambiguous distinctions are down to human interpretation placed upon nature and therefore kinds with

vaguer boundaries are not natural. Mental states would plausibly fall foul of such a criteria which would rule them out as candidates for natural kindhood.

However, it could be questioned whether this is too strong a requirement. Perhaps the less perfectly natural kinds have less well defined categories or properties which may not all be shared alike by every single member of the kind as long as there is enough of a `family resemblance'? That is to say, imperfectly natural kinds such as special science kinds will have fuzzy boundaries. But special science kinds are generally agreed to feature in laws. So, even if mental states cannot be said to have very distinct boundaries (a claim which seems plausible) this doesn't seem to rule out that they can appear in exceptionless laws. In summary, this is not a criterion I will use to judge whether mental and brain states can qualify for natural kindhood.

To summarise, I will not be using Ellis' categorical distinction criteria because although it's not clear that mental states have such distinct boundaries, it's also not clear that a kind need have strict boundaries in order to feature in laws. Neither be I will using the criterion which claims featuring in laws of nature is required for a kind to be natural as to do so would be circular. I will however be using Lewis' conceptions of objective resemblance, perfectly natural and imperfectly natural kinds. This is so I can compare and contrast the naturalness of mental states and their corresponding brain states. At the very least, Ellis' criteria aside (although, again, we have reasons to think that this criteria is overly strong), no criteria for natural kindhood I have presented rules out that mental states can qualify as (at least imperfectly) natural kinds.

Natural Kind Roles in Laws

So, tying together the threads of the previous sections, it seems that causes are determined by reference to the laws of nature which feature natural kinds. And it appears that those perfectly or imperfectly natural kinds which appear in laws of nature are the sorts of things which can be causes.

The important point for me is that kinds, perfectly or imperfectly natural, can be both causes and effects.[20] And only natural kinds can feature in scientific laws, as opposed to arbitrary or gerrymandered kinds which cannot. The group of objects in my living room form a living room kind, but could never feature in a law of nature. Whereas, natural kinds such as electrons can and do. This is because in order to feature in a law there has to be projectability (in Goodman's (1983) sense of the word). Projectability relies on the instances of a kind being relevantly similar to each other such that generalisations can reliably be made about that group. If an entity is a member of a natural kind then the generalisability can hold, unlike with gerrymandered kinds which may have no lawful generalisations which hold true of them.

Before continuing to a deeper discussion of whether mental states can be considered more (imperfectly) natural than their corresponding physical states, there is a potential stumbling block I wish to consider. This is the question of multiple realizability.

[20] I will put the question of what exactly causal relata are to one side for now but I discussed this in section 2.3.

9.2 Mental Causation and Multiple Realizability

In this section I turn to multiple realisability and the relation this has with categorising natural kinds. Briefly, the thesis that the mental is multiply realisable by the physical is the idea that, contra identity theory, there's no one-to-one mapping from the physical to the mental. One mental state can be realised by more than one physical `realiser' or base such as a human brain or a silicon robot brain or by different underlying neural brain states.

Why is this topic of relevance to my argument? If mental states are multiply realisable then it may be argued that they cannot form a genuine kind. This would preclude them featuring in scientific laws which means my argument for them being causally efficacious won't go through. To explore this topic I will now discuss the debate between Kim and Fodor on the topic of multiple realisability and laws. However, before launching into that debate it might help to lay out Kim's views about multiple realizability.

9.2.1 Kim and Multiple Realisability

At this point it will be helpful to focus in on Kim's views in order to contrast them with mine. Kim makes the argument in various places (see for example (1992) and (1993*b*)) that multiple realizability, rather than leading to the demise of reductionism and type identity theory, leads to the opposite conclusion. Indeed Kim claims, if the consequences of multiple realizability are fully considered then it is fully compatible with, if not suggestive of, a form of local reduction.

In "The Myth of Nonreductive Materialism" (1993*b*) Kim starts by examining Putnam's (1975) multiple realizability argument against reductionism. In summary Putnam argues that an identity theorist would need to specify a particular physical brain state such that "*any* organism (not just a mammal) is in pain if and only if (a) it possesses a brain of a suitable physical-chemical structure; and (b) its brain is in that physical-chemical state" ((1975) p.436) and so on for every other type of mental state. Furthermore Putnam states it must also be the case that organisms which are not capable of feeling pain are also not capable of having that physical corresponding brain state. It's vanishing unlikely that this will be possible and we have empirical reasons for thinking it is not that case. Indeed, Kim concedes then that multiple realizability does refute global reductionism. Global or uniform reductionism is "a reduction of every psychological state to a uniform physical-biological base across all actual and possible organisms" ((1993*b*) p.274).

Kim then moves on to discuss Fodor's antireductionist argument in favour of special science autonomy. To reduce a higher level theory to a lower one (at least in a Nagelian sense (see (1979)) then there must exist a series of bridge laws connecting the predicates of the two theories in an appropriate way. If all the predicates of the reducing theory are nomologically coextensive with predicates of the reduced theory then the two theories are what Kim calls "strongly connected" ((1993*b*) p.272). What this means is that there is a biconditional connecting every predicate of the first theory with one of the second which act as bridge laws. These universally quantified biconditionals

would allow the reduced theory to be rewritten in the terms of the reducing theory. Multiple realizability puts this picture in doubt however, as it shows that there are no physical states which can be coextensive with the higher level property. Rather there will be a (potentially infinite) list of physical states which will be nomologically sufficient to bring about the higher level property.

An obvious solution to this issue would be to take the higher level property to be coextensive with the disjunction of physical states. Putnam dismisses this idea as "ad hoc" stating it "does not have to be taken seriously" ((1975) p.437). However, Kim disagrees with Putnam on this point, as his discussion of jade shows. While on the surface jade may be thought of as a natural kind it is not because there are actually two different minerals (with different intrinsic properties) which are both referred to as 'jade'. These two minerals are jadeite and nephrite. Jade therefore, Kim claims, is a genuinely disjunctive property because the underlying microphysical structures of jadeite and nephrite are different enough to classify as different chemical kinds. He argues that 'jade is green' cannot be a law as laws are confirmed by their instances but there are circumstances where 'jade is green' is not confirmed by its instances. We would require instances of jadeite and nephrite to confirm or falsify whether they were all green. However, if it turns out that all our previous jade samples were actually all jadeite then at best we have confirmed that jadeite is green, not that jade is green ((1992) p.12), 'jade is green' has not been confirmed by its instances and therefore is not a law.

In other words, 'jade' is not projectible. Kim says it is projectibility or "this kind

of instance-to-instance accretion of confirmation that is supposed to be the hallmark of lawlikeness" ((1992) p.11). Further it is this lack of projectibility which marks 'jade' out as a "true disjunctive kind" ((1992) p.12). Mental states such as pain should be treated analogously. If the non-nomicity of jade shows that 'jade' is not a kind then shouldn't the non-nomicity of the disjunction of the various physical brain states also show that mental states are non-nomic? This is how Kim argues against Fodor's autonomy of the special sciences.

To return to Putnam, Kim claims that "in rejecting the disjunction move, however, Putnam appears to be assuming this: *a physical state that realizes a mental event is at least nomologically sufficient for it*" ((1993*b*) p.273, emphasis in original). Kim continues though to claim that Putnam actually has stronger laws than this in mind; namely ones which are both necessary and sufficient relative to a biological species. That is any member of species S is such that it is in mental state M if and only if it is also in physical state P. This gives laws of the form "$S_i \rightarrow (M \leftrightarrow P_i)$" ((1993*b*) p.273). Importantly, Kim thinks that biological species may be too broad to fit this bill so he relativises instead to physical-biological structure[21] types.

All Kim now needs to do, he claims, to breathe "new life into psychophysical reductionism" ((1993*b*) p.274) is to claim that "the phenomenon of multiple realization is *consistent* with species (or physical-biological structure) specific strong connectibility" which seems to him to be "plainly true" ((1993*b*) p.274).

[21] A physical-biological structure is a "neural structure that subserves a psychological state or function" ((1993*b*) p.38).

The end result is a series of local reductions. The upshot of multiple realizability in the end then is not that it rules out reductionism, rather it leads to "multiple local reductions" ((1993*b*) p.275).

Lastly, Kim thinks of this not as a form of eliminativism, but rather a form of "mental property irrealism" ((1992) p.26). This is because he does not deny that our mental property concepts (like our concept of jade) have meaning. We can still pick out bits of green stone which we call jade just as we can experience sensations that hurt and label them pain. It is just that there is no 'pain' or 'jade' out there in the world. As Kim puts it there is no "pain as such" ((1992) p.25). What there are though are species specific mental properties (or even more specific physical-structure relative properties). So while there is no 'pain' there is 'human pain' and 'martian pain' and so on. These are just reduced to their underlying physical microstates.

9.2.2 Debate with Fodor

With Kim's views on multiple realizability in mind, I will move onto his debate with Fodor. Jade as we know it is multiply based by jadeite and nephrite, that is 'jade' is not a kind in itself but is rather "a disjunction of two heterogeneous nomic kinds" ((1992) p.12). In "Multiple Realization and the Metaphysics of Reduction" (1992) Kim uses this fact to argue that kinds like jade which are multiply realised cannot feature in laws because they are not projectible. Fodor (1997) disagrees with Kim in his paper "Special Sciences; Still Autonomous After All These Years". Take again the example of a putative law 'jade is green'. As jade is multiply

realised it forms a disjunctive kind of jadeite and nephrite and is in itself not projectible. If Kim is right, this would be bad news for me because he takes mental states to be multiply realisable. And as the jade example shows this would seem to mean that they form a disjunctive kind rather than a natural kind which can feature in laws. This strips them of any causal power.

You could view mental states such as pain as functional kinds as Block and Fodor (1972) do. On Fodor's view, because pain is a functional kind whereas jade is not, pain is multiply realizable where as jade is a mere disjunctive kind. This would mean that pain can still be a kind where jade is not. Fodor makes the distinction between "a multiply based property that is *disjunctive*, and a multiply based property that is disjunctively *realised*" ((1997) p.153). A property is the former (i.e. disjunctive) if all its metaphysically possible realisers are present in the actual world. Jade falls into this category ((1997) p.153). A property is disjunctively realised if not all its metaphysically possible realisers are present in the actual world. If you take a functionalist stance such as Fodor's, mental states such as pain would fall into this category because it can also be realised in ways not present in the actual world for example by silicon or by Martians' brains ((1997) p.154). The functionalist however does not think that just because pain is multiply realised that it is merely disjunctive. Many functionalists would agree that pain states can be homogeneous with respect to their function (either between different entities or within the same entity over different times) even though their realisers are heterogeneous. Therefore, pain is not a mere disjunctive kind and it can (unlike jade) feature in laws of nature. And, they

claim, what is true about pain should apply to other mental states as they too are functional kinds picked out by their functional roles.

9.2.3 Yablo and Multiple Realisability

Stephen Yablo (1992) disagrees with Kim that the physical can never leave causal room for the mental and thinks that there is a case to be made for the mental having genuine causal power. Yablo thinks the CEA is unsound because he doesn't think that mental causation would involve widespread systematic overdetermination. His argument rests on the concept of multiple realisability and the idea of proportionate causal explanation.

Yablo discusses the colour red, which is multiply realised by its different shades. The base necessitates the realised property but not vice versa exactly because more than one base can realise the same property. In other words, there is "asymmetric necessitation" ((1992) p.250) in such cases. Holding supervenience along with this asymmetry leads Yablo to characterise the mental/physical relation as one of determinate/determinable. So, crimson is the determinate of the determinable, red. Yablo defines this Determination Relation between properties as;

> "P determines Q iff: for a thing to be a P is for it to be a Q, not simpliciter, but in some specific way" ((1992) p.252)

where P and Q are properties. In other words;

> "P determines Q (P < Q) only if: (i) necessarily, for all x, if x has P then x

has Q; and (ii) possibly, for some x, x has Q but lacks P" ((1992) p.252).

Yablo constructs an analogy between mental properties and mental events claiming that "we find that the relation between mental and physical events effectively duplicates that of mental to physical properties" ((1992) p.270) which yields this principle;

> "A mental event m occurs iff some physical determination p of m occurs" ((1992) p.271)

where the determination relation for events is;

> "*p* determines *q* iff: for *p* to occur is for *q* to occur, not *simpliciter*, but in a certain way" ((1992) p.260).

For example "Icurus's flying too near the sun determines his flying *per se*." ((1992) p.261). And there exists a world in which Icurus flies but does not do so too close to the sun.

Yablo's idea is to try to hone in on the most proportionate causal explanation for any given effect. "It seems clear" ((1992) p.277) to him that a more proportionate cause will be preferred as an explanation for an event than a less proportionate cause. The most proportionate cause is one which doesn't leave out causally relevant information but at the same time doesn't include too much. This requires that the cause be both contingent and adequate. Being contingent is defined as "If x had not occurred, then y would not have occurred either" ((1992) p.274). Being adequate is defined as "If x had not occurred, then if it

had, y would have occurred as well ((1992) p.274). Being both guarantees that the effect would not occur without the cause and, supposing the effect hadn't occurred, the effect would occur if the cause did.

He gives the death of Socrates by guzzling hemlock as an example ((1992) p.275). To say that his guzzling the hemlock caused his death is to be overly specific. Presumably a drinking event could have occurred even if a guzzling event had not and this would still lead to death by poisoning.[22] This is a therefore a violation of contingency. As a violation of adequacy, Yablo gives the example of a safety valve which, due to a freak malfunction, stops opening at the correct speed. This allows pressure to build which eventually causes the boiler to explode. The opening in itself is not adequate for the effect; the slowing of the door opening is needed to cause the explosion. To see this, think of the scenario in which the malfunction had not happened but the opening had. This is just the scenario in which the valve functions as normal. In that case it is to be hoped that the boiler would not have exploded.

There are two further requirements for proportionality; the cause must be required and enough for the effect, which is to say that the cause shouldn't contain any extraneous causal factors, but that it should include enough to ensure the event. To take Yablo's examples, Socrate's guzzling the hemlock was contingent for his death but was not required because drinking the hemlock would be sufficient. Similarly, the valves opening was adequate for the explosion

[22]Unless you take a more finely grained view of events and think that this would not actually be the same death.

(supposing this time however that this was always a slow moving valve). But its opening per se is not enough, it must open slowly in order for the explosion to occur. See ((1992) pp.276-277).

Yablo introduces Sophie the pigeon who has been trained to peck at red objects. One day, her guardians show her a scarlet object so she pecks at it.

> "Assuming that the scarlet was causally sufficient for the pecking, we can conclude by the exclusion principle that every other property was irrelevant. Apparently then the redness, although it looked to be precisely what Sophie was responding to, makes in reality no causal contribution whatever" ((1992) p.257).

Yablo thinks this result is nonsensical. Of course it is the redness which is causing the pecking, it's the very thing she's been *trained* to peck at. Just as in the hemlock example, the pecking was not contingent on the scarlet, because if scarlet had not been presented, but some other shade of red had, then the pecking still would have happened. Further, the redness seems to be an adequate cause of the effect because, on the presumption that the object was in fact green (and thus remained un-pecked) had the object been red then it would have been pecked. In this case then, the more proportionate cause of the pecking was the objects redness, not its scarletness. If this holds for mental and physical properties, as Yablo claims it does, then the CEA fails to hold. This is because if there is a single, more proportionate mental cause, then there will be no widespread and systematic overdetermination.

Will there be times when it is more proportionate to explain a given action by mental states rather than the underlying physical states? Yablo thinks so, in cases where the effect in question does not depend too sensitively on its physical implementation. He gives as a final example, his ringing a doorbell ((1992) p.278). He hypothesises that there are many physical implementations that can instantiate the decision to ring a doorbell. This means that despite having a specific physical determination, that particular brain state is akin to scarletness in this case. The effect is not contingent on the brain state. Other brain states are possible which would have instantiated the decision and would have caused the doorbell ringing. Without the decision altogether though, it seems as though there would be no bell ringing. So the decision is contingent for the doorbell ringing. So, here we have an example of a mental cause being more proportionate for an effect than a physical one and therefore it should be the preferred explanation of the cause.

To return to why Yablo thinks the CEA is invalid, I take it that he thinks the CEA fails to go through because there will be only one most proportionate explanation for any given event which may be mental. If this is the case then there will not be any widespread systematic overdetermination by mental causes of physical effects despite the fact that there will always be a sufficient physical cause for those physical effects. I'll now consider an objection to his argument.

9.2.4 Wrong Grain Objection

I agree with Yablo's conclusion that mental states can be non-overdeterming causes of physical events. Rather my objection to Yablo lies with his argument for that conclusion. The problem, in my view, is that he compares different grains in the physical and mental cases. The mental state 'pain' is a much more coarse grained mental phenomena than 'specific set of neural firings' is a physical state. There are many ways of being in pain, compare a paper cut to a headache (or even compare a throbbing headache to a searing headache) whereas there's only one way to exemplify a certain set of neural firings. I think it would be more accurate to compare 'pains' to 'sets of brain states which correspond to pains' and 'specific searing headaches' to 'specific sets of neural firings'. If we were to individuate these pains ever more finely, the idea of multiple realisability and asymmetric determination become more questionable. Therefore, while Yablo may be right that the most proportionate explanation is the preferable explanation, he is wrong in which explanations he picks out as being proportionate in that he thinks there's a mental explanation which will be more proportionate than the physical one.[23]

It is time to return to the psychological laws I mentioned earlier in Section 9.1.1. I used the general terms 'sadness', 'anger' and 'stress' in the same way I have been using the term 'pain'. This is as an umbrella term including a vast array of more specific types of sub-emotions. Take 'sadness'. Just as I take it there are

[23]Rather, I will argue, it is the more perfect naturalness of mental states when compared to their corresponding physical state which is the reason to prefer mental states as causes.

various kinds of pains (throbbing, aching, sharp and so on), I take it that there are various kinds of sadness. Examples can include but are not limited to grief, depression, disappointment, loss, mild unhappiness and so on. These seem to me to be phenomenologically distinct just like the different pains are. This is important to note because just as I argue that specific pain states (for example throbbing headache pain) are not multiply realisable or merely disjunctive, so I want to argue that other qualitative experiences of emotions are not. When it comes to the example laws featuring these emotions I gave them at a coarse grain level because this is how Dr Bao gave them to me. And these coarse grain level generalisations do seem to hold in a lawlike way. In this way the pain case is disanalogous to the jade case because there can't be any laws about 'jade' at all.

Different pain states, even when considered on the coarser levels will all have more objective resemblance as a group than will examples of jadeite and nephrite as a group. This resemblance of instances of pain shows that pain constitutes an imperfectly natural kind. The higher the level of specificity of pain state the more objective resemblance there will be in the group (perhaps concluding with perfect naturalness at the highest level of specificity). Generalisations which include the more specific pain kinds will perhaps become lawlike with lower degrees of exceptionlessness as the generality of the group becomes higher. Hence, 'pain' is, unlike 'jade', not a 'merely disjunctive' kind. And this is why, while there cannot be any laws about 'jade', there can be laws about 'pain'.

What about propositional attitudes? As I have stated before, I do not intend to focus on propositional attitudes. I don't need to argue that all mental states are causally efficacious as long as I can show that some mental states are efficacious. Furthermore, as qualitative mental states are, after Chalmers (1996), considered the "hard problem" of consciousness then the propositional attitudes should present a more straightforward case.

If mental states are not multiply realisable by their physical bases then pain, rather than being one perfectly natural kind, is an imperfectly natural kind including many types of (resembling) pain. But which ever grain you chose to examine, there will always be a corresponding physical state or set of physical states.[24] So, the mental states will never be the most proportionate cause or at least there will always be an *equally* proportionate brain state.[25]

This brings us back to the debate between Fodor and Kim with regard to natural kinds in the special sciences and whether these can be related one-to-one with physical natural kinds. If the different pain kinds do correspond one-to-one with neurological kinds then multiple realisability looks less plausible, at least in this case.

Given all these various considerations, I do not find Yablo's argument convincing. While I agree that the most proportionate explanation should be preferred, I do not think he picks out the uniquely most proportionate response in token cases.

[24]This is an empirical hypothesis. Therefore my argument is open to empirical refutation if this turns out not to be the case.

[25]Although, if the mental and the physical present equally proportionate causes then a further argument would be required as to why the physical should be preferred.

Pain may be the most proportionate cause of reaching for the paracetamol *in general* (although that claim could possibly also be had by the appropriately selected group of brain states). But, the brain state will be an equally proportionate cause of any *particular* reaching for the paracetamol. Or at least, there will be a brain state or set of brain states which is *equally* proportionate to the mental state at hand (that is to specific type of pain as opposed to pain in general). The same could be said for laws of the type suggested by Dr Bao such as "experiencing psychological trauma reliably increases your chances of suffering a mental health condition". As phrased, this is a very coarse grained law. The more general mental state 'trauma' will correspond to a set of brain states which, when taken together, are equally proportionate. So, proportionality won't help us pick out which cause is to be philosophically preferred. Rather, I claim, it is the *naturalness* of the mental states which make them more appropriate for causal roles (when opposed to the bundles of physical brain states which correspond to them).

It is plausible that more general mental states (say headaches in general) constitute imperfect natural kinds. Can the same be said for their underlying physical states? It is much less plausible that bundles of physical states which underlie general mental state can form even highly imperfect natural kinds. While different types of headaches resemble each other (a throbbing headache resembles a sharp headache and so on) albeit imperfectly, the same cannot be said for the disjunctive bundle of physical states. Laws which feature such general mental states will be higher level and will not be exceptionless but so

are many special science laws. Thus, I claim, my having a headache is more plausibly a cause, and not merely a more proportionate cause, of my reaching for the paracetamol than my having a disjunctive brain state.

What about the case of more specific mental states (say having a throbbing headache)? This may be a harder case to make since the specific physical brain state itself plausibly constitutes a natural kind. Even if such specific brain states do constitute natural kinds however, such a kind would be imperfect, and I claim, *more* imperfect than the mental kind. I make this claim on the basis that the mental state is simpler than the corresponding brain state (although I postpone full discussion on this point until section 9.3.2). So, a case can be made, based on naturalness, that specific mental states are better candidates to feature in laws and thus play causal roles than specific corresponding brain states.

However, even if you are not convinced that specific mental states are more suited to featuring in laws than their corresponding brain states, an appeal can be made to Yablo's argument from proportionality. Perhaps, it is my having a headache in general (rather than having a throbbing headache, or its underlying brain state, specifically) which causes me to reach for the paracetamol. Even more generally speaking, having pain in general may be the most proportionate cause of my reaching for paracetamol. The higher and more general the mental state the less plausible it is that the set of corresponding brain states form any kind of natural kind. Thus, the stronger the case that it is the mental state which is playing the causal role.

How does this sit with what I've said above about the placebo effect and the effectiveness of talking therapies? I've presented the mental states in these cases to be the most proportionate (in the sense of most natural) cause even in token cases. That is, not only would counselling[26] be the most proportionate response to behaviours caused by depression in general, one particular session of counselling (or set of sessions) would be the most proportionate response to one particular set of depressed behaviours. Yablo's argument would therefore go through. I posit that with no mental causation there would also be no placebo effect and no effective counselling. This is precisely because the phenomenal aspect of these therapies seem to so integral to their effectiveness. It is only because we believe the medicine will work that it works. I'm not sure that without mental causation talking therapies would be efficacious because phenomenal experience without mental causation would be passive.

9.3 Natural Kinds, Laws and Mental Casuation?

Why have I spent time discussing these issues? In other words, why are these topics relevant to the debate on mental causation? The answer is that I believe when pieced together, these topics can help lend weight to the idea that mental states are causally efficacious over and above the brain states which underlie them. As I have said, only natural kinds can feature in scientific laws,

[26]It's important to note that I take counselling to be a case of mental to physical causation because the aim is to intervene directly on mental states to bring about (among other things) a change in behaviour. For example, certain behaviours like excessive alcohol consumption can often be motivated by depression. It is these behaviours that counselling seeks to prevent by intervening on the mental states of the patient.

not arbitrary kinds. If mental states can be classified as natural kinds, and at least sometimes be more natural candidates than the brain state kinds, then mental states can feature in scientific laws. It doesn't always have to be the case that the mental state plays the more suitable natural kind role. It is sufficient that it is sometimes the case.

At first blush it certainly seems as though mental states can be classified as natural kinds and can feature in both causal relations and laws. Take the example of pain. If I am in pain then the mental state I am in belongs to the imperfectly natural kind of pain states.[27] It can be an effect of putting my hand accidentally on the hot oven top and can be the cause of my arm moving away from the oven. And it can take a role in psychological laws such as that 'pain prevents people from repeating the same painful behaviour in the future'.

When put into the context of scientific laws, mental states become a much better candidate for the role of featuring in them than the associated brain states. If they feature in scientific laws (in place of their corresponding brain states) then they must be playing a causal role. At least it would place the burden of proof onto the denier of mental causation to explain why mental states should be better suited to feature in laws if is actually always the underlying brain state which is always doing to the causing. If the brain states which underlie mental states were the only causers (that is, if mental states were never causally efficacious) you would expect them to be featuring in the scientific

[27] Or if I am in a specific kind of pain, say 'throbbing headache pain', the state may even belong to a perfectly natual kind. Again, I will discuss this below.

laws describing and explaining this phenomena. So, that we can put forward an argument for the opposite does carry some weight.

9.3.1 Mental Kinds and Brain Kinds

So, do mental states meet the criteria for natural kindhood? Do brain states also meet them? Which can more plausibly argued to play the natural kind role in scientific laws? I will argue that mental states are at least imperfectly natural and therefore suitable for inclusion in scientific laws. Moreover, although the brain states corresponding to the mental states in question may also be suitable to feature in laws, they may be less suitable than mental states at least in some cases. Therefore, mental states can play a role which the corresponding physical brain states cannot, leaving them room to be causally efficacious.

Whether you view mental states as perfectly or imperfectly natural will depend on your view of the mental in general. For example a panpsychic such as Goff (2017) or Chalmers (2016) would say that at least some mental states are perfectly natural as they are a basic feature of all things in the world. A genuine emergentist might also view mental states as perfectly natural. This is because they think that as soon as a structure (in the world a brain structure) becomes complex enough then it can give rise to a genuinely new fundamental property. A non-reductive physicalist who thinks that the mental is intimately related to, but not identical or reducible to, the physical, may also think that mental states are natural whether perfectly or imperfectly so. Those who view mental states as multiply realisable as Fodor (1974) does for example, might view them as

imperfectly natural. An eliminitavist such as Churchland (1981) will take the opposite approach that mental state talk will eventually be eliminated in favour of a more sophisticated understanding of neuroscience. He would therefore make the claim that mental states do not form a natural kind.

I think it's highly plausible to regard mental states as at least imperfectly natural. And from this I can argue that they can feature in scientific laws. Imperfectly natural kinds can be both causes and effects. Take again Lewis' example of "being metallic" ((1983) p.347) as an example of an imperfectly natural property. The can's being metallic could be the cause of it being placed in the recycling bin or it could be the effect of the can manufacturer wanting its can to be recyclable. Or, the rock sank because it was made of basalt and the rock was made of basalt because a volcano erupted. So long as the brain states (or bundles thereof) which underlie mental states cannot also be classed this way, then we should take mental states to be the things featuring in laws thus paving the way for mental causation.

How do mental states fare in relation to the criteria for natural kindhood given above? Do they share a natural property in common? For example, do all pain mental states share a natural property? It seems like they do if pain can be classified as a natural property, there certainly seems to be a unifying unpleasant quality to pain. But this isn't specific enough. Pain itself is a bundle of different types of sensations. Aching pains, throbbing pains, sharp pains and so on. Each of these subtypes are going to share in whatever common property makes us distinguish them. The pain experiences feel very similar each time we

experience them, though the different types of pain experiences feel distinct from each other. And I think it's highly plausible that not only do other humans feel similar types of pain sensations but that some animals do too (given their similar reactions, behaviours and physiologies). This is especially true of those animals most closely related to us such as chimpanzees and other primates. The same can be said for other kinds of phenomenal mental state such as happiness or sadness (or rather their associated emotional sensations) or about seeing colours.

Do they permit inferences? I would argue that they do. If we know someone is experiencing a throbbing headache then we can infer how they might behave, reaching for asprin or avoiding bright light. Conversely, if they are taking these actions, then we can infer that they are suffering from a headache particularly if one of the actions they take is stating this.

Do they resemble each other? I would argue that again they do. If my counterpart was experiencing throbbing unpleasant sensations in their head then they would be experiencing a throbbing headache. I've ruled out there being anything perfectly natural about general pain states. But, perhaps belonging to the kind `throbbing headache' at the appropriate level of specificity will actually be perfectly natural. Without ruling this possibility out, it is at least the case that Lewis specifies that an imperfectly natural kind is so "in virtue of a close-knit family of genuine universals one or another of which is instantiated" by its members ((1983) p.347). As I'm interested in mental states featuring in laws then they don't *have* to be perfectly natural. Take `general pain' as an

example again. Does this meet Lewis' criteria for imperfectly natural kindhood? Is there a close-knit family of universals which is instantiated by its members? As I've argued throughout this chapter it seems as though pain could well meet this criteria. Therefore, on Lewis' view mental states such as pain can meet the criteria for imperfectly natural kindhood and therefore feature in laws.

It seems then that mental states can at least be viewed as imperfectly natural kinds. What about the corresponding brain states? If they can also be considered to form a kind then they can take the place in the scientific laws which I argue mental states can fill. In analogy to the CEA, if the underlying brain states can take the place of mental states in said laws, then there is no work left for the mental states to do without being overdetermining. On the other hand, if the underlying brain states can't be said to form a natural or imperfectly natural kind then they cannot take mental states place in scientific laws. Mental states would not be overdetermining and the CEA wouldn't go through.

So how do the brain states which correspond to mental states fare in relation to the criteria for kindhood? It is not essential to my argument that they can never plausibly form a kind. All I require is that there are at least some cases where the mental states make a better candidate for kindhood than the corresponding brain states. By 'better candidate' I mean being more perfect natural. Therefore, if a mental state is more perfectly natural than its corresponding brain state then it qualifies as the better candidate for playing the natural kind role and thereby being the cause. So, to be clear, I do not need it to the be the case that brain states never qualify for kindhood. Just that they do not do so in every

case.

It seems to me that in cases such as a throbbing headache that the mental state will be the better candidate for playing a causal role. What underlies such a headache in the brain? A jumble of neurons firing, some signals received from the wider central nervous system. What property could the neurons be sharing with other jumbles of neurons firing at different times in response to different token pains? They're unified by the mental state they underlie more than by anything they share in and of themselves. It would appear that in terms of at least one criteria mental states make more plausible natural kinds than the underlying mental states.

That being said, as my discussion of Yablo above showed, I have rejected multiple realisability. If I am right, there will be a higher degree of resemblance between the physical brain states than would be the case for other thinkers such as Fodor and Yablo who do endorse multiple realizability. At least this will be the case if we're talking about very similar pain types (say `searing headaches', as opposed to `pain' in general). But, what will be true in every case is that there will be a high level of conjunctiveness and complexity in the physical state underlying any given pain at any level of generality. In comparison the mental state is plausibly much more simple than its corresponding physical state. Could this simplicity be a basis to argue that mental states make the more natural kinds than the corresponding brain states and therefore play causal roles?

9.3.2 Simplicity as a Criteria for Natural Kindhood

There is another criteria for natural kindhood which I haven't discussed yet. It merits its own section because I believe it is a very promising path to follow. Furthermore, this criteria could be used to make a stronger argument than I have hitherto been aiming for. All I need for my argument to go through is that there are at least some cases where mental states are more natural (albeit imperfectly so) than their corresponding physical states. If simplicity can be considered a good criterion for natural kinhood, then perhaps a case could be made that at least some mental kinds are perfectly natural. If so, then this would be good reason indeed to think that they (and not their corresponding physical states) feature in laws and play causal roles. Once again, my argument does not rest on a claim this strong, but it is an intriguing avenue to explore. I will begin now by discussing the concept of simplicity.

Fodor claims that "wildly disjunctive" ((1974) p.103) physical states which can underlie special science kinds means that there can be no reduction and that therefore special sciences are autonomous. But there seems to be an issue of an overabundance of *conjunctiveness* in physical states too. Complexity seems to be the heart of the problem for underlying brain states here. They're very conjunctive given that they're groups of neuronal firings and electrical signals and so on not just from the brain but from the whole central nervous system. Mental states of the same grain on the other hand are not conjunctive at least to the same degree. Even setting aside the problem of multiple realizability,

this fact on its own would arguably make them more perfectly (or rather less imperfectly) natural kinds than their corresponding physical brain state jumbles. So mental states are the more plausible candidates for playing the imperfectly natural kind role in psychological laws and therefore must have causal power.

This brings up questions about the role of simplicity in regards to natural kinds. Can simplicity be a criteria for natural kindhood, or more specifically, perfectly natural kindhood? Recall Lewis' idea of objective resemblance. In order to perfectly resemble each other two objects must share in the same perfectly natural property which Lewis takes to be a universal ((1983) p.357). What universals are and how they should be conceived are questions which have received many different answers. Lewis's conception of universals is of "classes of *possibilia*" ((1983) p.344) making it a nominalist approach. Armstrong's conception on the other hand is realist about universals. As Lewis defines them, universals are "repeatable entities, wholly present wherever a particular instantiates it" ((1983) footnote 2, p.343). A unifying feature of our ideas of universals though is that they should be simple.[28] Armstrong speaks of atomic states of affairs:

> "An atomic state of affairs exists if and only if a particular has a property, or a relation holds between two or more particulars. These properties and relations are, of course, universals" ((1997) p.20)

He distinguishes between "atomic states of affairs *strictly so-called*, and atomic states of affairs in a *loose sense*" ((1997) p.19). The former are simple in and of

[28]This seems to be implicit in Lewis' view.

themselves whereas the latter are "ontologically equivalent to conjunctions of simpler, if not simple, states of affairs" ((1997) p.20). Orilia and Swoyer (2020) highlight a distinction sometimes drawn between structured and unstructured properties. So, perhaps an argument can be made that mental properties are simpler in that they are unstructured or at least less structured than their corresponding physical brain states. Furthermore, Grossmann (1983) is another philosopher who argues that all universals are simple. The reason to press this point is that if mental states or properties can be considered as simpler than their corresponding physical brain states then this can further my argument that it's mental states which appear in laws. I think it's plausible that mental states can be considered simpler in this sense.

Why do I think this is plausible? Introspection provides the answer. Often pains, for example, do not feel particularly complex or structured. There may be mental states which are more complex or more structured than others. Certain emotional states, say depression for example, can produce complex, perhaps even confusing, phenomenal experiences. But, what it's important to remember is that simpler here is relative to the associated physical brain states. My introspective experiences lead me to believe that mental states, at least phenomenal mental states, often have a unity and simplicity to them which cannot be rivalled by the associated physical states.

To return to Lewis, if you agree with his view of perfectly natural kindhood as sharing in a universal then perfectly natural kinds should be simple. At the very least they should be simpler than imperfectly natural ones. When comparing

mental states and their corresponding physical neuronal brain states, it appears to me that the mental states are simpler than their physical counterparts. Once again, so long as mental states can be considered simpler than their physical counterparts, then they can be considered to be more perfectly natural and therefore as the more suitable natural kind to feature in psychological laws.

Does this mean that we can consider mental kinds to be perfectly natural per se? That is, can we consider them to be perfectly natural rather than just less imperfectly natural than the physical kinds associated with them?[29] If mental states are individuated at the fine grain level, as I have argued they should be, then they are certainly simple. It's hard to see how they could break down further than the finely individuated phenomenal feeling they are. It may be possible to consider them as perfectly natural kinds in which case they look certain to be better candidates for law roles than the physical kinds underlying them. It should be noted that if you take this view then you will have to endorse a Lewisian view of kinds or at least one similar to it. However, once again, I think my argument still holds even if you view mental properties as only ever being imperfectly natural. Furthermore, it is possible to endorse simplicity as a criteria for natural kindhood without also endorsing the theory of universals.

Again, I don't need to make a claim this strong for my argument to work[30]. After all I just need mental states to make more perfectly natural kinds than their

[29]My thanks to Dr Luke Fenton-Glynn for this suggestion.

[30]One consideration which may speak against mental states being perfectly natural kinds is that they don't seem to function in exceptionless laws.

corresponding physical states at least some of the time.[31]

Therefore, to summarise, a case can be made that mental states, at least some of the time, are the more plausible candidates for being natural kinds and therefore that they're better placed to feature in scientific laws. Therefore, it seems plausible that they are causally efficacious.

9.4 The Natural Kinds Argument and the CEA

How do the arguments I have put forward in this chapter relate back to the CEA? They could be interpreted as questioning either the no-overdetermination premise or the causal closure premise. Alternatively, they could be seens as putting the validity of the CEA into question.

Firstly, the argument from natural kinds may give us reason to reject the no-overdetermination premise. The argument from natural kinds gives us good reason to think that mental states can appear in laws and thus play causal roles. It also gives us good reason to think that mental states can more plausibly be suited to feature in laws than their corresponding physical states. So, even though mental states will always be accompanied by corresponding physical states we have reason to prefer, or view as more proportionate, the mental cause. This could be viewed as suggesting either that no overdetermination is occurring, or that no *bad* overdetermination is occurring. So, this could give us

[31] Reflex movements for example might not be 'mentally caused' despite having an experiential element.

reasons to reject the no overdetermination premise.

Alternatively, on the face of it, an argument which posits mental causation can be viewed as calling the causal closure premise into question. However, recall that the weak causal closure principle states that any physical event has a sufficient physical cause, without ruling out the possibility of it also having a non-physical cause. My argument from natural kinds is consistent with this in that the mental state will be accompanied by a (less natural) corresponding physical state. So it need not violate causal closure. Of course, if you hold a stronger form of causal closure, which rules out any non-physical causation of phsyical effects, then the argument from natural kinds could give you reason to reject this.[32]

Lastly, similarly to Yablo's argument in "Mental Causation" (1992), the argument from natural kinds could be interpreted as questioning the validity of the CEA. While Yablo's argument rests on the idea of proportionality, mine rests on the concept of naturalness. However, both rely on the idea that there may exist a 'better' mental cause while there also exists a sufficient physical cause without there being (bad) overdetermination. I believe a Bennett (2003) style argument can be made that, given the relationship between the mental and the physical, that mental causation of physical effects will not amount to overdetermination, or at least won't amount to 'bad' overdetermination. And, as just mentioned, because the mental natural kind also co-exists with a sufficient physical cause,

[32]Although, as noted in section 8.1.2, holding strong causal closure begs the question of the CEA, which is why Kim rejects it.

it's possible that causal closure can still hold. So, depending on how strong you hold the causal closure principle to be and your views on overdetermination, it's possible to argue that the conclusion of the CEA simply does not follow from its premises.

Furthermore, a benefit of the natural kinds argument when compared to the argument against the principle of causal closure from my previous chapter, is that nothing in it rests on whether the world is deterministic or probabilistic. So, no matter what your commitments are, or, if you think that it's too counterintuitive that so much should rest on whether the world is probabilistic or deterministic, you can still hold the argument from natural kinds.

10

Conclusion

"Everything has an end, and you
get to it if you only keep all on"

The Railway Children - E. Nesbit

(1993)

In this thesis I have set out to show the CEA is not sound in an indeterministic setting and may well not be sound in a deterministic setting either. First, I will recap the two versions of the CEA; Kim's original deterministic version and the probabilistic analogue. I will then briefly recap my main argument and draw together final conclusions including a short discussion on how my work on mental causation relates to causation by other special science properties. Lastly, I will ponder some final unresolved issues which could be areas for future

work.

To recap for the final time, Kim's version of the original deterministic CEA goes as follows;

(P1) Causal Closure of Physics
Every physical event has a sufficient physical cause.

(P2) No Systematic Overdetermination
It is not usually the case that there are multiple minimally sufficient causes of any given event which exist simultaneously.

(P3) Non-Identity of the Physical and the Mental
There are physical states and mental states and these are not identical to each other.

(C) There is no Mental Causation
Mental states cannot be causes of physical effects.

Evaluating the deterministic CEA in the deterministic setting, there are still some reasons to doubt that it is sound. Firstly, there are positive arguments such as the argument from the MMI, the argument from evolution, the argument from inference to the best explanation, the argument from natural kinds and potentially even arguments against overdetermination being problematic. So, though I want to remain somewhat neutral on this issue, we may have reasons to doubt the deterministic version of the CEA is sound even in deterministic settings (assuming that it would be uncharitable and unsuitable to assess the

strength of the deterministic CEA from a probabilistic setting).

However, we have compelling reasons from physics to think that our world is not a deterministic one. Rather, it is plausible that ours is a probabilistic world. Therefore, we need to adapt the CEA to suit this setting. I present this adapted version below as the Probabilistic Analogue CEA'. It is this version which needs deeper analysis because this is the version whose premises seem more likely to be true in our world. I think we have even more reason to doubt it is sound than we did for the deterministic version.

Probabilistic Analogue CEA':

(P1') Probabilistic Causal Closure of Physics
Every physical event has a physical cause which is sufficient to fix its probability.

(P2') No Systematic Overdetermination of Probabilities
It is not usually the case that there are multiple sets of events that are minimally sufficient to fix the probability of a further event which exist simultaneously.

(P3) Non-Identity of the Physical and the Mental
There are physical states and mental states and these are not identical to each other.

(C) There is no Mental Causation
Mental states cannot be causes of physical effects.

Having laid out some groundwork, including introducing key theories of causation and probability I moved on to discuss the CEA. I agreed with (P3) that

the mental and the physical are not identical using Levine's (1983) explanatory gap argument, Nagel's (1974) argument from "What it's like to be a Bat" and Frank Jackson's (1986) Knowledge Argument. I also presented Sider's (2003) and Bennett's (2003) arguments as to why overdetermination is not problematic and showed how they could be extended to the case of the overdetermination of probabilities.

However, my main argument comes from rejecting the first premise of the probabilistic CEA, that is, I reject the causal closure of physics. This is because, in a probabilistic world even once all the physical probabilities have been fixed, there can still be work left for the mental to 'top up' (or indeed perhaps even lower) the probability of a given event occurring. I argue that this picture could be consistently held with a world view incorporating a best system interpretation of probability and a probability raising counterfactual theory of causation. However, I think my arguments can also be made consistently with other theories, for example an interventionist account of causation.

Additionally, even if the reader does not want to reject causal closure I present an argument for the existence of mental causation depending on natural kinds. Even if underlying brain states can be said to form a kind, if their corresponding mental states are more perfectly natural then it is plausible that it is the mental and not the physical states which should play roles in causal laws. Thus mental events can be causally efficacious as a mental event.

In conclusion then, I argue that I have shown the CEA to be ineffective at

proving that the mental can not cause physical effects at least in so far as probabilistic worlds like ours go.

10.1 Special Sciences, Higher Level Properties and Downward Causation

How does my work with mental causation interact with causation by broader special science properties? Indeed Kim notes that mental causation is commonly supposed to be a "special case" of downward causation ((2010) p.38). So, could my arguments in defence of mental causation also apply to special science properties, thus providing an argument for their autonomy? I will very briefly discuss this topic now.

Take as an empirical example the collapse of the wave function on certain orthodox interpretations of quantum mechanics. On such views the wave function collapses on measurement. The wave collapse of a particle is a physical micro-level event. Measurement is a macro-level event. This would appear to be an example of a higher-level property causing a fundamental physical event. If the measurement was an observation made a person as in Schrödinger's (1935) thought experiment this would be a case of mental causation on some interpretations of quantum mechanics.[1]

This is however, a special case. What about those more `run of the mill' cases of

[1]Namely the von Neumann-Wigner interpretation (see Von Neumann (1955) and Wigner (1962)).

putative downward causation?[2] Paoletti and Orilia broadly define downward causation as "causation of lower-level effects by higher-level entities" ((2017) p.1). Kim defines downward causation as when a higher-level property "causes the instantiation of a lower-level property" ((2010) p.28). So, for example, a biological process like respiration can cause effects on the molecules in the air by removing oxygen and increasing the amount of carbon dioxide.

There can also be cases of downward causation in terms of higher-level special science properties bringing about macro-physical effects. For example, the breeding of cattle and their digestion processes create a large amount of methane gas which contributes to climate change. It could also be possible for properties from very high-level social sciences such as economics to have physical effects. For instance, the economic property of value causes transactions and distributions of goods to occur.

Exactly which phenomena will qualify as cases of downward causation depends on how `levels' are understood within a given model. Kim gives the example of a vase with a mass of 1 kilogram. If the vase were to be dropped it would smash into pieces causing "myriads of molecules of all sorts to violently fly away in every direction" ((2010) p.32). If levels of your model were stratified by scientific domain then this may not count as a case of downward causation as both mass and the atoms which constitute the vase fall within the purview of physics. On the other hand it does count as a case of downward causation if

[2]For some defences of downward causation see LePore and Loewer (1987), Baker (1993), Kroedel (2015) and Zhong ((2011), (2012)) among others.

the scale of the entities in question are what stratify the layers within your model. This is because it is the property of the macro-object (the mass of the whole vase) which causes the micro-level entities (the atoms) to be strewn around. In so far as my arguments relate to mental causation specifically, and possibly to higher level special science properties, for example, economic or pyschological properties, I think I can sidestep this issue. This is because, given I have rejected the identity of the mental and the physical, it is plausible that cases of mental causation (or higher special science property causation) will qualify as cases of downward causation however you build your model.[3]

So, how do my arguments pertaining to mental causation apply to wider cases of downward causation? As irreducible downward causation is analogous to mental causation so these higher-level events are analogous to mental events in their ability to bring about lower-level events. They are therefore also potentially susceptible to CEA type arguments, depending on how strictly you define physics. I have included special sciences such as biology into my definition of physics (see Section 8.1.1). Therefore, biological and other such higher-level properties are not excluded by causal closure. However, if you are more restrictive than me in your definition of physics, a CEA type problem may

[3]To clarify Kim's position, he himself is not a proponent of a "single hierarchy of connected levels, from higher to lower, in which every object and phenomenon of the natural world finds its `appropriate place'" ((2002) p.16). Indeed he calls attempts to create such a hierarchy "rather pointless if not hopeless" ((2002) p.16). Likewise, due to exclusion type worries he suggests we don't consider downward causation "real" ((2010) p.40) in the sense of latching onto real ontological levels in the world. Rather he suggests "we may try to salvage downward causation by giving it a *conceptual* interpretation" ((2010) p.40) interpreting talk of levels as representational or levels of description.

arise.[4] If the higher-level and lower-level properties are not identical and the lower-level is causally closed, then there is no overdetermining causation by higher-level properties.

One solution to the problem of downward causation would be to simply identify higher-level properties and events with their lower-level instantiation bases. It's not the cattle digestion which is causing the release of methane which leads to climatic change. It's not even the chemical processes which take place in the cows digestive system which cause the breakdown of food and production of methane. Rather the cause is actually the microphysical events underlying these chemical processes.

However, there are problems with this solution. Block (2003) questions whether this reliance on situating causal powers on lower levels is problematic. Block claims that at best, relying on this solution strips all but the lowest physical level of any causal power. He sees this position as unpalatable (I agree with Block on this point). At worst though, this position could lead to an infinite regress if there is no fundamental or bottom physical level. Block argues that such an infinite regress would result in causation draining completely away. It is of course an "open question from the point of view of the core of contemporary physical theory" ((2003) p.138) as to whether there is a fundamental or bottom level of physics. However, as Block notes, the problem is serious even if there is a bottom level of physics, as this kind of causal `drainage' would still lead to the stripping

[4]Likewise, if the CEA type argument on higher-level properties don't apply then my arguments will not be relevant. But, if you restrict physics then my replies can be used to defend higher-level properties in the same way as mental properties.

of higher-level properties, including mental properties, of their causal power. This is obviously a conclusion I would like to avoid.

Kim's solution to this problem rests on his identifying the higher-level physical properties with their lower-level bases such that the causal draining is not problematic. However, as I have argued already in Chapter 6, we have good reasons to think that the mental is not identical to the physical: Levine's (1983) explanatory gap, Jackson's (1986) knowledge argument and Nagel's (1974) argument in "What it's Like to be a Bat".

The Natural Kinds Argument and Downward Causation

I will now briefly examine if my natural kinds argument from the previous chapter can apply to higher-level special science properties as it can to mental ones. In my previous chapter I argued that if mental properties can constitute kinds then they can be the better candidate than their associated physical properties for playing a role in laws. I based my argument on the idea that mental states are plausibly more perfectly natural than their physical correlates. If they play a role in laws then they are causally efficacious. Can the same be said for more general special science higher-level properties?

I would suggest in fact that the argument is easier to make for the special science properties than for mental properties given that the literature on natural kinds has often focused on special science properties. Special science examples from chemistry and biology are often put forward as exemplars of kinds, those examples being elements and species respectively. To apply this line of

argument to my cattle example, if `digestion' can be considered a natural kind in the same way that `headache' can be[5] then this higher level property is a much better candidate to feature in laws than whatever the very conjunctive physical base would be. And if they feature in laws then they are causally efficacious.

Does what I've said in Chapter 8 about causal closure also apply to the problem of downward causation? I would argue it does at least in so far as exclusion type arguments go. If the world is probabilistic and I am right that causal closure doesn't hold, then the probabilistic CEA doesn't hold. If the causal closure of physics is wrong then that means there is no causal closure of the lowest-level science. This makes way for the higher-level properties to top up or lower the probabilities as set by the lower-level causes. Once again, at the least, this places the burden of proof back on the sceptic to show why the higher-level properties cannot be causally efficacious.

So to conclude this section, which has been a mere sketch, I argue that my arguments apply as much to the wider problem of downward causation as they do to the more special case of mental causation.

[5]While I have used `headache' as an example of a kind here it should be noted that this is actually a group of more precise natural kinds; `searing headache', `throbbing headache' and so on. I am confident the same can be said for many other higher-level properties.

10.2 Unresolved Issues

There are of course many questions my work will raise which will not be answered in this thesis. I will comment on two of them now; how far my argument pushes me from physicalism and the potentially strange implication my argument has for mental causation in deterministic worlds. These could be fruitful avenues for further work.

10.2.1 Departing from Physicalism

Given that I argue that causal closure does not hold in probabilistic worlds, what consequences does this have for physicalism? Or, in other words, how far do we have to depart from physicalism given my arguments? This is an important question because it cuts to the heart of how the mental and the physical interact and therefore how the mechanism behind mental states raising the probability of physical events coming about actually works.

It may be possible that future physics will be able to 'fill in the gaps' and fully explain how the mental and the physical relate without any violation of causal closure. However, the explanatory gap, knowledge argument and so on, all lead me to think that this is not going to be the case.

It appears on the face of it then, that some kinds of property dualism holds. But, where exactly my argument lies in relation to physicalism, and whether I have to depart from it to an objectionable extent, are interesting questions for future

work.

10.2.2 Peculiar Conclusion?

It may be argued that my argument could lead to a strange implication for the existence of mental causation. That is, it seems odd and counterintuitive to say that the existence of mental causation depends on whether the world is deterministic or probabilistic. As I have argued, there is more reason to think that the deterministic CEA holds in deterministic worlds than we have to think that the probabilistic analogue CEA holds in probabilistic worlds. This could lead to the odd conclusion that mental causation exists in probabilistic worlds but not in deterministic worlds. The strangeness of this conclusion might lead someone to question my argument.

I don't think the counterintuitive nature of such an outcome would be reason enough to dismiss my arguments. Partly this is because sometimes things just are counterintuitive. More importantly though, the status of mental causation in deterministic worlds is irrelevant from our point of view given my assumption that we live in a probabilistic world. We have no experience of deterministic worlds and our experience of this world cannot contradict the notion that there is no mental causation in deterministic worlds. It may well be, for all we know, the case that there is no mental causation in deterministic worlds.

What would a `mental life' be like in a deterministic world then? Perhaps there would be no experience of free will such as we have if we examine our inner

mental lives. Perhaps we would have no experience of mental causation. All told, our 'mental lives' could be radically different in a deterministic world. Indeed it's even possible that a deterministic world may be a zombie world. This would render all my arguments null as applied to the deterministic world. The MMI wouldn't hold, or at least it would be empty, there would be no mental phenomenology to accommodate into our theories. However, this is not problematic, because in a philosophical zombie world there is no mental causation to explain in the first place. Ultimately, if our deterministic mental lives would be different from our probabilistic mental lives then it doesn't speak against my arguments. In the end it might have to be the case that I embrace a counterintuitive conclusion.

Of course it would be preferable if I could avoid it, after all, Bohmian Mechanics could be true and therefore my assumption that our world is not deterministic false. Let's say that our world does turn out to be deterministic, what are the consequences? The MMI still holds, we're not philosophical zombies and our mental lives would remain to be explained. This is not to say that I'm back to square one though. Importantly, I argue that it is very possible that mental causation could exist even in deterministic worlds. We still have reasons to think the deterministic CEA is unsound, at least if we suppose that our world is actually deterministic or that MMIs are relevantly similar in deterministic worlds. These include, but are not limited to, the argument from the MMI, the argument from evolution, the argument from inference to the best explanation, the argument from natural kinds and potentially even arguments against overdetermination

being problematic. Further examination on this topic would be an interesting avenue for future work.

BIBLIOGRAPHY

Abrams, M. (2017), Probability and chance in mechanisms, *in* e. Stuart Glennan & P. M. Illari, eds, 'The Routledge handbook of mechanisms and mechanical philosophy', Routledge Handbooks Online, pp. 169–183.

Adams, D. (1995), *The Hitch Hiker's Guide to the Galaxy: A Trilogy in Five Parts*, Random House, UK.

Albert, D. & Loewer, B. (1988), 'Interpreting the many worlds interpretation', *Synthese* **77**(2), 195–213.

Albert, D. Z. (2000), *Time and chance*, Harvard University Press, Cambridge, MA, Cambridge, Mass. ; London.

Alzheimer's disease - Symptoms (2018). Library Catalog: www.nhs.uk Section: conditions.
URL: *https://www.nhs.uk/conditions/alzheimers-disease/symptoms/*

Anscombe, G. E. M. (1971), *Causality and Determination: an Inaugural Lecture*, CUP Archive. Google-Books-ID: RFw4AAAAIAAJ.

Armstrong, D. M. (1997), *A World of States of Affairs*, Cambridge studies in philosophy, University Press, Cambridge.
URL: *http://dx.doi.org/10.1017/CBO9780511583308*

Bach, T. (2012), 'Gender Is a Natural Kind with a Historical Essence *', *Ethics* **122**(2), 231–272.

Baker, L. R. (1993), Metaphysics and mental causation, *in* J. Heil & A. R. Mele, eds, 'Mental causation', University Press, Oxford, pp. 75 – 95.

Baumgartner, M. (2009), 'Interventionist Causal Exclusion and Non-reductive Physicalism', *International Studies in the Philosophy of Science* **23**(2), 161–178.
URL: *http://www.tandfonline.com/doi/abs/10.1080/02698590903006909*

Baumgartner, M. (2010), 'Interventionism and Epiphenomenalism1', *Canadian Journal of Philosophy* **40**(3), 359–383,509. Place: Edmonton.
URL: *http://search.proquest.com/docview/763128502/?pq-origsite=primo*

Beissner, F., Brandau, A., Henke, C., Felden, L., Baumgärtner, U., Treede, R.-D., Oertel, B. G. & Lötsch, J. (2010), 'Quick Discrimination of A(delta) and C Fiber Mediated Pain Based on Three Verbal Descriptors', *PLoS ONE* **5**(9).
URL: *http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2944851/*

Bennett, J. (1987), 'Event Causation: The Counterfactual Analysis', *Philosophical Perspectives* **1**, 367–386.

Bennett, K. (2003), 'Why the Exclusion Problem Seems Intractable, and How, Just Maybe, to Tract It', *Noûs* **37**(3), 471–497.

Bennett, K. (2008), Exclusion Again, *in* J. Hohwy & J. Kallestrup, eds, 'Being Reduced: New Essays on Reduction, Explanation, and Causation', Oxford University Press.

Bird, A. (1998), 'Dispositions and Antidotes', *Philosophical Quarterly* **48**(191), 227–234. Place: Oxford, UK and Boston, USA.

Bird, A. & Tobin, E. (2018), Natural Kinds, *in* E. N. Zalta, ed., 'The Stanford Encyclopedia of Philosophy', spring 2018 edn, Metaphysics Research Lab, Stanford University.

URL: *https://plato.stanford.edu/archives/spr2018/entries/natural-kinds/*

Bishop, R. C. (2006), 'The hidden premiss in the causal argument for physicalism', *Analysis* **66**(1), 44–52.

URL: *http://www.jstor.org/stable/25597700*

Blanchard, T. & Schaffer, J. (2017), Cause without Default, *in* H. Beebee, C. Hitchcock & H. Price, eds, 'Making a Difference: Essays on the Philosophy of Causation', Oxford University Press. ISBN: 9780198746911.

Block, N. (2003), 'Do Causal Powers Drain Away?', *Philosophy and Phenomenological Research* **67**(1), 133–150. Place: Oxford, UK.

Block, N. (2013), *Readings in Philosophy of Psychology, Volume I*, Harvard University Press.

Block, N. J. (2007*a*), *Consciousness, Function, and Representation: Collected Papers.*, A Bradford Book Ser, Cambridge: MIT Press.

Block, N. J. (2007*b*), Troubles with functionalism, *in* 'Consciousness, Function, and Representation: Collected Papers.', A Bradford Book, Cambridge: MIT Press, pp. 63–101.

Block, N. J. & Fodor, J. A. (1972), 'What Psychological States are Not', *The Philosophical Review* **81**(2), 159–181.

Bohm, D. (1952), 'A suggested interpretation of the quantum theory in terms of "hidden" variables. II', *Physical Review* **85**(2), 180–193.

Bradley, D. (2015), *A critical introduction to formal epistemology*, Bloomsbury critical introductions to contemporary epistemology, first edn, Bloomsbury Academic, London, England.

Bub, J. (2017), Quantum Entanglement and Information, *in* E. N. Zalta, ed., 'The Stanford Encyclopedia of Philosophy', spring 2017 edn, Metaphysics Research Lab, Stanford University.
URL: *https://plato.stanford.edu/archives/spr2017/entries/qt-entangle/*

Burns, R. (1788), *Poems, chiefly in the Scottish dialect. By Robert Burns*, Printed for, and sold by Peter Stewart and George Hyde, the west side of Second-Street, the ninth door above Chesnut-Street, Philadelphia.

Butler, J. (1993), *Bodies that matter: on the discursive limits of "sex"*, Routledge, New York ; London.

Butler, J. (1999), *Gender trouble feminism and the subversion of identity*, Ebook Central (Collection), 10th anniversary edition. edn, Routledge, New York ;

London, New York.
URL: *http://dx.doi.org/10.4324/9780203902752*

Campbell, J. (2010), 'Control Variables and Mental Causation', *Proceedings of the Aristotelian Society* **110**, 15–30.

Carlson, A. (2010), 'Gender and Sex: What Are They? Sally Haslanger's Debunking Social Constructivism', *Distinktion: Scandinavian Journal of Social Theory* **1**(1), 61–72.

Carroll, L. (1856), *Alice's Adventures in Wonderland*, Macmillan. Google-Books-ID: Y7sOAAAAIAAJ.

Casati, R. & Varzi, A. (2015), Events, *in* E. N. Zalta, ed., '*The Stanford Encyclopedia of Philosophy*', winter 2015 edn, Metaphysics Research Lab, Stanford University.
URL: *https://plato.stanford.edu/archives/win2015/entries/events/*

Chalmers, D. J. (1996), *The conscious mind: in search of a fundamental theory*, Philosophy of mind series, Oxford University Press, New York.

Chalmers, D. J. (2016), Panpsychism and Panprotopsychism, *in* G. Brüntrup & L. Jaskolla, eds, 'Panpsychism: Contemporary Perspectives', Oxford University Press. ISBN: 9780199359943.

Chisholm, R. M. (1954), 'Law Statements and Counterfactual Inference', *Analysis (Oxford)* **15**(5), 97–105. Place: Oxford.

Choi, S. & Fara, M. (2018), Dispositions, *in* E. N. Zalta, ed., 'The Stanford Ency-

clopedia of Philosophy', fall 2018 edn, Metaphysics Research Lab, Stanford University.
URL: *https://plato.stanford.edu/archives/fall2018/entries/dispositions/*

Churchland, P. M. (1981), 'Eliminative Materialism and the Propositional Attitudes', *The Journal of Philosophy* **78**(2), 67–90.
URL: *http://www.jstor.org/stable/2025900*

Cohen, J. & Callender, C. (2009), 'A better best system account of lawhood', *Philosophical Studies* **145**(1), 1–34.
URL: *http://dx.doi.org/10.1007/s11098-009-9389-3*

Cohen, J. & Hansel, M. (1956), *Risk and gambling: the study of subjective probability*, Longmans, Green, London.

Cottingham, W. N. & Greenwood, D. A. (2007), *An Introduction to the Standard Model of Particle Physics*, 2 edn, Cambridge University Press, Cambridge.
URL: *http://ebooks.cambridge.org/ref/id/CBO9780511791406*

Courage, K. H. (2013), 'How the Freaky Octopus Can Help Us Understand the Human Brain'.
URL: *https://www.wired.com/2013/10/how-the-freaky-octopus-can-help-us-understand-the-human-brain/*

Crane, T. (1995), 'The mental Causation debate', *Proceedings of the Aristotelian Society, Supplementary Volumes* **69**, 211–253.
URL: *https://www.jstor.org/stable/4107076*

Curtis, V., Aunger, R. & Rabie, T. (2004), 'Evidence that disgust evolved to protect from risk of disease.', *Proceedings of the Royal Society B: Biological Sciences* **271**(Suppl 4), S131–S133.
URL: *http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1810028/*

Davidson, D. (2001*a*), *Essays on Actions and Events*, Oxford University Press.

Davidson, D. (2001*b*), Mental Events, *in* '*Essays on Actions and Events*', Oxford University Press, pp. 207–228.

Davidson, D. (2001*c*), Psychology as Philosophy, *in* 'Essays on Actions and Events', Oxford University Press. ISBN: 9780199246274.

Davidson, D. (2005), *Thinking Causes*, Oxford University Press, pp. 185 – 200. ISBN: 9780198237570.

DeRubeis, R. J., Siegle, G. J. & Hollon, S. D. (2008), 'Cognitive therapy vs. medications for depression: Treatment outcomes and neural mechanisms', *Nature reviews. Neuroscience* **9**(10), 788–796.
URL: *http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2748674/*

Dowe, P. (1995), 'Causality and Conserved Quantities: A Reply to Salmon', *Philosophy of Science* **62**(2), 321–333.
URL: *http://www.jstor.org/stable/188437*

Dowe, P. (2000), *Physical Causation*, Cambridge studies in probability, induction, and decision theory, University Press, Cambridge.
URL: *http://dx.doi.org/10.1017/CBO9780511570650*

Dowe, P. (2001), `A Counterfactual Theory of Prevention and `Causation' by Omission', *Australasian Journal of Philosophy* **79**(2), 216–226.
URL: *http://www-tandfonline-com.libproxy.ucl.ac.uk/doi/abs/10.1080/713659223*

Dowe, P. (2008), Causal Processes, *in* E. N. Zalta, ed., `*The Stanford Encyclopedia of Philosophy*', fall 2008 edn, Metaphysics Research Lab, Stanford University.
URL: *https://plato.stanford.edu/archives/fall2008/entries/causation-process/*

Dupré, J. (1984), `Probabilistic Causality Emancipated', *Midwest Studies In Philosophy* **9**(1), 169–175.

Dupré, J. (1993), *The disorder of things: metaphysical foundations of the disunity of science*, Harvard University Press, Cambridge, Mass ; London.

Eagle, A. (2004), `Twenty-One Arguments against Propensity Analyses of Probability', *Erkenntnis* **60**(3), 371–416. Place: Dordrecht.

Eagle, A. (2016), Chance versus Randomness, *in* E. N. Zalta, ed., `The Stanford Encyclopedia of Philosophy', fall 2016 edn, Metaphysics Research Lab, Stanford University.
URL: *http://plato.stanford.edu/archives/fall2016/entries/chance-randomness/*

Earman, J. (1986), *A primer on determinism*, number v. 32 *in* `University of Western Ontario series in philosophy of science', D. Reidel Pub. Co. ; Sold and distributed in the U.S.A. and Canada by Kluwer Academic, Dordrecht ; Boston : Norwell, MA, U.S.A.

Earman, J. & Norton, J. (1987), `What Price Spacetime Substantivalism? The Hole

Story', *The British Journal for the Philosophy of Science* **38**(4), 515–525.

Edgington, D. (1997), 'Mellor on Chance and Causation (Book Review)', *The British Journal for the Philosophy of Science* **48**(3), 411–433.

Elder, C. L. (2008), 'Biological Species Are Natural Kinds', *Southern Journal of Philosophy* **46**(3), 339–362. Place: Oxford, UK.

Eliot, T. S. (2010), *The Waste Land and Other Poems*, Broadview Press. Google-Books-ID: tUVrhzxW3xIC.

Ellis, B. (2001), *Scientific essentialism*, Cambridge studies in philosophy, University Press, Cambridge.

Ereshefsky, M. (2016), Natural kinds in biology, *in* 'Routledge Encyclopedia of Philosophy', 1 edn, Routledge, London.
URL: *https://www.rep.routledge.com/articles/thematic/natural-kinds-in-biology/v-1*

Everett, H. (1973), *The many-worlds interpretation of quantum mechanics: a fundamental exposition; with papers by J. A. Wheeler (and four others) edited by Bryce S. DeWitt and Neill Graham*, Princeton legacy library, Princeton University Press, Princeton, New Jersey.

Fair, D. (1979), 'Causation and the Flow of Energy', *Erkenntnis* **14**(3), 219–250.
URL: *http://www.jstor.org/stable/20010665*

Fenton-Glynn, L. (2009), A Probabilitic Analysis of Causation, DPhil, University of Oxford, Oxford.

Fenton-Glynn, L. (2017), 'A Proposed Probabilistic Extension of the Halpern and Pearl Definition of 'Actual Cause'', *The British journal for the philosophy of science* **68**(4), 1061–1124.

Fenton-Glynn, L. (M.S.*a*), Causation. Unpublished Manuscript.

Fenton-Glynn, L. (M.S.*b*), Probabilistic actual causation. Unpublished Manuscript.

Fleming, G. R. & Scholes, G. D. (2014), 'Quantum biology: introduction'. ISBN: 9781107010802.

Flowers, P., Theopold, K. & Langley, R. (2016), 'Quantum-Mechanical Tunneling', https://chem.libretexts.org/Courses/University_of_California_Davis/UCD_Chem_107B%3A_Physical_Chemistry_for_Life_Scientists/Chapters/4%3A_Quantum_Theory/4.09%3A_Quantum-Mechanical_Tunneling.

FN1 Gene - GeneCards | FINC Protein | FINC Antibody (n.d.).
URL: *https://www.genecards.org/cgi-bin/carddisp.pl?gene=FN1*

Fodor, J. (1974), 'Special sciences (or: The disunity of science as a working hypothesis)', *Synthese* **28**(2), 97–115.

Fodor, J. (1989), 'Making Mind Matter More', *Philosophical Topics* **17**(1), 59.

Fodor, J. (1997), 'Special Sciences: Still Autonomous after All these Years', *Noûs* **31**, 149–163.

Fodor, J. A. (1991), 'You Can Fool Some of The People All of The Time, Every-

thing Else Being Equal; Hedged Laws and Psychological Explanations', *Mind* **C**(397), 19–34.

Foucault, M. (1991), *Discipline and punish: the birth of the prison*, Penguin social sciences, reprint edn, Penguin Books, London. OCLC: 45870209.

Foucault, M. (2010), *The birth of the clinic: an archaeology of medical perception*, Routledge classics, 1st reprinted edn, Routledge, London. OCLC: 845118786.

Foucault, M. & Khalfa, J. (2006), *History of madness*, Routledge, New York.

Fraassen, B. C. V. (1989), *Laws and Symmetry*, Oxford University Press.

Frankfurt, H. G. (1969), 'Alternate Possibilities and Moral Responsibility', *The Journal of Philosophy* **66**(23), 829–839.

Frigg, R. & Hoefer, C. (2015), 'The Best Humean System for Statistical Mechanics', *Erkenntnis* **80**(Supplement 3), 551–574. Place: Dordrecht.

Gary E. Bowman (2008), *Essential quantum mechanics / Gary E. Bowman.*, Oxford scholarship online, University Press, Oxford.
URL: *http://dx.doi.org/10.1093/acprof:oso/9780199228928.001.0001*

Gibbon, L. G. (2001), *Smeddum: a Lewis Grassic Gibbon anthology (1933)*, number 97 *in* 'Canongate classics', Canongate Books, Edinburgh. OCLC: ocm42274394.

Gillies, D. (2000), 'Varieties of Propensity', *The British Journal for the Philosophy of Science* **51**(4), 807–835.

Glennan, S. (1996), 'Mechanisms and the nature of causation', *Erkenntnis* **44**(1), 49–71. Place: Dordrecht.

Glennan, S. (2011), Singular and general causal relations: A mechanist perspective, *in* P. M. Illari, F. Russo & J. Williamson, eds, 'Causality in the Sciences', Oxford University Press.
URL: *https://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780199574131.001.0*
9780199574131-chapter-37

Glennan, S. (2017), *The new mechanical philosophy*, Oxford scholarship online, first edition. edn, University Press, Oxford.
URL: *http://dx.doi.org/10.1093/oso/9780198779711.001.0001*

Glennan, S. & Illari, P., eds (2017), *The Routledge handbook of mechanisms and mechanical philosophy*, Routledge handbooks in philosophy, first edition. edn, Routledge, New York.
URL: *https://www.routledgehandbooks.com/doi/10.4324/9781315731544*

Godfrey-Smith, P. (2014), *Philosophy of biology*, Princeton foundations of contemporary philosophy, University Press, Princeton.
URL: *https://www.jstor.org/stable/10.2307/j.ctt5hhnq6*

Goff, P. (2017), *Consciousness and Fundamental Reality.*, Philosophy of Mind Ser, Oxford: Oxford University Press USA - OSO.

Goldacre, B. (2009), *Bad science*, fourth estate paperback edition. edn, Fourth Estate, London.

Goldszmidt, M & Pearl, J (1992), Rank-based systems: A simple approach to belief revision, belief update, and reasoning about evidence and actions, *in* W. S. Bernhard Nebel, Charles Rich, ed., `Principles of knowledge representation and reasoning: proceedings of the Third International Conference on Principles of Knowledge Representation and Reasoning', Morgan Kaufmann series in representation and reasoning, Morgan Kaufmann, San Mateo, Calif., pp. 661 – 672.

Goodman, N. (1983), *Fact, Fiction and Forecast*, 4th ed. edn, Harvard University Press, Cambridge, Mass. ; London, Cambridge, Mass.

Goodman, N. (2000), The New Riddle of Induction, *in* S. Bernecker & F. I. Dretske, eds, `Knowledge: Readings in Contemporary Epistemology', OUP Oxford.

Grossmann, R. (1983), *The Categorical Structure of the World*, Indiana University Press.

Hacking, I. (1983), *Representing and Intervening: Introductory Topics in the Philosophy of Natural Science*, University Press, Cambridge.
URL: *http://dx.doi.org/10.1017/CBO9780511814563*

Hacking, I. (1999), *The social construction of what?*, Harvard University Press, Cambridge, Mass. ; London.

Hájek, A. (2012), Interpretations of Probability, *in* E. N. Zalta, ed., `The Stanford

Encyclopedia of Philosophy', winter 2012 edn, Metaphysics Research Lab, Stanford University.
URL: *http://plato.stanford.edu/archives/win2012/entries/probability-interpret/*

Hall, N. (1994), `Correcting The Guide to Objective Chance', *Mind* **103**(412), 505–517.

Hall, N. (2004), Two Concepts of Causation, *in* J. Collins, N. Hall & L. Paul, eds, `Causation and Counterfactuals', The MIT Press, pp. 225–276.

Halpern, J. Y. (2016), *Actual causality*, The MIT Press, Cambridge, Massachusetts.
URL: *http://cognet.mit.edu/book/actual-causality*

Halpern, J. Y. & Hitchcock, C. (2011), Actual causation and the art of modeling, *in* R. Dechter, H. Geffner & J. Halpern, eds, `Heuristics, Probability and Causality: A Tribute to Judea Pearl', College Publications.
URL: *http://arxiv.org/abs/1106.2652*

Halpern, J. Y. & Pearl, J. (2011), `Causes and explanations: A structural-model approach'.
URL: *https://escholarship.org/uc/item/3p69p8wk*

Handler, D. (2005), *A Series of Unfortunate Events; The Penultimate Peril*, Egmont UK Limited.

Harré, R. (2002), *Cognitive science: a philosophical introduction*, SAGE Publications, London ; Thousand Oaks, Calif.
URL: *https://ebookcentral.proquest.com/lib/ucl/detail.action?docID=334339*

Hart, H. L. A. & Honoré, T. (1985), *Causation in the law*, Oxford scholarship online Y, 2nd edn, Clarendon, Oxford.
URL: *http://dx.doi.org/10.1093/acprof:oso/9780198254744.001.0001*

Haslanger, S. (2012*a*), 'Gender and Race: (What) Are They? (What) Do We Want Them to Be?'.

Haslanger, S. (2012*b*), 'Social Construction: The "Debunking" Project'.

Haslanger, S. A. (2013), *Resisting reality: social construction and social critique*, Oxford scholarship online, University Press, Oxford.
URL: *http://dx.doi.org/10.1093/acprof:oso/9780199892631.001.0001*

Heisenberg, W. (1927), 'Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik', *Zeitschrift fur Physik* **43**, 172–198.
URL: *http://adsabs.harvard.edu/abs/1927ZPhy...43..172H*

Hempel, C. G. (1965), *Aspects of scientific explanation: and other essays in the philosophy of science*, Free Press ; Collier Macmillan, New York : London.

Hitchcock, C. (1996), 'The mechanist and the snail', *Philosophical Studies* **84**(1). Place: Dordrecht.

Hitchcock, C. (2001*a*), 'The Intransitivity of Causation Revealed in Equations and Graphs', *The Journal of Philosophy* **98**(6), 273–299.

Hitchcock, C. (2001*b*), 'A Tale of Two Effects', *The Philosophical Review* **110**(3), 361–396.
URL: *http://www.jstor.org/stable/2693649*

Hitchcock, C. (2004*a*), 'Causal Processes and Interactions: What Are They and What Are They Good For?', *Philosophy of Science* **71**(5), 932–941.

Hitchcock, C. (2004*b*), Do All and Only Causes Raise the Probabilities of Effects?, *in* J. Collins, E. J. Hall & L. A. Paul, eds, 'Causation and Counterfactuals', MIT Press.

Hitchcock, C. (2007), 'Prevention, preemption, and the principle of sufficient reason', *Philosophical Review* **116**(4), 495–532.

Hitchcock, C. (2009), 'Problems for the Conserved Quantity Theory: Counterexamples, Circularity, and Redundancy', *The Monist* **92**(1), 72–93.

Hitchcock, C. (2012), Probabilistic Causation, *in* E. N. Zalta, ed., 'The Stanford Encyclopedia of Philosophy', winter 2012 edn, Metaphysics Research Lab, Stanford University.
URL: *http://plato.stanford.edu/archives/win2012/entries/causation-probabilistic/*

Hitchcock, C. (2019), Causal Models, *in* E. N. Zalta, ed., 'The Stanford Encyclopedia of Philosophy', summer 2019 edn, Metaphysics Research Lab, Stanford University.
URL: *https://plato.stanford.edu/archives/sum2019/entries/causal-models/*

Hitchcock, C. & Knobe, J. (2009), 'Cause and Norm', *The Journal of Philosophy* **106**(11), 587–612.

Hitchcock, C. R. (1995), 'Salmon on Explanatory Relevance', *Philosophy of*

Science **62**(2), 304–320.

Hitchcock, C. & Woodward, J. (2003), 'Explanatory Generalizations, Part I: A Counterfactual Account', *Noûs* **37**(1), 1–24. Place: Boston, USA and Oxford, UK.

Hájek, A. (1996), '"Mises redux" — Redux: Fifteen arguments against finite frequentism', *Erkenntnis* **45**(2), 209–227. Place: Dordrecht.

Hájek, A. (2007), 'The reference class problem is your problem too', **156**(3), 563–585. Dordrecht, Netherlands, Springer Nature B.V.
URL: *http://search.proquest.com/docview/196666793/abstract/CCCFD49164A34711PQ/1*

Hájek, A. (2009), 'Fifteen Arguments Against Hypothetical Frequentism', *Erkenntnis* **70**(2), 211–235. Dordrecht, Netherlands.

Hölldobler, B. & Wilson, E. O. (2011), *The leafcutter ants: civilization by instinct*, Norton, New York.

Hoefer, C. (2016), Causal Determinism, *in* E. N. Zalta, ed., 'The Stanford Encyclopedia of Philosophy', spring 2016 edn, Metaphysics Research Lab, Stanford University.
URL: *http://plato.stanford.edu/archives/spr2016/entries/determinism-causal/*

Honderich, T. (1982), 'The Argument for Anomalous Monism', *Analysis* **42**(1), 59–64.

Hoover, T. (2018), 'Oobleck', *Science Scope; Washington* **41**(9), 14–16. Washing-

ton, United States, National Science Teachers Association.
URL: *http://search.proquest.com/docview/2062656318/abstract/7712D7CF1E6B4A6FPQ*

Howson, C. (1984), 'Probabilities, propensities, and chances', *Erkenntnis* **21**(3), 279–293. Place: Dordrecht.

Hume, D. (1988), *An enquiry concerning human understanding (1748)*, Prometheus Books, Buffalo, N.Y. OCLC: 19868862.

Hume, D. & Norton, D. F. (2009), *A treatise of human nature (1739)*, Oxford Philosophical Texts, Oxford University Press, Oxford. OCLC: 934530118.

Humphreys, P. (1985), 'Why Propensities Cannot be Probabilities', *The Philosophical Review* **94**(4), 557–570.

Hunt, G. R. & Gray, R. D. (2003), 'Diversification and cumulative evolution in New Caledonian crow tool manufacture', *Proceedings of the Royal Society of London B: Biological Sciences* **270**(1517), 867–874.
URL: *http://rspb.royalsocietypublishing.org/content/270/1517/867*

Illari, P. K. & Williamson, J. (2011), Mechanisms are real and local, *in* P. Illari, F. Russo & J. Williamson, eds, 'Causality in the sciences', Oxford University Press.

Ioannidis, S. & Psillos, S. (2017), Mechanisms, counterfactuals, and laws, *in* e. Stuart Glennan & P. M. Illari, eds, 'The Routledge handbook of mechanisms and mechanical philosophy', Routledge Handbooks Online.
URL: *https://www-routledgehandbooks-com.libproxy.ucl.ac.uk/doi/10.4324/9781315731*

Ismael, J. (2008), 'Raid! Dissolving the Big, Bad Bug', *Noûs (Bloomington, Indiana)* **42**(2), 292–307. Place: Oxford.

Jackson, F. (1986), 'What Mary Didn't Know', *The Journal of Philosophy* **83**(5), 291–295.

Jeffreys, H. (1948), *Theory of probability*, International series of monographs on physics, second edition edn, Clarendon Press, Oxford.

John Venn (1876), *The logic of chance: an essay on the foundations and province of the theory of probability, with especial reference to its logical bearings and its application to moral and social science.*, 2nd ed., re-written and greatly enl. edn, Macmillan, London.
URL: *http://content.apa.org/books/2008-08461-000*

Johnson, W. E. (1921), *Logic*, University Press, Cambridge.

Johnston, M. (1992), 'How to speak of the colors', *Philosophical Studies* **68**(3), 221–263. Place: Dordrecht.

Jones, K. (2008), 'The Causal Closure of Physics: An Explanation and Critique', *World Futures* **64**(3), 179–186.

Joseph Bertrand (1888), *Calcul des probabilités / par J. Bertrand.*, Gauthier-Villars, Paris.

Joseph Y Halpern & Judea Pearl (2005), 'Causes and Explanations: A Structural-Model Approach. Part I: Causes', *The British journal for the philosophy of science* **56**(4), 843–887. Place: Oxford.

Kam-Hansen, S., Jakubowski, M., Kelley, J. M., Kirsch, I., Hoaglin, D. C., Kaptchuk, T. J. & Burstein, R. (2014), 'Altered placebo and drug labeling changes the outcome of episodic migraine attacks', *Science Translational Medicine* **6**(218), 218ra5.

Kaptchuk, T. J., Friedlander, E., Kelley, J. M., Sanchez, M. N., Kokkotou, E., Singer, J. P., Kowalczykowski, M., Miller, F. G., Kirsch, I. & Lembo, A. J. (2010), 'Placebos without Deception: A Randomized Controlled Trial in Irritable Bowel Syndrome', *PLOS ONE* **5**(12), e15591.
URL: *http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0015591*

Keynes, J. M. (1921), *A treatise on probability*, Macmillan and Co, Ltd, Macmillan, London.

Kim, J. (1966), 'On the Psycho-Physical Identity Theory', *American Philosophical Quarterly* **3**(3), 227–235.

Kim, J. (1984), 'Epiphenomenal and Supervenient Causation', *Midwest studies in philosophy* **9**(1), 257–270. Place: Oxford, UK.

Kim, J. (1992), 'Multiple Realization and the Metaphysics of Reduction', *Philosophy and Phenomenological Research* **52**(1), 1–26.

Kim, J. (1993*a*), Mechanism, purpose, and explanatory exclusion, *in* E. Sosa, ed., 'Supervenience and Mind: Selected Philosophical Essays', Cambridge Studies in Philosophy, Cambridge University Press, p. 237–264.

Kim, J. (1993*b*), The myth of nonreductive materialism, *in* E. Sosa, ed., 'Super-

venience and Mind: Selected Philosophical Essays', Cambridge Studies in Philosophy, Cambridge University Press, p. 265–284.

Kim, J. (1993*c*), Supervenience and mind, *in* E. Sosa, ed., `Supervenience and mind: selected philosophical essays', Cambridge studies in philosophy, Cambridge University Press, Cambridge.

Kim, J. (1998), *Mind in a physical world: an essay on the mind-body problem and mental causation*, Representation and mind, MIT Press, Cambridge, Mass.

Kim, J. (2002), `The Layered Model: Metaphysical Considerations', *Philosophical Explorations* **5**(1), 2–20.

Kim, J. (2008), *Physicalism, or something near enough*, Princeton monographs in philosophy, 3rd print., and 1st paperback print edn, Princeton Univ. Press, Princeton. OCLC: 611751299.

Kim, J. (2010), Making Sense of Emergence, *in* `Essays in the Metaphysics of Mind', Oxford University Press. ISBN: 9780199585878.

Kitcher, P. (1984), `Species', *Philosophy of Science* **51**(2), 308–333.

Knobe, J. & Fraser, B. (2008), Causal Judgment and Moral Judgment: Two Experiments, *in* W. Sinnott-Armstrong, ed., `Moral Psychology', MIT Press.

Kolmogorov, A. N. (1956), *Foundations of the theory of probability / translation edited by Nathan Morrison ; with an added bibliography by A. T. Bharucha-Reid.*, 2nd english edn, Chelsea Publishing Co, New York.

Kripke, S. A. (1980), *Naming and necessity*, Library of philosophy and logic Y, Blackwell, Oxford.

Kroedel, T. (2008), 'Mental causation as multiple causation', *Philosophical Studies* **139**(1), 125–143.

Kroedel, T. (2015), 'A simple argument for downward causation', *Synthese* **192**(3), 841–858. Dordrecht, Netherlands.

Kvart, I. (2004), Causation: Probabilistic and Counterfactual Analyses, *in* N. Hall, L. Paul & J. Collins, eds, '*Causation and Counterfactuals*', MIT Press, Cambridge, Mass., pp. 359–387.

Laplace, P. S. (1902), *A Philosophical Essay on Probabilities*, Project Gutenberg. **URL:** *http://www.gutenberg.org/ebooks/58881*

Le Pore, E. & Loewer, B. (1987), 'Mind Matters', *The Journal of Philosophy* **84**(11), 630–642.

Lemmon, E. (1996), Comments on D. Davidson's "The Logical Form of Action Sentences", *in* N. Rescher, ed., '*The Logic of Decision and Action*', University of Pittsburgh Press, Pittsburgh, pp. 96–103.

Levine, J. (1983), 'Materialism and Qualia: The Explanatory Gap', *Pacific Philosophical Quarterly* **64**(October), 354–61.

Lewis, D. (1966), 'An Argument for the Identity Theory', *The Journal of Philosophy* **63**(1), 17–25.

Lewis, D. (1973), `Causation', *The Journal of Philosophy* **70**(17), 556–567.

Lewis, D. (1983), `New work for a theory of universals', *Australasian Journal of Philosophy* **61**(4), 343–377.
URL: *http://www.tandfonline.com/doi/abs/10.1080/00048408312341131*

Lewis, D. (1986*a*), *On the Plurality of Worlds*, Basil Blackwell, Oxford.

Lewis, D. (1986*b*), *Postscripts to 'Causation'*, Oxford University Press.

Lewis, D. (1987*a*), *Philosophical Papers Volume II*, Oxford University Press.

Lewis, D. (1987*b*), A Subjectivist's Guide to Objective Chance, *in* `Philosophical Papers Volume II', Oxford University Press, pp. 83 – 132. ISBN: 9780195036466.

Lewis, D. (1994), `Humean Supervenience Debugged', *Mind* **103**(412), 473–490.
URL: *http://www.jstor.org/stable/2254396*

Lewis, D. (1997), `Finkish dispositions', **47**(187), 143–158.
URL: *https://www.jstor.org/stable/2956325*

Lewis, D. (1999), Psychophysical and theoretical identifications, *in* `Papers in Metaphysics and Epistemology', Cambridge University Press, pp. 248 – 261. ISBN: 9780521582483.

Lewis, D. (2000), `Causation as Influence', *The Journal of Philosophy* **97**(4), 182–197.

Lewis, D. (2001), *Counterfactuals*, (rev. ed.) edn, Blackwell Publishers, Malden, Mass. ; Oxford.

Lewis, D. (2004), Void and object, *in* J. Collins, N. Hall & L. A. Paul, eds, 'Causation and Counterfactuals', MIT Press, pp. 277–290.

Locke, J. (1970), *An essay concerning human understanding / Facsimile reprint of edition published by Thomas Basset, London, 1690*, Scolar Press, Menston.

Loewer, B. (2001), 'Determinism and Chance', *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics* **32**(4), 609–620.
URL: *http://www.sciencedirect.com/science/article/pii/S1355219801000284*

Marais, A., Adams, B., Ringsmuth, A. K., Ferretti, M., Gruber, J. M., Hendrikx, R., Schuld, M., Smith, S. L., Sinayskiy, I., Krüger, T. P. J., Petruccione, F. & Van Grondelle, R. (2018), 'The future of quantum biology', *Journal of the Royal Society, Interface* **15**(148).

Marcus, E. (2005), 'Mental causation in a physical world', *Philosophical Studies* **122**(1), 27–50.
URL: *https://link.springer.com/article/10.1007/s11098-005-2204-x*

Martin, C. B. (1994), 'Dispositions and Conditionals', *The Philosophical Quarterly (1950-)* **44**(174), 1–8.

Mathers, E. P. (1934), *Cain's Jawbone*, Unbound, London (2019 ed).
URL: *https://unbound.com/books/cains-jawbone-paperback/*

Maxwell, J. C. (1863), 'A Dynamical Theory of the Electromagnetic Field', *Pro-*

ceedings of the Royal Society of London **13**, 531–536.
URL: *http://www.jstor.org/stable/112081*

McGinn, C. (1989), *Mental content*, Basil Blackwell, Oxford.

Menzies, P. (1989), 'Probabilistic Causation and Causal Processes: A Critique of Lewis', *Philosophy of Science* **56**(4), 642–663.

Mill, J. S. (1846), *A system of logic, ratiocinative and inductive: being a connected view of the principles of evidence and the methods of scientific investigation*, Harper, New York.
URL: *http://content.apa.org/books/2008-12811-000*

Mises, R. V. (1939), *Probability, statistics and truth; translated by J. Neyman, D. Scholl, and E. Rabinowitsch.*, (2nd ed.) edn, Hodge, London.

Montero, B. (2003), Varieties of Causal Closure, *in* S. Walter & H.-D. Heckmann, eds, 'Physicalism and Mental Causation', Imprint Academic, pp. 173–187.

Montero, B. G. & Papineau, D. (2016), Naturalism and Physicalism, *in* K. J. Clark, ed., 'The Blackwell Companion to Naturalism', Wiley Blackwell, p. 182–195. ISBN: 9781118657775.

Nagel, E. (1979), *The structure of science: problems in the logic of scientific explanation*, 2nd ed edn, Routledge & Kegan Paul, Hackett PubCo, London, Indianapolis.

Nagel, T. (1974), 'What is It Like to Be a Bat?', *Philosophical Review* **83**(October), 435–50.

Nesbit, E. (1993), *The Railway Children (1905)*, Wordsworth classics, Ware, Hertfordshire : Wordsworth Editions, Ware, Hertfordshire.

Norton, J. (2008), 'The Dome: An Unexpectedly Simple Failure of Determinism', *Philosophy of Science* **75**(5), 786–798.

Orilia, F. & Swoyer, C. (2020), Properties, *in* E. N. Zalta, ed., 'The Stanford Encyclopedia of Philosophy', summer 2020 edn, Metaphysics Research Lab, Stanford University.
URL: *https://plato.stanford.edu/archives/sum2020/entries/properties/*

Paoletti, M. P. & Orilia, F. (2017), Downward Causation: An Opinionated Introduction, *in* M. P. Paoletti & F. Orilia, eds, 'Philosophical and Scientific Perspectives on Downward Causation', New York: Routledge, pp. 1–21.

Papineau, D. (1993), *Philosophical naturalism*, Blackwell, Oxford, UK ; Cambridge, Mass.

Papineau, D. (2002), *Thinking about consciousness*, Clarendon Press ; Oxford University Press, Oxford : New York.

Pearl, J. (2000), *Causality: models, reasoning, and inference / Judea Pearl.*, University Press, Cambridge.
URL: *http://dx.doi.org/10.1017/CBO9780511803161*

Popper, K. R. (1957), The propensity interpretation of the calculus of probability, and the quantum theory, *in* S. Körner, ed., 'Observation and Interpretation', Butterworths, pp. 65–70.

Popper, K. R. (1959), 'The Propensity Interpretation of Probability', *The British Journal for the Philosophy of Science* **10**(37), 25–42.

Popper, K. R. (1990), *A world of propensities*, Thoemmes, Bristol.

Psillos, S. (2011), The idea of mechanism, *in* P. M. Illari, F. Russo & J. Williamson, eds, 'Causality in the Sciences', Oxford University Press.
URL: *https://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780199574131.001.0001/acpr*
9780199574131-chapter-36

Puccetti, R. (1977), 'The Great C-Fiber Myth: A Critical Note', *Philosophy of Science* **44**(2), 303–305.
URL: *http://www.jstor.org/stable/187355*

Putnam, H. (1975), The nature of mental states, *in* H. Putnam, ed., 'Philosophical Papers Volume 2 Mind, Language and Reality', University Press, Cambridge.
URL: *http://dx.doi.org/10.1017/CBO9780511625251*

Quine, W. V. (1969), *Ontological Relativity and Other Essays*, 423.15 edn, Columbia University Press, New York Chichester, West Sussex.

Quine, W. V. (2014), *Word and object; foreword by Patricia Smith Churchland; preface to the new edition by Dagfinn Føllesdal.*, new ed. edn, The MIT Press, Cambridge, Massachusetts.
URL: *http://cognet.mit.edu/book/word-and-object*

Ramsey, F. P. (1978), *Universals of Law and of Fact*, International library of psychol-

ogy, philosophy, and scientific method, edited by d. h. mellor edn, Routledge and Kegan Paul, London.

Reichenbach, H. (1949), *The theory of probability: an inquiry into the logical and mathematical foundations of the calculus of probability*, University of California Press, Berkeley, Calif.

Reichenbach, H. (1956), *The Direction of Time*, Dover.

Richards, W. G. (1983), *Quantum pharmacology / W.G. Richards.*, 2nd ed. edn, Butterworths, London ; Boston, London.
URL: *https://www.sciencedirect.com/science/book/9780408709507*

Rorty, R. (1965), 'Mind-Body Identity, Privacy, and Categories', *Review of Metaphysics* **19**(September), 24–54.

Rudolf Carnap (1951), *Logical foundations of probability*, Routledge and Kegan Paul, London.

Révész, P. (1968), *The laws of large numbers*, number 4 *in* 'Probability and mathematical statistics', Academic Press, New York, New York ; London.
URL: *https://www.sciencedirect.com/science/book/9781483230559*

Ryle, G. (1988), *The concept of mind*, repr edn, Penguin Books, London. OCLC: 248432191.

Salmon, W. (1997), 'Causality and explanation: A reply to two critiques', *Philosophy of Science* **64**(3), 461–477. Chicago.

Salmon, W. C. (1980), 'Causality: Production and Propagation', *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association* **1980**(2), 49–69.

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

Salmon, W. C. (1998), Probabilistic Causality, *in* 'Causality and Explanation', Oxford University Press, pp. 208 – 229. ISBN: 9780195108644.

Schaffer, J. (2000), 'Causation by Disconnection', *Philosophy of Science* **67**(2), 285–300.
URL: *http://www.jstor.org/stable/188725*

Schaffer, J. (2001), 'Causes as Probability Raisers of Processes', *The Journal of Philosophy* **98**(2), 75–92.

Schaffer, J. (2007), 'Deterministic Chance?', *The British Journal for the Philosophy of Science* **58**(2), 113–140.

Schizophrenia - Treatment (2017). Library Catalog: www.nhs.uk Section: conditions.
URL: *https://www.nhs.uk/conditions/schizophrenia/treatment/*

Schneider, S. (2017), 'Events', *Internet Encyclopedia of Philosophy*, ISNN 2161-0002'.
URL: *http://www.iep.utm.edu/events/*

Schrödinger, E. (1926), 'An Undulatory Theory of the Mechanics of Atoms and

Molecules', *Physical Review* **28**(6), 1049–1070. American Physical Society.
URL: *https://link.aps.org/doi/10.1103/PhysRev.28.1049*

Schrödinger, E. (1935), 'Die gegenwärtige Situation in der Quantenmechanik', *Naturwissenschaften* **23**, 807–812.
URL: *http://adsabs.harvard.edu/abs/1935NW.....23..807S*

Schwitzgebel, E. (2015), Belief, *in* E. N. Zalta, ed., '*The Stanford Encyclopedia of Philosophy*', summer 2015 edn, Metaphysics Research Lab, Stanford University.
URL: *https://plato.stanford.edu/archives/sum2015/entries/belief/*

Selective serotonin reuptake inhibitors (SSRIs) (2017). Library Catalog: www.nhs.uk Section: conditions.
URL: *https://www.nhs.uk/conditions/ssri-antidepressants/*

Sellars, W. S. (1963), Philosophy and the Scientific Image of Man, *in* R. Colodny, ed., 'Science, Perception, and Reality', Humanities Press/Ridgeview, pp. 35–78.

Shelley, M. W. (1818), *Frankenstein; or, The Modern Prometheus*, Open Road Media Integrated Media, 2014 edition, New York.
URL: *https://search.proquest.com/pq1lit/legacydocview/EBC/1799658*

Sider, T. (2003), 'What's So Bad About Overdetermination?', *Philosophy and Phenomenological Research* **67**(3), 719–726.

Smart, J. J. C. (1959), 'Sensations and Brain Processes', *The Philosophical Review* **68**(2), 141–156.

Spirtes, P. & Scheines, R. (2004), 'Causal Inference of Ambiguous Manipulations', *Philosophy of Science* **71**(5), 833–845.

Strawson, P. F. (2006), *Individuals: an essay in descriptive metaphysics*, Routledge, London.

Suppes, P. (1970), *A Probabilistic Theory of Causality*, North-Holland Publishing Company. Google-Books-ID: Ff4HAQAAIAAJ.

Taleb, N. N. (2008), *The black swan: the impact of the highly improbable*, Penguin, London.

Thau, M. (1994), 'Undermining and Admissibility', *Mind* **103**(412), 491 – 503. Place: Oxford.

Tobia, K. P. (2017), 'Phineas Gage'.

Types of talking therapies (2019). Library Catalog: www.nhs.uk Section: conditions.
URL: *https://www.nhs.uk/conditions/stress-anxiety-depression/types-of-therapy/*

Veit, L. & Nieder, A. (2013), 'Abstract rule neurons in the endbrain support intelligent behaviour in corvid songbirds', *Nature Communications* **4**, ncomms3878.
URL: *https://www.nature.com/articles/ncomms3878*

Vicente, A. (2006), 'On the Causal Completeness of Physics', *International Studies in the Philosophy of Science* **20**(2), 149–171.

Vicente, A. (2011), `Current Physics and `the Physical", *The British Journal for the Philosophy of Science* **62**(2), 393–416.

Von Neumann, J. (1955), *Mathematical foundations of quantum mechanics / John Von Neumann / translated from the German edition by Robert T. Beyer*, Investigations in physics ; no. 2, Princeton UP; Oxford UP.

von Wright, G. H. (1971), *Explanation and understanding.*, Contemporary philosophy, Cornell University Press, Ithaca, NY.

Weir, A. A. S., Chappell, J. & Kacelnik, A. (2002), `Shaping of Hooks in New Caledonian Crows', *Science* **297**(5583), 981–981.
URL: *http://science.sciencemag.org/content/297/5583/981*

Wigner, E. (1962), Remarks on the mind-body question, *in* I. J. Good, ed., `The Scientist speculates: an anthology of partly-baked ideas', Basic Books, New York.

Wittgenstein, L. (2010), *Philosophical investigations*, 4th ed. edn, Wiley, Hoboken.
URL: *http://UCL.eblib.com/patron/FullRecord.aspx?p=514408*

Woodward, J. (2003), *Making things happen: a theory of causal explanation*, Oxford scholarship online, Oxford University Press, Oxford.
URL: *http://dx.doi.org/10.1093/0195155270.001.0001*

Woodward, J. (2013), Causation and Manipulability, *in* E. N. Zalta, ed., `The Stanford Encyclopedia of Philosophy', winter 2013 edn, Metaphysics Research

Lab, Stanford University.

URL: *http://plato.stanford.edu/archives/win2013/entries/causation-mani/*

Woodward, J. (2015), 'Interventionism and Causal Exclusion', *Philosophy and Phenomenological Research* **91**(2), 303–347.

Woodward, J. (2017), Intervening in the Exclusion Argument, *in* H. Beebee, C. Hitchcock & H. Price, eds, 'Making a Difference', Oxford University Press, pp. 251 – 267. ISBN: 9780198746911.

Wray, K. B. (2015), 'Pessimistic Inductions: Four Varieties', *International studies in the philosophy of science* **29**(1), 61–73.

Yablo, S. (1992), 'Mental Causation', *The Philosophical Review* **101**(2), 245–280.

Yablo, S. (2002), 'De-Facto dependence', *Journal Of Philosophy* **99**(3), 130–148.

Zhong, L. (2011), 'Can Counterfactuals Solve the Exclusion Problem?', *Philosophy and Phenomenological Research* **83**(1), 129–147.

Zhong, L. (2012), 'Counterfactuals, regularity and the autonomy approach', *Analysis* **72**(1), 75–85. Oxford University Press.