ON THE QUA PROBLEM QUA QUA PROBLEM:
THE EPISTEMOLOGY AND METAPHYSICS OF MENTAL CAUSATION

David James Frost

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Philosophy.

Chapel Hill
2011

Approved by:
Alan Nelson
William Lycan
Simon Blackburn
Doug Long
James Lesher
ABSTRACT

David Frost: On the Qua Problem Qua Qua Problem
(Under the direction of Alan Nelson and William Lycan)

Mental causation is one of the most discussed topics in the philosophy of mind. The well-known “exclusion problem” is, however, often confused with a different problem called the “qua problem.” Contrary to conventional wisdom, Davidson’s anomalous monism does not suffer from the exclusion problem, although it may seem to engender the qua problem. That is, a given mental event can be causally related to the effect to be explained (ruling out the exclusion problem). But it won’t be related in virtue of its mental properties. It won’t be related qua mental. The critics of anomalous monism say the mental event c is a cause of the effect e. But, the critics ask, in virtue of what is c a cause of e? This question, I argue, contains an unnoticed equivocation. There is an ambiguity in the critics’ formulation of the qua problem. There are in fact two different “in virtue of” questions, one epistemological and one metaphysical, which are confused in the literature. I review some of the relevant parts of the philosophical literature on mental causation, causation and causal laws. And then after surveying the interpretative debate about what the historical Hume really said, I put forth my own view on the disambiguated metaphysical “in virtue of” question. The “Reverse view,” as I call it, answers a Euthyphro question about causes and laws in a novel, quasi-Humean way.
Dedicated to Indigo Michael Frost
ACKNOWLEDGEMENTS

I owe special thanks to the members of the Department of Philosophy at the University of North Carolina, Chapel Hill for their help and guidance, especially Alan Nelson, Bill Lycan, Simon Blackburn, James Lesher, Doug Long, Ram Neta, Allan Silverman, Laurie Paul, Matt Kotzen, Jesse Prinz, Geoff Sayre-McCord, David Reeve, and Jan Boxill. And I wish to acknowledge the edifying friendship I enjoyed with my professors and colleagues at University of California, Irvine, including David W. Smith, Martin Schwab, Nicholas P. White, William Bristow, Ermanno Bencivenga, Michelle Montague, and “The Group,” Hera Arsen, Christopher Lay and Jeffrey Ogle.

I also thank my undergraduate teachers Gayatri Spivak, Mic Taussig, and Patricia Kitcher, and my high school mentor, Ross Duran.

I thank my colleagues at Chapel Hill for their friendly and helpful philosophical discussion, especially Kristen Bell, Davis Bennett, Seth Bordner, Jason Bowers, Patrick Connolly, Corey Cusimano, Antti Kauppinen, Cathay Liu, Ted Parent, Amber Ross, and Carla Saenz.

And I thank the people who have supported me throughout it all, especially my parents Patricia Frost and Terry Frost, as well as the late Doris Clerin, Gary Clerin, Andrea Ford, Sophie Forrester, John Frost, Stacey Frost, Dorothea Greene, Rob Houston, Jasmine Hudepohl, Linda Morse, and Mike Roach.
# TABLE OF CONTENTS

List of Figures .................................................................................................................. vii

Prologue ............................................................................................................................ 1

Overview ........................................................................................................................... 4

Chapters and Sections

1. Reason Explanations and Mental Causation ............................................................... 13
   1.1 A Mental Causation Catastrophe ................................................................. 13
   1.2 Causalism about Reasons ........................................................................... 18
   1.3 Actions, Reasons and Causes .................................................................... 22
   1.4 New Arguments against Causalism ........................................................... 27
   1.5 A Euthyphro Question regarding Reasons and Causes .............................. 30
   1.6 Conclusion .................................................................................................... 37

2. Causation and Causal Laws ......................................................................................... 39
   2.1 Introduction .................................................................................................... 39
   2.2 Causal Accounts and Analyses ................................................................. 48
   2.3 Adopting a View ........................................................................................... 63
   2.4 On Causal Laws ............................................................................................ 66
   2.5 Conclusion .................................................................................................... 76

3. The Metaphysics of the Exclusion Problem ............................................................... 77
   3.1 Introduction .................................................................................................... 77
   3.2 Brief Formulation of the Exclusion Problem .............................................. 77
   3.3 Brief History of the Mind-Body Problem .................................................... 79
   3.4 Formulation of the Exclusion Problem ...................................................... 84
   3.5 What is Supervenience? .............................................................................. 90
   3.6 Kim’s Exclusion + Supervenience Argument ............................................ 93
   3.7 An Initial Pragmatist Rejoinder ................................................................. 99
   3.8 Canonical Solutions ..................................................................................... 102
       3.8.1 Eliminativist Reductive Physicalism ................................................... 102
       3.8.2 The Dual Explananda Response ......................................................... 103
       3.8.3 Yablo’s Proportionality Constraint ..................................................... 108
       3.8.4 Horgan’s Quausation ................................................................. 115
   3.9 Anomalous Monism ...................................................................................... 119
   3.10 Anomalous Monism and the Identification Problem ................................. 127
   3.11 Conclusion .................................................................................................... 130

4. Anomalous Monism and the Qua Problem ............................................................... 131
   4.1 Introduction .................................................................................................... 131
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.2</td>
<td>The Criticism</td>
<td>133</td>
</tr>
<tr>
<td>4.3</td>
<td>Canonical Formula for Causal Explanations</td>
<td>135</td>
</tr>
<tr>
<td>4.4</td>
<td>Epistemically Adequate Explanatory Generalizations</td>
<td>137</td>
</tr>
<tr>
<td>4.5</td>
<td>Two Conditions on Causal Explanations’ Adequacy</td>
<td>144</td>
</tr>
<tr>
<td>4.6</td>
<td>What Metaphysically Undergirds the Causal Relation?</td>
<td>147</td>
</tr>
<tr>
<td>4.7</td>
<td>Two Distinct “In Virtue of” Questions</td>
<td>147</td>
</tr>
<tr>
<td>4.8</td>
<td>My Criticism of the Critics’ Objection</td>
<td>148</td>
</tr>
<tr>
<td>4.9</td>
<td>The Shattering Glass Objection</td>
<td>152</td>
</tr>
<tr>
<td>4.10</td>
<td>LePore and Loewer on Two Types of Causal Relevance</td>
<td>153</td>
</tr>
<tr>
<td>4.11</td>
<td>Davidson on “Cause” and “Explain”</td>
<td>155</td>
</tr>
<tr>
<td>4.12</td>
<td>Conclusion</td>
<td>156</td>
</tr>
<tr>
<td>5.</td>
<td><strong>Coda: New Hume and the Quasi-Humean Reverse View</strong></td>
<td>158</td>
</tr>
<tr>
<td>5.1</td>
<td>The New Hume Debate</td>
<td>158</td>
</tr>
<tr>
<td>5.2</td>
<td>The Reverse View</td>
<td>164</td>
</tr>
<tr>
<td>5.2.1</td>
<td>Introduction</td>
<td>164</td>
</tr>
<tr>
<td>5.2.2</td>
<td>The Second “In Virtue Of” Question and Law</td>
<td>165</td>
</tr>
<tr>
<td>5.2.3</td>
<td>No Catastrophe</td>
<td>167</td>
</tr>
<tr>
<td>5.2.4</td>
<td>Wider Ramifications</td>
<td>167</td>
</tr>
<tr>
<td>5.2.5</td>
<td>The Reverse View Versus Armstrong</td>
<td>168</td>
</tr>
<tr>
<td>5.2.6</td>
<td>Armstrong’s Counterintuitive Realism</td>
<td>169</td>
</tr>
<tr>
<td>5.2.7</td>
<td>Conclusion</td>
<td>171</td>
</tr>
</tbody>
</table>

**Conclusion** ...................................................................................................................................173

**Epilogue** .......................................................................................................................................177

**Bibliography** ................................................................................................................................178
LIST OF FIGURES

Figure

1. Figure 1: The Common Cause Error .................................................................41
2. Figure 2: Nietzsche’s Cornarism .................................................................41
3. Figure 3: Epiphenomenalism ........................................................................47
4. Figure 4: Atmospheric Pressure, Barometer Reading, and Storm ..............59
5. Figure 4: Kim’s Framework ............................................................................94
PROLOGUE

Davidsonian philosophy of mind… enables us to treat both physics and poetry evenhandedly.

In what way does Davidson’s philosophy of mind help us to treat physics and poetry evenhandedly? When Rorty says “us” he means we philosophers. And the way he suggests we treat physics and poetry (we may substitute Geisteswissenschaften) is to basically leave them alone. Indeed, I take it that Rorty’s Davidsonian pragmatism says that the special sciences do not need nor admit of any metaphysical justification.¹ As one commenter says, “Rorty scorns the very project of giving philosophical accounts of [subjects of traditional philosophical interest] or pretty much any other phenomena,” (Manning 2006: 468). Rorty himself writes that, “Part of my ambition… is to help it come to pass that where epistemology and metaphysics were, sociology and history shall be,” (2000: 103). Physics, biology, psychology, sociology, and, yes, poetry and the humanities are “vocabularies” that stand on their own without the need for metaphysical foundation. Rorty has an agenda for philosophy that is

¹ Brandom (2000) writes that, according to Rorty, democracy “does not need nor admit of metaphysical justification.” The same goes for other human endeavors such as the special sciences.
against any attempt to give metaphysical justification for the special sciences as well as for basic physics.

In the metaphysics of mind, the problems of mental causation – especially “the exclusion problem” and “the qua problem” – are good cases by which to test Rorty’s pragmatist, anti-metaphysical agenda for philosophy. The mental causation problem for the higher-order special sciences, and especially for psychology, is a request for a justifying metaphysical account of the possibility of those sciences and of their product: knowledge-giving causal explanations.

The explanations offered by the sciences are adequate explanations (or they are not) depending on methods of assessment internal to the sciences themselves – and without philosophy’s help. However, I argue that there is a “minimal metaphysics” that one is committed to upon affirming the explanatory adequacy of those explanations. That is to say, there is a certain relation between epistemology and metaphysics that even the pragmatist is committed to. Simply having the confidence in the explanations coming from the special sciences – that these sanguine pragmatists do have – imputes to them certain metaphysical commitments, however minimal.

On this view, metaphysics is not necessarily a discussion about what is really real. It is not a discussion about the fundamental constituents of reality. In some important sense, we may let the physicists and the rest of the scientists tell us that. In that way, we naturalize epistemology. They will tell us what we know to be real. And then metaphysics becomes the speculative inquiry into the necessary conditions of knowledge as delivered by science. Metaphysics becomes what else we are committed to beyond the vocabulary of the sciences when we accept the causal explanations of those sciences.
The way of doing metaphysics I have in mind is a matter of asking and answering a “how-possible” question of actual phenomena. For example, we come to have knowledge in such and such a way; how is it possible that we come to knowledge in that way? This kind of account of our actual epistemological practices is meant to establish metaphysical legitimacy for those practices, which are of paramount salience in our most global self-conception as human beings.

With respect to our case study – the mental causation debate – a tension arises if we allow that causation somehow involves necessity and simultaneously note that special science explanations involve laws or generalizations that are non-strict and not exceptionless: i.e., they fall short of making the effect necessary given the cause. Accepting this generates the problem. So, going forward I am going to assume that causation involves necessity, although just how that cashes out is open for discussion. I am also going to take it as given that sometimes our sciences – physics and the special sciences – provide genuine causal explanations. The task is to reconcile these two assumptions. And this task demands – even of the pragmatist – a minimal amount of speculative metaphysics. Or so I would argue.
OVERVIEW

1. When I walk into the grocery store, I like to snap my fingers just before the automatic doors open. It gives the impression that my snapping caused the doors to open, like I’m the Fonz from *Happy Days*.

   Or consider this possible case: In a time of inefficient information distribution, a court astronomer could discover how to calculate when the next eclipse is going to occur – and down to the second. This would make it possible for the King to call a public assembly at that day and time and – just before the eclipse begins – announce, “In the King’s name!” gesturing at the sun. This would give the impression that the King had caused the eclipse.

   In each of these cases, we recognize that a purported cause is actually not a cause at all. The King’s gesture doesn’t cause the eclipse. Had the King done nothing, the eclipse would have occurred all the same. And my snapping my fingers doesn’t cause the doors to open. It is sufficient that I walk in front of the automatic doors; they’ll open whether or not I’m snapping my fingers.

   Now consider a case from the philosophy of mind, specifically couched in mental terms. When a student raises her hand because she wants to make a point, we might say that her *wanting* caused the action. Yet it also seems true that said wanting – a movement of the mind, a mental event – is *realized* by a physical event in the nervous system. So we might also say that some state of her (complex) physical system antecedent to the action caused the action. These two purported causes – the wanting and the physical realizer of the wanting –
some philosophers say compete with respect to being the cause of the action. Then, because of orthogonal commitments to a certain kind of physicalism, it can seem to follow that the mental event does not cause the action at all. The purported physical cause is sufficient according to these physicalist commitments and “excludes” the mental event from being a cause.2

2. This exclusion appears to be a problem because the mental event we were comfortable appealing to as a cause of an action is actually excluded from being the cause of the action by plausible claims about the physical world itself. Mental causation involves an apparently fundamental tension between the conception of ourselves as minded, causally efficacious agents and our scientific conception of how the world (of which we are a part) works. The reason the failure to establish mental causation seems devastating is because the salience of our practice of reason explanation commits us very deeply to a causal explanatory role for mental properties.3 If they are not causally explanatory then that important practice falls apart.4

---

2 The analogy between the automatic door, astronomy cases and the mental case breaks down at some point, of course. The first two cases offer potential causes that are not really believable while it is supposed to be a deeper quandary with respect to which one is causal in the mental causation debate, M or P. However, we might assume that the witnesses to these events believe in the causality of the finger-snap and the King’s gesture. When they learned of the deceit they would “exclude” the purported causes not on their knowledge of sensor technology and astronomy but simply on an exclusion or non-overdetermination principle.

3 I may sometimes refer to properties (properly, property instances) or descriptions. While these terms are certainly not interchangeable I do not intend to make much of the differences between them. Some have defended Davidson against his critics by remarking that Davidson was a nominalist and spoke almost always of predicates and descriptions and rarely if ever of properties. For instance, see Gibb (2006). For my purposes, however I do not wish to let Davidson off the hook by invoking his nominalism. Instead I wish to remain at the level that almost everyone else operates at, namely the level of properties as opposed to predicates. I may sometimes say “descriptions” but I am assuming those descriptions cannot be true unless the relevant properties are instantiated by what is described. I do however count on the difference between properties and property bearers, namely event tokens. As per Sections 1.3.2 and 4.3 the extensional relation of causation holds between event tokens and not properties.
Fodor says that failing to establish mental causation would be “the greatest intellectual catastrophe” (2002: 19) for humankind ever. I believe, although Fodor speaks of “causal responsibility” – what other words could he use? – that he means to be speaking of causal explanation. If our mental properties do not causally explain our behavior then it is the end of the world. I will call this – alluding to Fodor’s assessment – the Catastrophe thesis.

So it is a matter of the all-important “because.” The eclipse happened when the King willed it but not because the King willed it. The door opened when I snapped my finger, but not because I snapped my finger. And if the story about exclusive, sufficient physical causes is correct, then her arm went up when the student willed it, but not because she willed it. The purported mental cause seems epiphenomenal. It may always be correlated with the action while never causing it.

3. Any theory of mind and body must contend with the exclusion problem or show that the theory does not engender it. Davidson’s anomalous monism does not engender the exclusion problem. Instead it seems to suffer from the importantly different qua problem. According to the exclusion problem, the given reason or mental event is excluded from being a cause; while according to the qua problem the mental event is a cause but not in the right way. The

---

4 Causalism about reasons has its echt-defense in Davidson (1963). I defend Causalism from contemporary anti-causalists in Chapter One.

5 Indeed, a central claim of Davidson (1967) is that people say “cause” and mean “causally explain.” See Davidson (1967: 162). See my Section 4.11 below.

6 The exclusion problem and the qua problem are importantly different although they look similar and are confused in the literature. I’ll argue below that the exclusion problem’s Causal Closure Principle is an existential claim about the existence of a certain kind of cause, while anomalous monism’s Cause-Law Thesis from the qua problem is a claim about the nature of causation.
critics of anomalous monism allow that on Davidson’s theory the mental event is a cause but they argue it is not a cause qua mental event – it is not a cause in virtue of its mentality. That’s the qua problem. “Agreed, c is a cause of e but in virtue of what is c a cause of e?” That is the “in virtue of” question that gets the qua problem going. I argue that the critics of anomalous monism do not understand what they do when they propose the qua problem. I show them what is really going on: I show them the qua problem qua qua problem. By the end of this dissertation, I will have argued that there are two – as yet to be explained – “in virtue of” questions, one epistemic and one metaphysical, which are often confused in the mental causation literature. This pragmatist or “Kantian” difference between epistemology and metaphysics is a point that will recur throughout the dissertation. I carefully distinguish the two “in virtue of” questions and, then at the end in a kind of coda, I propose a novel answer to the newly disambiguated metaphysical “in virtue of” question.

4. In the first chapter of this dissertation, I argue for Causalism about reasons, i.e., that reasons are causes. I lay out a novel interpretation of Davidson’s “multiple reasons argument.” While there are many reasons that could be the reason for which an agent acted – that is, while there are many reasons that rationalize the agent’s action – the one that the agent actually acted on is the cause of the action. Implied in the multiple reasons argument is, I argue, a certain answer to a Euthyphro question about reasons and causes. Surveying what I call The Strict Qualification Thesis and The Priority of the Phenomenon Thesis, I consider the ramifications of the Euthyphronic answer that says a reason is a cause in virtue of being a reason (this answer simultaneously denies that a cause is a reason in virtue of being a cause). Here too there is an interesting upshot with respect to the relationship between metaphysics and epistemology.
5. If we are to argue that reasons are causes and that mental events are causes, it would be wise to have some understanding of the nature of causation. It is easy to say that something is causal. But we will not know what it means to say so until we get some understanding of the nature of causation per se. To that end, in Chapter Two I conduct a literature review of causation and causal explanation, including scientific explanation. The sharp distinction I make between causal explanation and causation, one which is epistemic and the other metaphysical, is backed up by the philosophical literature which itself makes a sharp distinction between metaphysical causation and epistemic causal explanation. Finally, in advance of moves I make in Chapter Five, it is also necessary to review the literature on causal laws and the difference between accidental generalizations or de facto regularities and laws that possess an aspect of necessity. In Chapter Five I (merely) outline a novel position on the relation between causes and laws.

6. In the third chapter of this dissertation, I offer a brief history of the mental causation debate. The contemporary roots are in the 1950s identity arguments, but a cousin or sister problem would certainly be Descartes’ problem with interaction of mind and matter. I present and discuss the contemporary formulations of the variety of mental causation problems. The canonical attempted solutions are also presented and criticized. I discuss the required background and useful concepts such as supervenience. Then I look in detail at the exclusion problem for non-reductive physicalists as proposed by Kim. Burge and Baker’s “explanatory practice” objection to Kim’s exclusion problem is found wanting. I conclude in Kim’s favor that indeed the problem of mental causation is a “metaphysical” problem and needs a “metaphysical” solution. Finally I survey a variety of responses to Kim’s exclusion
problem, including Horgan (1989), Yablo (1992) and Davidson (1970). Here we see that Davidson does not suffer from the exclusion problem, although he does apparently encounter the qua problem.

7. In the fourth chapter, I intervene in one small sub-debate in the mental causation discussion. Almost everyone agrees with Kim (1989, 1993, 1998) that Davidson’s anomalous monism suffers from the qua problem. I show that the critics have made a mistake. Kim et al equivocate over the words “in virtue of” in the question they put to Davidson which supposedly engenders the qua problem. To repeat: The qua problem is that, while it may be the case that the given mental event token causes the effect event token, it does not cause it in virtue of its mental properties – it does not cause it qua mental event. Rather, according to anomalous monism’s commitment to Nomological Causation, it causes the effect in virtue of its physical properties. If Davidson is asked, “The mental event token causes the physical event, but does it cause it in virtue of its mental properties or in virtue of its physical properties?” then he must reluctantly say, “in virtue of its physical properties.” Or so say the critics.

---

7 There are other sub-debates in the mental causation literature. See Robb & Heil (2008) for a discussion of “the problem of exclusion,” “the problem for anomalous monism,” (what I call the qua problem) and “the problem of externalism,” also called “wide causation.” With respect to the latter, “Suppose, as many philosophers do suppose, that externalism is true: the contents of representational states of mind—propositional attitudes, perceptual experiences, mental images, and so on—depend, not merely on intrinsic features of those states, but on relations—in particular, on the causal, historical, and social relations agents bear to their surroundings,” (Section 7.1). In that case, then there is a worry about the causal role of our beliefs in what we do. “…If the content of Lilian’s thought that there is a tree in the quad is “broad”, if the significance of her thought depends on factors outside Lilian’s body (Lilian’s standing in an appropriate causal relation to the tree, for instance), then it is indeed hard to see how this content could figure in a causal account of Lilian’s actions,” (7.1).


9 And thus does not engender the exclusion problem.
I show that there are two fundamentally different questions to which some properties and not others answer in this ballpark. That is, when we are philosophizing about causal explanations and we ask “in virtue of what is c a cause of e?” we might mean at least two distinct things. I sometimes call one the epistemic “in virtue of” question and the other the metaphysical “in virtue of” question and at other times simply the first “in virtue of” question and the second “in virtue of” question, for the misleadingly vague use of “epistemic” and “metaphysical” in the philosophical conversation.10

With these two distinct questions in hand, I show that neither of anomalous monism’s answers to these two questions entails Catastrophe. Importantly, with Davidson’s supposed Principle of the Nomological Character of Causation (PNCC) he can answer the second “in virtue of” question about what makes a cause be the cause it is, without fear of entailing Catastrophe. The qua problem has no teeth, at least not as usually formulated against non-reductive token-identity views like anomalous monism.

8. Having isolated the second “in virtue of” question about what makes a cause be the cause it is, I turn to examine a variety of possible answers to this question. In the mental causation debate, one of the most prominent answers to the second “in virtue of” question – if not indeed the most prominent answer – is the answer provided by the nomic subsumption view. According to the nomic subsumption account of causation, a cause event token, c, is related

10 What I show is that on the first or epistemic “in virtue of” question, we are asking in virtue of what is reference to c’s being F an explanation of e’s being G? In other words, in virtue of what is it explanatory to refer to c’s-tokening-property-F with respect to the explanandum of e’s tokening G? On the second or metaphysical “in virtue of” question, we are asking in virtue of what properties, H and J, is c a cause of e, or, in other words, in virtue of what is c related by the causal relation to e? These are different questions provided one avoids what I call Procrusteanism about causal explanation where F=H and G=J, which I argue is untenable anyway. See Section 4.7 below.
to an effect event token, e, in virtue of c and e tokening, respectively, properties H and J which together instantiate a strict or exceptionless law. I make the observation that in Davidson’s original formulation of his *Principle of the Nomological Character of Causation* he stayed neutral as to what was in virtue of what, laws or causes. The PNCC just says that when there are causal relations there are laws. It does not say that there are causal relations *because* there are laws or that there are laws *because* there are causal relations. There remains after all these years a Euthyphronic question to be posed to Davidson’s PNCC.

On the one hand, you could interpret the PNCC with a metaphysical order of explanation *from* nomic properties to the causal relation, such that the laws metaphysically explain or “govern” causes. The PNCC would thereby be read as Davidson’s critics actively interpret it: “When two events are related as cause and effect *this is because* they have descriptions under which they instantiate a strict law of physics.” But, on the other hand, the order of explanation could conceivably go the other way, such as to imply a metaphysical order of explanation going from causes to laws. Then we might say that it is *because* certain causal relations obtain that certain laws obtain. I call this the “Reverse View.”

What I outline is a broadly Humean (although considerable historical, poetic and neologistic license is taken using this name) view modeled on discussions in the laws of nature literature such that the law between H and J is what it is in virtue of the causal relation between c and e being what it is. Having explored the so-called New Hume debate, I outline my so-called Quasi-Humean “Reverse View” and argue for one of its main virtues. I also make one attack against the main alternative to it, namely an Armstrong backed nomic subsumption view.
9. On the Reverse View, laws depend in a sense on causal relations which themselves are, in a sense to be explained, brute. But the bruteness of causal relations in this sense is not as “implausible” as McLaughlin (1993: 23) and others take it to be, or so I argue in the final chapter. The worry about causal relations being brute is that if that were the case then, “there would be nothing in virtue of which c is a cause of e,” (23). I examine this worry in order to show that those who would put forth this worry find it absurd that “there would be nothing in virtue of which c is a cause of e.” I argue, however, that as a metaphysical view bruteness is nowhere near absurd. It’s only on an epistemological understanding of “in virtue of” there that some sense of bruteness seems absurd.

Finally I develop problems for the traditional Armstrong-backed nomic subsumption view. It answers the Euthyphronic question about causes and laws in the traditional way – causes depend on laws – which commits the view to its own sort of brute fact. That is, it must assert, what was called in the New Hume debate, a “straitjacketing” fact. I outline a more empiricist location for brute primitive facts.
CHAPTER ONE

REASON EXPLANATIONS AND MENTAL CAUSATION

1.1 A Mental Causation Catastrophe

It has been a longstanding project of the metaphysics of mind to vindicate the explanatory status of our commonsense, folk psychological explanations of intentional behavior.\(^{11}\) These explanations appeal to an agent’s mental events such as beliefs, desires and/or reasons in order to explain the agent’s action.\(^{12}\) Consider an elementary example: we say that Georgia was thirsty and believed there was beer in the fridge in order to explain why Georgia walked to the fridge. Some philosophers say that explanations of this sort do their explaining by making the action-to-be-explained seem a reasonable thing to do given the beliefs and desires the agent had. In this sense, the explanations are rationalizing explanations, making appeal to the agent’s purposes or reasons.

Philosophers in the theory of action disagree about whether or not reason explanations are causal explanations. For a long while in the history of recent philosophy,

---

\(^{11}\) An action is a behavior that is intentional, i.e., done for a reason. This follows Anscombe’s account of what distinguishes actions that are intentional from those that are not: “Intentional actions are the actions to which a certain sense of the question ‘Why?’ is given application; the sense is of course that in which the answer, if positive, gives a reason for acting,” (1957: 9).

\(^{12}\) Participants in the debate I am engaging with all agree that a reason is a mental event. For our purposes, it will do to say the event that is the reason is the “coming to be” of a pro-attitude paired with a belief. Recently, some have denied that reasons are mental events. See Dancy (2004) and Horgan (2007) for example. See below Section 1.4 for a discussion.
many philosophers held that reasons could not be causes. But Davidson’s (1963) widely influential “multiple reasons argument” suggests that rationalizing explanations must also be causal explanations. According to Causalism, the explanatory reason for an action must be the cause of the action. Causalism in the philosophy of action then hooks up with the mental causation debate in the philosophy of mind. Causalism argues that reasons had better be causes; the mental causation debate is about understanding how that is possible.

The multiple reasons argument, which we will discuss in greater detail below, only shows, it seems to me, that reasons had better be causes. Further work is necessary to show how it is possible for reasons to be causes. As Kim correctly observes the mental causation problem in the philosophy of mind is a how-possible question. The question “whether or not” there is mental causation, Kim says, must be answered in the affirmative.

The phenomenon of mental causation is the thing to be captured by our philosophical work in this area of the philosophy of mind. Nevertheless, Kim is only establishing mental causation by stipulation. I follow him and the rest of the debate in that. However, I can imagine a philosophical account of intentional behavior that, for instance, held the conscious experience of willing or acting intentionally to be an epiphenomenon of a deeper cause of our behavior. Here there would be no mental causation. Or much less of it.

---

13 For instance, Anscombe (1957), Dray (1957), Ryle (1948), and Taylor (1968). The so-called Neo-Wittgensteinians may have been too diverse for such a grouping to be appropriate. Some, for instance, may have taken it that reasons could not be causes, while others may have taken it that reasons need not be causes. In any case, Davidson took them to say that reasons could not be causes.

14 Davidson (1963) argued that rationalizing explanations must also be causal explanations using his “multiple reasons argument,” namely, that there might be any number of reasons that rationalize the action that stands in need of explanation. But the reason for which the agent actually acted will have been the cause of the action (as well as being a reason that rationalizes the action). See Section 1.3 below.
And I do not believe it would absolutely wreck our picture of ourselves. But it is out of the scope of this project to discuss it here.

Participants in the mental causation debate are tasked with vindicating psychological explanations against any eliminative reductionism. If our commonsense psychological explanations – those that appeal essentially to reasons – are to be considered legitimate, it must be the case that the reason an explanation appeals to actually be the cause of the agent’s action. However, the problem is that most of our standard physicalist theories of causation end up ruling out the causal efficacy of mentality. The problem of mental causation is to show how it is possible that mentality be causal so that mentality can be causally explanatory.15

As said above in the “Overview,” the reason the failure to establish mental causation seems devastating is because the salience of our practice of reason explanation commits us very deeply to a causal explanatory role for mental properties.16 As Fodor has written,

…if it isn't literally true that my wanting is causally responsible for my reaching, and my itching is causally responsible for my scratching, and my believing is causally responsible for my saying…. If none of that is literally true, then practically everything I believe about anything is false and it's the end of the world (1990: 156).

15 Admittedly, there are some philosophers who talk about apparently similar matters (mental causation, reduction, etc.) without the aim of underwriting our everyday practices. New-compatibilists or “overlap” theorist (in the phrasing of Harbecke (2008) and Dardis (2009), respectively) are involved in a metaphysical undertaking not essentially tied to the project of underwriting our everyday practices. They include Shoemaker (2007) and Yablo (1992).

16 The Dravians and the Wittgensteinians would rightly dispute this point. After all, they too articulate their non-causal position as being more in line with the structure and nature of specifically human intentional action. As I will discuss below, they were working in a tradition that argued that explanation of human actions in history and psychology are different than explanations in the natural sciences and yet no worse off for it. Nevertheless, almost everyone in the mental causation debate takes it that the phenomenon of mental causation is something established for which we need a metaphysics. And many if not all cite our image of ourselves as casually efficacious agents for why mental causation must be the case.
Elsewhere Fodor says, somewhat less hyperbolically, not that the world would end, but that failing to establish mental causation would be “the greatest intellectual catastrophe” (2002:19) for humankind ever. I take it that, although Fodor speaks of “causal responsibility” – what other words could he use? – he means to be speaking of causal explanation. If our mental properties do not causally explain our behavior then we have what I will call the *Catastrophe* thesis.

*Catastrophe*: Mentality plays no causal role in the causal explanation of intentional action.

If *Catastrophe* is true then the practices mentioned above, which are very salient for us, would fall apart. This is a situation worth resisting.

### 1.1.1 Dissenters

On the other hand, perhaps the philosophers are right who affirm *Catastrophe*. After all, the *Catastrophe* thesis is not catastrophic merely in virtue of its name. It can be endorsed; it’s just that it comes at a price. Who would affirm it and at what cost? Any philosopher who denies causation in the manner of Russell (1913) would thereby deny *mental* causation and affirm

---

17 Indeed, a central claim of Davidson’s view is that people say “cause” and mean “causally explain.” See Davidson (1967/1980 p. 162). See Section 4.11 below.

18 As many have noticed, Kim’s arguments generalize to any domain of properties supervening in the basic physical domain, which is all of them. So by Kim’s argument, more about which see below, geological properties, biological, social and psychological properties are epiphenomena. All the more reason to resist such an account. For more on the generalization argument against Kim see Bontly (2002), Schoeder (2002) and Moore (2010).

19 Russell colorfully writes: “All philosophers of every school, imagine the causation is one of the fundamental axioms or postulates of science, yet, oddly enough, in advanced science such as gravitational astronomy, the word ‘cause’ never occurs… The law of causality, I believe, like much that passes muster among philosophers, is a relic of a bygone age, surviving, like the
Catastrophe. The cost in that case would be the high cost of denying causation entirely (see Price and Corry 2007) much less being able to underwrite our causal explanatory practices with respect to human agency and behavior. This latter cost is shared by eliminative reductionists like Churchland (1981), who does not deny causation per se but denies mental causation because he denies the existence of propositional attitudes entirely and thereby affirms Catastrophe. A view that affirms Catastrophe would have to give an explanation for why our explanatory practices make reference to beliefs and desire as causally effective. It would also have to replace or build back up our self-image as causally effective agents.20

Another group of philosophers endorse Catastrophe, namely the Neo-Wittgensteinians. Not every philosopher was convinced by Davidson (1963) and they remain parties to the debate in the theory of action. But in the philosophy of mind’s mental causation debate the Neo-Wittgensteinians are not parties to the debate, which stipulates that endorsing Catastrophe is more than expensive; indeed it’s catastrophic.

Finally, there has been a movement in empirical psychology to suggest that there is no such thing as mental causation. For example, Wegner (2002) denies the causal efficacy of the conscious will and Wilson (2002) goes a long way toward concluding that most of our actions are automatic and the feeling of willing is a post-hoc construction by the mind.21 On these views, and on certain readings of Nietzsche (see Leiter 2002, 2007), an action and the feeling

monarchy, only because it is erroneously supposed to do no harm.” For a sympathetic but ultimately negative assessment of Russell see Price & Corry (2007).

20 Churchland attempts both of these tasks in his classic (1981). Bontly (2005), as I’ll discuss later, says that mental causation is a Moorean fact, which means that there is more in its favor than could be said for any theory that denied it.

21 See also Haidt (2001) and Greene (2008) for discussion of post-hoc rationalization. Haidt’s work suggests that Wegner’s and Wilson’s work can be generalized from willing to the case of reasons, i.e., beliefs and desires. See Frost (in preparation).
of willing (including a post-hoc construction of a reason for action) are effects of a common cause in the unconscious. On this view, a reason and an action may always be counterfactually dependent in the right way without being causally related. However, I believe this is counterintuitive with respect to our self-image as causally effective agents. There may be ways to build back up an image of the self that is lost when mental causation is denied and Catastrophe is affirmed, but that is not the project in this dissertation.

In mental causation debate that we engage with here, almost all participants take it that reasons are causes. The mental causation debate itself usually stipulates that reasons (mental events) are causes and that we need a metaphysics for how this actuality is even a possibility. But the philosophy of action, starting with Davidson (1963), gives the arguments for the causality of reasons that Kim et al. assume. So, before starting in on the mental causation debate we need to engage with the thesis that reasons are causes, i.e., Causalism about reasons.\footnote{Finally let us say, “we refer to reasons as causes,” and mean by this that Causalism about reasons is true. For Causalism about reasons to be true, mental causation had better be true. Let us say, “Some mental events cause some physical events,” and mean by this that there is truly mental causation.}

\section{1.2 Causalism about Reasons}

In the following sections, I argue for a particular and, I daresay, unique reading of Davidson’s “multiple reasons” argument. The view I lay out is the view I think Davidson was getting at all along, although the secondary literature does not read him this way. If the reader disputes what I say as good Davidson exegesis, I can easily give up the claim that Davidson holds the...
views I discuss here. I will nevertheless assert them myself as a Davidsonian account. They hang together as an account of the relation between actions and the reasons that explain those actions.  

We begin by considering the history of antagonism between proponents of causal explanations and those of “making-intelligible” explanations. This historical sketch leads us from the European debates about the *Geisteswissenschaften* to the anti-causalist milieu in which Davidson wrote “Actions, Reasons, and Causes,” (1963, hereafter ARC). We then consider what is, and what is not, actually demonstrated by Davidson in that article. ARC shows that the explanatory reason must be a cause but ARC does not offer a “causal criterion” by which to discover the causal reason. We discuss my Strict Qualification Thesis and my Priority of the Phenomenon Thesis, which if true suggest a certain pragmatist relation between epistemology and metaphysics.

1.2.1 **History of *Geisteswissenschaften* versus *Naturwissenschaften***

Davidson’s intervention in ARC changed the debate and set the agenda for “a whole generation of philosophers,” (Leist 2007: 1). In order to understand his contribution, we must first properly understand how the dialectic stood prior to Davidson’s contribution. Davidson’s target in ARC – namely, the anti-causalist thesis that rationalizing explanations cannot be causal explanations – has a rich genealogy in European philosophy of social science.

Consider, again, some example explanations:

Jill returned to the bank because she believed she left her glasses there.
Georgia walked to the fridge because she wanted a beer.

---

23 Along the way I articulate a specific position on the relation between epistemology and metaphysics which again I think is in line with Davidsonian pragmatism. But if it’s not, it remains an interesting view on its own.
Let’s assume these are true, accurate and adequate explanations of Jill’s and Georgia’s intentional actions. Many philosophers have thought that explanations of this kind do their explaining by “making intelligible” the agent’s action in light of the beliefs and desires attributed to the agent in the explanation.\(^\text{24}\) Sometimes called “rationalizing” explanations, these explanations of actions show how the action was the rational thing to do given the beliefs and desires of the agent.\(^\text{25}\) They are also often called “reason explanations,” or “intentional explanations.” A certain tradition in the philosophy of social science debates whether or not there is a unified structure of explanation shared by reason explanations and explanations found in natural sciences, such as physics.

In fact, the debate in the philosophy of social science is broader. It tends to contrast explanation in the natural sciences (or, natural scientific explanation) with not only reason explanations but any explanation from a wide range of disciplines including psychology, history, economics, anthropology and sociology. These are sometimes grouped together under the heading of “sciences of man” (more appropriately discriminated as “sciences of the mind”), or Geisteswissenschaften and les sciences d’esprit in German- and French-speaking circles respectively. Geisteswissenschaften, as I will call them, includes psychology while excluding

\(^{24}\) The example explanations I have offered are enthymematic insofar as one references a belief while only implying a relevant desire, and the other references a desire while only implying a relevant belief. That is to say, Jill had to also desire to recover her glasses for our reference to her belief that they were at the bank to be able to explain her returning to the bank. Her desire, however, is plausibly implied. It would be the other way around for Georgia. It is standard since Davidson (1963) to take reasons for action to be a belief and desire pair. (Actually, because of certain counterexamples involving desires the agent does not identity with or value, “desire” is best replaced by “pro-attitude.”) Taking a reason as a pro-attitude/belief pair remains foundational in, for example, Smith’s seminal “Humean Theory of Motivation,” introduced in his (1987) and defended more recently in Smith (2004).

\(^{25}\) Rationalizing explanations can be misleading if it is taken to mean that explanations of this kind only work if an agent acts in a rational manner with respect to her beliefs and desires. Irrational actions can be explained by “taking the perspective” of the irrational actor.
physiology of the brain. They are to be contrasted with Naturwissenschaften, such as biology, chemistry and especially physics. A strict criterion to differentiate is not necessary, won’t delay us and probably wouldn’t be forthcoming anyway.

However, we can make an initial stab at a potential distinguishing difference between Geisteswissenschaften and Naturwissenschaften by simply letting the list of differing attempts at a definition wash over us. Perhaps the top things Geisteswissenschaften has been said to study are: Seelenleben, the life of the soul as opposed to the body; consciousness and events understandable only qua conscious events; society, practical reason, agency and morality. Indeed, Geisteswissenschaften was the word the first German translator of Mill took for Mill’s use of the term “moral sciences.”

We might also discuss our difference in terms of the difference between Erklären and Verstehen. In Europe there was a debate between the proponents of the unity of science, and those who denied a monism of acceptable methodology, i.e., those who defended an autonomous and sui generis “science of man.” The latter preferred to think that explanation of human action provided not Erklären (“clarification/explanation”) but Verstehen (literally, “understanding”). Weber took the word “Verstehen” from Jaspers (Descombes 2001: 37). Dilthey and Gadamer (and later Winch and Taylor) emphasize the affinity between the social sciences and the humanities and the disaffinity between, say, psychology and physics.

1.2.2 Hempel and Dray

The distinction between causal explanations and “making intelligible” explanations was passed into our contemporary discussion from the Vienna Circle via Hempel. Hempel (1942)

---

26 See Descombes (2001: 30).

27 The history and relevance of the Verstehen – Erklären debate is discussed in Apel (1984), Descombes (2001), Marras (1993), Friedman (1999). Relevant are Gadamer (1960), and Dilthey (1883) and von Wright (1971).
argues, for instance, that historical explanation shares the deductive-nomological (DN) structure that physics employs (see Section 2.2 below). An event is successfully explained if and only if it follows deductively from a set of law-like generalizations, together with initial conditions. To use Hempel's example, the Dust Bowl farmers migrated to California because – and here's the law – “populations tend to migrate to areas offering better living conditions” and California offered better living conditions – there's the initial condition. Their migration is predicted by the explanatory law.²⁸

Opposing Hempel was Dray (1957) who argued that the social sciences had a different explanatory goal and different method than the natural sciences. The goal was “understanding” or “intelligibility” not causal explanation, that is, not the kind of explanation offered by natural sciences like physics. The method would be, for example, to make clear the rationality of an action as one done for reasons and to uncover the meaning of, or significance of, the action for the agent who performs it.

§

1.3 Actions, Reasons and Causes

Going forward let Causalism about Reasons be the thesis that the explanatory reason for a given action is the cause of that action. Anti-Causalism about Reasons would hold that the explanatory reason for an action cannot be the cause of the action. It is a good argument for Causalism about Reasons to note that what someone would do in a situation is different than what

²⁸ Hempel augmented and made his theory more sophisticated by allowing that historical explanations were often only “explanation sketches” which would account for the poor predictive capabilities of the lessons of history. Except “Winter invasions of Russia always fail,” there are few laws in history capable of providing true predictions. Hempel also suggested statistical laws for historical explanations that would predict and explain the frequency of kinds of events.
someone did do. Lots of reasons rationalize a given action, but only one rationalizes and brings it about. The idea seems to have been in the air. John Passmore makes such a move:

Explanation by reference to a “principle of action” or a “good reason” is not, by itself, explanation at all…For a reason may be a “good reason” – in a sense of being a principle to which one could appeal in justification of one’s action – without having in fact the slightest influence on us,” (1958: 275).

Hempel has a paper targeting Dray’s contention that the popular method of explaining actions in terms of underlying reasons in the light of which the agent acts, cannot be construed as conforming to the covering-law pattern. Hempel writes that:

To show that an action was the appropriate or rational thing to have done under the circumstances is not to explain why in fact it was done… The presentation of an action as being appropriate to a given situation, as making sense, cannot, for purely logical reasons, serve to explain why in fact the action was taken, (1963: 102).

In his influential book Theory of Action Lawrence Davis\textsuperscript{29} writes:

While the explanation of Sam’s action aims at displaying the action’s intelligibility… explanation of… a “mere” event, aims at displaying its inevitability. Reasons-explanations and causal-explanations differ, then, in their aims and the battery of concepts that apply to them (1979: 85).

Davidson’s main point in his “multiple reasons” argument was that there are many reasons that would rationalize the action-to-be-explained. (Let’s call the given action-to-be-explained, “the primary explanandum.”) Regina may desire to annoy her neighbor by mowing her lawn in the early morning but it is conceivable that she actually mows it early in the morning for the reason that this is the most convenient time (cf. Mele 1997). There is an important difference between all those accurately rationalizing explanations that exist and the one that actually explains the primary explanandum. The difference, Davidson argued, is that

\textsuperscript{29} All quotations from Sandis (2006) whom I follow closely here.
the explanatory reason is the cause of the action. “When we offer the fact of the desire and belief in explanation,” Davidson says, “we imply not only that the agent had the desire and belief, but that they were efficacious in producing the action,” (1974: 232, italics mine). This is his multiple reason argument and it is meant to show that the explanatory reason is the reason that caused the action.

1.3.1 The Neo-Wittgensteinians’ Logical Connection Argument

It is hard to know whether or not, or to what extent, Wittgenstein himself was influenced by this European tradition of philosophy of social science, Geisteswissenschaft versus Naturwissenschaft, Erklären versus Verstehen. However, Wittgenstein’s discussion in Philosophical Investigations (and elsewhere) about reasons for action influenced the so-called Neo-Wittgensteinians’ “Logical Connection Argument,” which I will try to present as charitably as possible.

The anti-Causalists take it that a cause and its effect must be logically distinct entities. (They think they find this in Hume.) They then argue that a reason and the action it rationalizes are in some sense logically related, and thus not logically distinct. So the reason that rationalizes an action cannot be a cause of the action. The reason that rationalizes an action is logically connected to or dependent on it perhaps because they are interdefined. Thirst is defined as causing thirst-quenching behavior and thirst-quenching behavior is defined as what one does when thirsty. But causal claims must be unlike definitions. In order to be causally explanatory causal relations cannot be statements characterized as stating logical necessities. We cannot causally explain why Jones is a bachelor by asserting that he is an unmarried male. There is an explanatory generalization here “all bachelors are unmarried males,” but it is not a generalization ready to underwrite causal explanation. It’s a definition
that is making a logical connection between, or interdefining, “bachelor” and “unmarried male.” In fact, the supposed explanation is just a redescription of the explanandum. But nothing can causally explain itself. The logical connection argument says analogously that to give a reason for an action is just to redescribe the action, since that action is interdefined with the beliefs and desires that rationalize it.

1.3.2 Davidson’s “Descriptions Argument” Against the LCA

Against the logical connection argument Davidson gives an argument suggesting that both logical and causal relations are a matter of how we describe things. He points out that it is possible that an action and a reason can be (or, better, can be seen to be) logically related under one description but not others. For example, if we ask for what reason Oedipus gouged out his eyes (an action), it would not be acceptable to say, “because Oedipus lusted after the Queen,” since there is no logical or rationalizing relation there. However it would be explanatory to say, “because Oedipus lusted after his mother,” because if you find yourself in that state and you’re a Greek hero, it makes sense to poke out your eyes. Both rationalizing and causal explanations, Davidson says, are intensional in this sense: they do not preserve truth (or in our case epistemic adequacy of explanation) upon substitution of co-referential terms. Even though Jocasta, his mother and the Queen, are all the same entity, only under some descriptions does reference to Jocasta explain Oedipus’s gouging out his eyes.\[30\]

Similarly, causal explanation is intensional in that the very same event may be described as a hurricane and as “the event reported on page one of the News and Observer,” but only under the former description does reference to that event provide epistemic

---

\[30\] We have occasion to discuss “intensionality” again in Chapter Two’s review of the causation literature.
satisfaction as to why the Cape Fear basin is flooded. With these observations in mind, Davidson argued that it was possible that a reason and an action (as two events) could be both logically connected as well as causally connected, simply under different descriptions.

But don’t we want to say that relations between two entities hold no matter how described? If “Jocasta” and “the Queen” and “his mother,” are descriptions of the same entity, and that entity is related to Oedipus’s action, then that entity is so related no matter how described; we may just not come to know it under certain descriptions. Indeed this is Davidson’s position and he captures the “no-matter-how-described” phenomenon just discussed by calling causal relations “extensional” relations between events simpliciter.\(^{31}\) I’ll explain.

“The event reported on page one of the News and Observer caused the Cape Fear basin to be flooded” is a true singular causal statement but it is not explanatory, Davidson famously reminds us. This singular causal statement is true because there is indeed a causal relation between the event reported on page one of the News and Observer and Cape Fear being flooded. It holds true no matter how its relata are described. But causal explanation (and rationalization for that matter) is intensional and is only achieved under descriptions with the right intension.

The relata of the causal relation are events themselves and not events-under-a-description. However, causal explanation works or does not work depending on the given descriptions of the events. We return to the discussion of the causal relata below. The important point now is that causal and logical connections can obtain between the very same two events, contra what the proponents of the logical connection argument asserted.

\(^{31}\) What would an intensional theory of causation (and not merely causal explanation) look like? It would make mind-dependent the question of what was a cause and what was not. In this regard, see the criticism of Yablo in my Section 3.9.3 and 3.9.4 below.
1.4  New Arguments against Causalism

1.4.1  Reasons are not Mental Events nor Mental Events

Davidson (1963) is widely considered to have “demolished” (see e.g., Schueler 2003: 9) the anti-causalist position. Recently, however, the anti-causalist position has had something of a resurgence. The whole mental causation debate depends on the success of Causalism in the battle between Davidson and the Neo-Wittgensteinians.\(^{32}\) Perhaps the resurgence of anti-causalism can show that causalism is false; that would undermine the whole foundation on which the mental causation debate stands. So we had better take a look.

The thrust of the new anti-causalist argument is actually two pronged. They do not have an argument against causalism per se. Instead, they argue that reasons are not mental events nor mental events. That is, (1) they argue that a reason is not a psychological state of the person who has a reason for action. They argue that reasons are not inner mental events. They argue (2) that reasons are states of affairs out in the world. For instance, Dancy says “even the most cursory glance at the sorts of reasons we actually give, in explaining our actions or those of others, reveals that,” reasons are not psychological states (2000: 15). Bittner says we understand our actions primarily in terms of their place among things happening, not in terms of our attitude to things happening,” (2001: 110). Scanlon (2000) also has an externalist view such that reasons are states of affairs.

\(^{32}\) It is a fact that Kim overlooks Davidson’s work in philosophy of action as a possible source to ground the causality of reasons. Instead Kim stipulates that reasons are causes.
I think there is a debunking explanation for why we have the intuition with Dancy and others that reasons are not our psychological states. I’m going to explain away the intuition they seem to appeal to. And in the end causalism is going to stand.

Here’s why we have the intuitions we have. It is true that our own psychological states are rarely the thing that we appeal to in order to justify our actions. We do not say that my desire to help the child I believed was crying justified my helping the child. That the child is crying justifies my helping, not anything about me. Indeed, we appeal to states of affairs in the world. But it could be said that states of affairs in the world cannot be proximal causes of our intentional behavior. Only our beliefs about states of affairs in the world can cause us to act. This is most obviously the case in situations where we have a false belief about the state of affairs in the world and act on that false belief. In the non-veridical case, we would say that the belief – the false belief – caused us to act. So too in the veridical case we ought to say it is a belief that causes us to act. This belief represents the state of affairs; it has the state of affairs as its content.

Furthermore, notice, when we speak of action by referencing reasons qua inner causes it is not so counterintuitive to say that my belief she was crying and my desire to help caused me to help. Going the other way, the state of affairs that she was crying did not directly cause me to help (not without my believing this was the case), but is more comfortably talked about as a reason. These observations betray a tension between the justifying aspect and the causal aspect of reason explanations. In the former, I think, we have “one thought too many” if we take our own mental states as justifying helping the child. Mental states are better thought of as causes. In the latter, our sense of causation excludes states of affairs, which cannot have an impact on us without causing a belief in us via our perceptual apparatus.
Indeed, states of affairs function better as reasons. But any reason has to do double duty as justification and efficient cause.

After all, in the framework we’ve set up so far, an agent acts only if the reason to act includes both cognitive and conative aspects. Someone may desire a beer but not know that there is beer in the fridge. And someone may know that there is beer in the fridge but desire something else. If in either case one didn’t get up and go to the fridge at all we’d understand why. What's relevant to this discussion is that one needs to be in the proper state of mind with respect to the reason-as-states-of-affairs in order for the latter to be an explanatory reason.

But it is not enough to believe or have a representation of the state of affairs in order for it to be a consideration for or against acting. One’s representation needs to be accurate. Otherwise you don’t have a reason. Is that right? Not entirely. You would not have a justificatory reason but you would still have an explanatory reason. In the end, the mental causation debate is only talking about the causal status of actually correct and adequately explanatory explanations of behavior, made in the intentional idiom. So in the mental causation debate we are assuming a good case. There’s no point in proving mental causation for purported causal explanations that do not actually reference actual causes, or causal explanations that fail to adequately explain in virtue of positing non-veridical states of affairs. There’s lots of ways for a causal explanation to go wrong. We are interested in the metaphysics of good explanations. This will have further relevance later.

§

33 I do not mean “justificatory” in the sense in which a justified false belief remains justified. I mean “justificatory” in the sense in which some reasons justify and other do not justify a certain action.
1.5 A Euthyphro Question Regarding Reasons and Causes

In the following sections, I describe a number of theses that I believe Davidson endorses in the philosophy of action. However, if it could be shown that Davidson does not endorse these theses, I would still defend them. If he didn’t endorse them, he should have and I will.34

The importance of the collection of theses will become apparent only in Chapter Four where it will be used as a defense against the so-called “Identification Problem,” which is the problem of identifying which tokens are identical in a token identity theory.

1.5.1 The “Strict Qualification” Thesis

In this section, I’m going to argue for the truth of what I call the “strict qualification” thesis. This thesis says that the only reasons we are interested in, and the only reasons Davidson was interested in, are those reasons that are actually explanatory.

Davidson opens his (1963) this way: “What is the relation between a reason and an action when the reason explains the action by giving the agent’s reason for doing it?” (3). “A causal relation,” Davidson answers. But his answer and the implications thereof can be misunderstood. The first misunderstanding arises with respect to just what the question is asking. Davidson is not asking broadly about the relation between reasons and actions. Only a subset of, or a certain kind of, reasons interest him. We need to recognize the strict qualification inherent in the clause, “when it explains the action by giving the reason.” Truth be told, Davidson limits his interest in reasons to those appearing in successful explanations of intentional action, reasons that actually explain.

34 In other words, I have two goals, Davidson interpretation and correcting the mental causation dialectic. I think they go together but they can be considered separately.
To see how, recall that Davidson foregrounds the fact that citing a reason is not the same thing as citing the reason for an action. He asks us to imagine situations in which there are multiple reasons that justify the action in question. This is the multiple reasons argument that I mentioned earlier when discussing Davidson’s argument for the causal requirement. The thought is that there might be lots of reasons that rationalize or make intelligible a given action. But the reason, as opposed to simply a reason (the explanatory reason as opposed to a merely justifying reason), is the one that actually caused, produced or brought about, etc., the action in question.35 “When we offer the fact of the desire and belief in explanation,” Davidson says, “we imply not only that the agent had the desire and belief, but that they were efficacious in producing the action,” (1974: 232).

Here’s my argument for the “strict qualification” reading of the first line of Davidson (1963). Reasons appear in successful and unsuccessful explanations of intentional action. Only the explanation that appeals to the efficacious reason, we must admit, will count as successful. So we should focus our attention on only those justifying reasons that are actually explanatory if we wish to discover the relation between the reason and the action it explains, rather than any old justifying reason. If we want to know the relation between a reason and the action that it explains then we need to look at only reasons that appear in true and epistemically adequate explanations. This is what we should do and this is what Davidson does.

However, this raises the question of how we know what was the reason for which an agent acted. How do we come to know the explanatory reason?

---

35 He later amends this to “caused in the right way” in order to account for deviant causal chains. See Davidson (1973). In any case, deviant causal chains do not occur with respect to intentional action done for reasons.
1.5.2 The Epistemology of the Reason

In Davidson’s framework, the epistemology of the reason is essentially our standard folk psychological practice of intentional behavior explanation through mental state attribution, including the principle of charity. Roughly speaking, a reason explanation explains by positing or merely hinting at an effective mental-event that is related to the explanandum by some loosey-goosey lore.

The reason that is offered as explanatory is in fact attributed or ascribed to the agent by the explainer with evidence that is severely underdetermining of the agent’s possession of said reason. Some philosophers think this is perhaps analogous to the way the electron and other unseen theoretical posits are used in physical science. They are adopted for their explanatory power in relation to choice explananda. But it is always possible that new phenomena will be observed that make another posit more likely and the former one much less likely.

In any case, a reason explanation posits an agent’s reasons for the given action by assuming the agent is like the explainer or follows relatively rational behavior-guidelines and participates in the same form of life, or cultural ways and mores, as the explainer; and by assuming lots of detail about the context of the action and what has come before. In this way, reason explanations make actions intelligible. This last sentence was the rallying cry for Davidson’s *anti-causalist opponents* when his (1963) was published. But here too is another misunderstanding of Davidson. As far as the epistemology of the reason goes, it is not widely recognized that Davidson does not actually (mean to) advance the discussion beyond his

---

36 All I need to show, I think, is that Davidson does not hold that reason explanation has the DN form. I regard it as a family dispute whether Davidson or I need to adopt a Quinean view, a Davidsonian view, a McDowellian view or a Brandomian view.
original interlocutors, the Neo-Wittgensteinians, on this point. Davidson’s rhetoric may suggest he has refuted the Neo-Wittgensteinians’ claim that reason explanations are not causal explanations. But on this point – on what might be called the methodological point about the structure of reason explanations – they agree that the strict DN explanations in physical science differ from the hermeneutic explanations of folk psychology. Reason explanations for Davidson and the Neo-Wittgensteinians rationalize the behavior according to our essentially normative notions of what would be appropriate for an agent to do who had the beliefs and desires that we holistically attribute to her in the context. (Then that reason is designated as the cause, according to Davidson.)

Again: it is a misunderstanding to take it that Davidson overturned the anti-causalist hegemony by arguing that reason explanations had the form of DN explanations. But this misunderstanding persists. For an example, Anthony (1989) writes that a reason explanation gets its “explanatory force” in virtue of a covering law (23). Sandis (2006) lists Davidson with other commentators who take reason explanations to be DN explanations:


Many will disagree with my claim about Davidson’s position on how folk psychology explains. My view is certainly in disagreement with a popular misconception. But Davidson did not think folk psychological explanation was DN, as some philosophers persist in thinking.

37 Sandis (2006: 2) is clearly wrong to place Davidson in that group. He does so perhaps because Sandis confuses epistemological and metaphysical matters.
It is clear that we accept as successful or true many explanations in folk psychology that do not have the same form as physical explanations; their explanatory structure is interpretative, an exercise in hermeneutical *Verstehen* appropriate to an object of the various *Geisteswissenschaften*, just as it is for the Neo-Wittgensteinians. The difference might be that Davidson seems to want to make that type of explanation fit into a naturalistic metaphysics, whereas his opponents may have felt they had come to the limits of naturalism or at least the limits of scientific explanation.

1.5.3 The Priority of the Phenomena Thesis

Furthermore, the reason is what you would naturally be interested in if you conceived yourself as engaged in one traditional type of metaphysical inquiry. The way of doing metaphysics I have in mind is, as I’ve said, a matter of asking and answering a “how-possible” question of an actual phenomenon. For example, we come to have knowledge in such and such a way; how is it possible that we come to knowledge in that way? This kind of account of our actual epistemological practices is meant to establish metaphysical legitimacy for those practices, which are of extreme importance in our self-conception as agents.

When Davidson asks, “What is the relation between a reason and an action when the reason explains the action by giving the agent’s reason for doing it?” he is engaging in a project we might call a metaphysical project and he is suggesting that it is distinct from projects we might call epistemological. It is a matter of epistemology, let’s agree, to come to know the reason for which an agent acted; it is a matter of metaphysics, let’s say, to make the inquiry into the relation between explanatory reasons and the actions they explain. According to Davidson, one condition of possibility for a reason to be an explanation of an action is that the reason be the cause of the action.
In the project proposed by “the relation question,” Davidson, as we said above, limits his interest to just those reasons that actually explain the action. (This is what we called the strict qualification thesis.) Once we have isolated the reason, i.e., once we have a successful explanation in hand, only then is a different question – the metaphysical question – raised: how is it that that reason can be a cause? That is to say, Davidson does not ask the metaphysical question, “How could a reason be a cause,” until he has got a fixed reason in hand. It gets fixed as our object of metaphysical inquiry by appearing in a successful folk psychological explanation, which, because it is successful, assumes a legitimate epistemology of the reason. Only after this stage is the metaphysical “how-possible” question asked. The problem is to say how the phenomenon of mental causation, which is an actuality, is even a possibility. As long as we are not eliminativists and we continue to consider epiphenomenalism a liability, the mental causation debate is about “saving the phenomena.” Establishing “mental causation” or achieving the metaphysical project called “mental causation,” is meant to metaphysically underwrite our folk psychological explanatory practices, which have priority.

1.5.4 The Euthyphro Question Regarding Reasons and Causes

Anytime you have two things, or two properties of one thing, and you want to find some relation of priority between them, then you can ask a Euthyphro question about the direction of priority. In this case, we have a mental event that is both the reason and a cause. We ask: Is it a cause because it is the reason or is it the reason because it is the cause? Answer: It is a cause because it is the reason. It is not the case that it is the reason because it is the cause. Let

38 The additional salience of my treating it this way will become apparent later when these observations are used as a defense against the so-called “identification problem.” See Section 3.10 below.
me be clear. I am not saying that some kind of epistemic Euthyphronic direction goes from reason to cause while some metaphysical Euthyphronic direction goes from cause to reason. I am saying that there is a Euthyphronic direction from epistemology to metaphysics. In other words, the reason is the cause in virtue of being the reason not the other way around. It is our best epistemology whose deliverance is the reason which imputes to us the metaphysical commitment that said reason is also the cause.

Why answer in this manner? I believe the *Priority of the Phenomenon Thesis* provides the answer to this Euthyphro question. The *Priority of the Phenomenon Thesis* says we first come to know the reason for which someone acted and we come to know it through “the epistemology of the reason,” as I have called it. So we may say that this event is the cause it is in (Euthyphronic) virtue of being the explanatory rationalizing reason it is.

Does anything interesting follow from this? Is anything at stake? Yes. This Euthyphronic order of explanation will be relevant any time a proponent of another theory tries to formulate causal criteria by which to identify the reason for an agent’s action. It’ll show that strategy to be wrong-headed. You’ll be getting things backwards if you try to discover the reason for action by seeking the cause of the action. The order of explanation goes the other way such that we may seek the cause of an action only through seeking the reason for the action.

So, Kim gets it exactly backwards when he says, “Reasons explain actions in virtue of being their causes,” (2006: 176). That is just wrong. Reasons are causes in virtue of epistemically explaining their actions. For another example, Campbell says: "Although Davidson acknowledges that rationality is an ideal that shapes our mental ascriptions and folk-psychological explanations, he has argued that we should not regard the rationalizing feature of reason explanation as the source of its explanatory force," (Campbell, 2005: 445)
This author is saying its being a cause is the source of a reason’s explanatory force. However, the author is wrong. Explanatory force comes from intensional matching. Truth comes from whether the causal relation is actual between the posited cause and the effect to be explained. (See Section 4.3 - 4.5 for more on explanatory force.)

The epistemology of the reason is not like an epistemology of causes. In the latter, we, assuming that every event has a cause, seek the cause of the given effect event using tried and true methods of causal discovery. Whereas, in the former we attribute mental states according to a normative system (described above in Section 1.5.2) and then designate that reason as the cause. The epistemology of causes of events is structured differently than the epistemology of reasons for action. A point that remains from our history of the intellectual milieu before ARC.

Even if it is the case that there are times when the epistemology of the reason looks like an epistemology of causes it is nevertheless still the case that we operate in a certain order. First we determine the reason using the epistemology of the reason, whatever it is, and then second, we designate (metaphysically-speaking) that reason as the cause. It does not go the other way around.

§

1.6 Conclusion

In this chapter we have discussed the history of the philosophy of Geisteswissenschaften and the philosophical milieu at the time Davidson wrote “Actions, Reasons and Causes,” (1963). I have argued that Davidson took it as a necessary precondition of genuine reason explanations that the reason appealed to in the explanation actually be the cause of the action explained. I
have registered my agreement. I have also outlined how this view fits into a systematic way of thinking, which includes a certain relation between epistemology and metaphysics. The *Strict Qualification Thesis* and the *Priority of the Phenomenon Thesis* highlight the direction of relation between metaphysics and epistemology.

I posed a Euthyphro question about reasons and causes and argued on the basis of those two theses that reasons are causes in virtue of being reasons not the other way around, i.e., it is not in virtue of being causes that reasons are causes. This will be relevant in Chapter Three where we defend anomalous monism from the so-called “identification problem.”

Next, in Chapter Two, we make a survey of the literature on causation and causal laws. For the sake of argument with his critics, I adopt a Davidsonian version of the nomological account of causation. But we should also explore how each account of causation fares with respect to the problem of mental causation. Perhaps an account of causation can avoid the exclusion of mental properties by physical properties even though Esfeld (forthcoming) says, “The problem of mental causation and the argument for token identity are independent of the stance that one takes in the metaphysics of causation,” (1).

It will have been worthwhile exploring the vicissitudes of causation and causal laws when, in Chapters Four and Five, we defend anomalous monism and go beyond it to reconsider the relation between causes and law.
CHAPTER TWO

CAUSATION AND CAUSAL LAWS

It is easy enough to say that reasons or mental events are causes. But what are causes? What is causation? Until we have some understanding of answers to these questions we cannot really know what it means to say that reasons are causes.

What we will see upon making a survey of the philosophical literature is that philosophers make, as I do, a strong distinction between metaphysical causation and epistemic causal explanation.

2.1 Introduction

We recognize causation any time we drop a glass and the impact with the floor causes the glass to break. Consider that smoking causes cancer. Striking the cue ball with the stick caused the ball to move. Philosophers say a cause makes something happen, brings something about or makes a difference; and “the difference it makes,” says David Lewis, one of the most influential thinkers on causation, “must be a difference from what would have happened without it,” (1973: 160-161). When we say that the financial crisis could have been avoided, we mean in part that there was an event, or were events, without which the crisis would not have happened. These are the causes of the crisis.

An analysis of causation aims to explain what it is for two events to be related as cause and effect – what must be the case, metaphysically speaking, for two events to be so
related. This kind of inquiry is different from the inquiry by which we come to know that two events stand in the causal relation, which is an epistemological project.

There are many kinds of analyses of causation in the literature. Despite of, or perhaps because of, the fact that causation is one of the most important and widespread issues in philosophy, there is little agreement about the correct approach. In a recent anthology, Helen Beebee et al. writes, “No theory has won univocal support in the literature. The discussion consequently is a little bit of a hodge-podge,” (2010: 1).

There are nevertheless issues that reappear in many of the discussions below. As we will see, differentiating between what’s causal and what’s only apparently causal is a difficult task for many theories.

2.1.1 Causation versus Correlation

Often the first thing to say about causation is “Correlation is not causation.” It’s true that the difference between genuine causal connections and mere correlations is central to an understanding of causation. However, the saying, “Correlation is not causation,” is incomplete. It would be more accurate to say, “Correlation is necessary but not sufficient for causation.” Two events, c and e, may be constantly correlated and not yet causally related. For instance, they may have a common cause in a third event, b.

Consider this example: sleeping with the lights on is strongly correlated in adults with waking up with cephalalgia, or headache. One might conclude, epistemologically speaking, that sleeping with the lights on caused the headache. However, if we can imagine a plausible common cause then this inference will seem illegitimate. After all, it is possible excess alcohol consumption caused both the headache and sleeping with the lights on. Making this epistemological error – the “common cause error” – amounts to mistaking an effect for a
cause. Metaphysically speaking, for the assertion that Event 2 is a cause of Event 3 to be true, there must *not* be another event, Event 1, causing both Event 2 and Event 3. In Figures 1 and 2 the dotted arrow represents an apparent connection, which is in fact no connection at all.

Consider a more concrete example. Nietzsche writes of the strange case of Cornaro, who thought that his meager diet caused longevity. Nietzsche says, "The worthy Italian thought that his diet was the cause of his long life, whereas the precondition for a long life, the extraordinary slowness of his metabolism, the consumption of so little, was the cause of his slender diet,"\(^{39}\) as well as of his longevity.

\(^{39}\) For more see Leiter (2002: 156-157).
2.1.2 Token and Type Causation

Consider these two statements:

(a) The short circuit caused the fire.
(b) Short circuits cause fires.

The former is an example of what philosophers call token or singular causation – an individual event causing another individual event. The latter is an example of type or general causation – causation between properties or kinds of events. Many philosophers hold that token and type causation are different kinds of causality. Sober (1985, 1986) and Eells (1991), for example, argue for theories of type and token causation that differ considerably. Cartwright (1989) and Hausman (1998), on the other hand, hold that type causation depends upon the more fundamental token causation. It is possible to hold the view that type-level relations are fundamental (see Tooley 1987). We proceed talking usually about token causation.

2.1.3 Causation versus Causal Explanation

Consider these two token causal claims:

(a) The hurricane caused the building to collapse.
(b) The event reported on page 5 of the Times caused the building to collapse.

Assuming (a) and (b) are both true, only (a) is explanatory; (b) is not, even assuming, as we do, that the hurricane is the very same thing as the event reported on page 5 of the Times.

True singular causal statements are called “extensional” because they remain true under substitution of all true descriptions of the relata (the things related by the causal relation). The Hurricane caused the building to collapse and so did the event on page 5 of the Times because they are one and the same thing. On the other hand, causal explanation can be called “intensional” because there are salience requirements on the proffered descriptions of the causal relata in order for the explanation to be genuinely explanatory.
If we ask again for an explanation of why Oedipus gouged out his eyes, recall that it would not be explanatory to say, “because Oedipus lusted after the Queen.” That description is somehow not relevant or salient. However it would be explanatory to say, “because Oedipus lusted after his mother,” because when you do that and you’re a Greek hero, it makes sense to poke out your eyes. Causal explanations, Davidson (1967) says, are intensional in this sense: they need not preserve explanatory adequacy upon substitution of co-referential terms (differing terms that nevertheless have the same referent). Even though Jocasta, his mother and the Queen, are all the same entity, only under some descriptions does reference to Jocasta explain Oedipus’s gouging out his eyes. Nevertheless, it’s true that he lusted after Queen Jocasta/his mother just as it is true that the event reported on page 5 of the Times caused the building to collapse. Each is true, but not explanatory. Causation and causal explanation are different insofar as it is possible for a sentence to be non-explanatory while still being causal.

Interest-relative, mind-dependent features of a situation are relevant to whether an explanation is more or less explanatory. Lewis’s (1979) theory of causal explanation includes the idea that an explanation describes (or gives information about) a part of the causal history of the event to be explained. Which part of the history is relevant will change relative to the context of the inquiry. Imagine a car crash. The road engineer will find the improper grading for the arc of the road’s curve to be explanatory. The policeman will find the driver’s speed to be explanatory. The driver himself, having safely navigated that curve at that same speed many times before, may blame the fresh rain on the road.

As Lewis says:

---

40 This example is found in Lewis (1986) and Kim (1998).
If someone says that the bald tire was the cause of the crash, another says that the driver's drunkenness was the cause, and still another says that the cause was the bad upbringing which made him so reckless, I do not think any of them disagree with me when I say that the causal history includes all three. They disagree only about which part of the causal history is most salient for the purposes of some particular inquiry, (1979: 215).

On the other hand, with respect to causation per se it is often considered a desideratum that causation be mind-independent. “Causal relationships are features of the world: they are ‘out there’ in nature. By contrast, explanation is an activity carried out by humans and conceivably by some other animals, having to do with the discovery and provision of information, information based on causal relationships,” (Woodward 2003: 23).

2.1.4 Causation, Possibility and Necessity

As we have observed, one way to think of causation is to think that a cause necessitates its effect. Given the cause you’d necessarily get (it would be impossible not to get) the effect. Causation is, thereby, intimately bound up with possibility and necessity.

Possibilities are things that could have happened, even if they never take place. Something is necessary when it is impossible that it fail to happen. Necessity and possibility are, thereby, like two sides of a coin. Furthermore, there is nomological or natural necessity and possibility and then there’s logical necessity and possibility. For example, a round square is logically impossible while faster than light travel and 3 cubic mile gold cubes are naturally impossible but logically possible. Insofar as the concept of causation involves necessity, it is natural necessity and not absolute necessity. That is to say, the relation of necessity between cause and effect is a relation of necessity only in this world and not in every possible world.

2.1.5 Theories of Scientific Explanation
Philosophers disagree about the extent to which philosophical theories of scientific explanation should make fundamental reference to causal concepts. The logical empiricists tried to do without causal concepts. More recently, problems with the deductive-nomological (DN) model and other theories suggest that theories of scientific explanation require causal concepts. We may briefly explore three popular theories of scientific explanation for their relation to a discussion of causation. There is the DN model as developed by Hempel & Oppenheim (1948) and Hempel (1965); the statistical relevance model of Salmon (1971); and the unificationist models as developed by Friedman (1971) and Kitcher (1981, 1989). The DN model of explanation has it that explanations are deductive arguments. The premises of the argument are the explanans (the thing doing the explaining) and the conclusion of the argument is the explanandum (the thing explained). The explanans contains at least one law of nature and it must contain it essentially such that the inference would not go through were it left out. An explanandum is explained if it is possible to deduce it from sentences about the antecedent conditions together with the law of nature.

Among many well-known problems, the DN has difficulty accounting for what we usually take to be the asymmetry of many scientific explanations. For instance, we normally take there to be an asymmetry pointing from the explanans to the explanandum: we derive the length of a shadow from the height of a flagpole and the angle of the sun (plus a law). The DN model captures this explanation perfectly. But it just as perfectly goes the other way: it purports to explain the height of the flagpole in terms of the angle of the sun and the length of the shadow (plus the law). But this is a counterintuitive result. DN is allowing too many things to count as explanations. We do not normally explain the height of a flagpole by referencing the length of its shadow. Normally, it goes the other way around. Kitcher (1989)

41 For a comprehensive discussion of the main problems for DN see Salmon (1989).
argues that what DN is missing is the asymmetry of causation. He diagnoses DN’s symmetry to be a result of the empiricists’ Humean desire to do without “mysterious” causal powers.

The central tenet of the statistical relevance model is roughly that statistically relevant features are explanatory features. Statistical relevance, then, is based on the conditional dependence of an event on other events. It says that given a class A, an attribute C will be statistically relevant to another attribute B if and only if \( P(B | A \& C) \neq P(B | A) \). In other words, C is statistically relevant to B if and only if the probability of B given A and C is different from the probability of B given only A. For an example, let’s say that whoever is infected by streptococcus (S) and takes the medicine penicillin (M) has a very high probability of quick recovery (R). This we would represent as follows: \( P(R | S \& M) \) is very high. We compare this to the probability of recovery given streptococcus without taking penicillin, or \( P(R | S) \). So the equation above says that if the probability of a certain outcome is different (higher or lower) given the streptococcus than the probability of an outcome given the streptococcus plus another factor (the medicine) then the medicinal factor is statistically relevant to the outcome.

One problem for the statistical model of causal explanation is a classic kind of trouble: namely with common causes and epiphenomena. A perfect description of statistical relevancies may not yet capture the actual causal relationships at work (Salmon 1971). For example, imagine Event 2 is causally relevant to Event 3 while Event 1 is an epiphenomenon necessitated (represented by the double arrow in Figure 3) by Event 2. In this situation, Event 1 and Event 2 have the same statistical relevance with respect to Event 3, even though only Event 2 causes Event 3. As counterfactual dependence alone is not causation, neither is statistical relevance alone causation, for there may be a common cause.
Friedman (1974) and Kitcher (1989) have put forth a unificationist account of scientific explanation. The essence of the unificationist account is that “scientific explanation is a matter of providing a unified account of a range of different phenomena,” (Woodward 2009b: Section 5.1). There is intuitive support for unification as part of what scientific explanation does. “Successful unification may exhibit connections or relationships between phenomena previously thought to be unrelated and this seems to be something that we expect good explanations to do,” (5.1). Furthermore, the unification of theories has clearly played a role in the history of scientific progress. “Paradigmatic examples include Newton’s unification of terrestrial and celestial theories of motion and Maxwell’s unification of electricity and magnetism,” (5.1). “The key question, however, is whether our intuitive notion (or notions) of unification can be made more precise in a way that allows us to recover the features that we think that good explanations should possess,” (5.1).

The purpose of this literature review is in part to demonstrate the striking distinction here between metaphysical causation and epistemic causal explanation. We looked at theories of scientific explanation. In the positivist tradition said theories were attempted without essential reference to causal concepts. But it has become apparent, to some in any case, that the theories of scientific explanation do not succeed without appeal to causal concepts. We now turn to a discussion of causation per se.
2.2 Causal Accounts and Analyses

An account of causation is meant to provide some understanding about causation. An analysis of causation is a specific kind of account of causation in which causality is analyzed or broken down into more fundamental, non-causal, parts. Analysis is thereby a reductive project, reducing causality to something having to do with possible worlds, or with laws of nature for example. The opposite of a reductive account would be a realist account.

We will consider each theory of causation with respect to its applicability to the mental causation debate.

2.2.1 Regularity and Nomological Theories of Causation

A standard regularity theory will start by saying: c caused e if and only if c preceded e, and events like c are always followed by events like e. This naïve definition captures the intuitions behind sayings like “The same cause is always accompanied by the same effect” and “If no cause is present, no effect occurs.” In a bit more detail:

c causes e if and only if:
1. c precedes e in time
2. c and e are contiguous in space
3. c is a C-type of event
4. e is a E-type of event and
5. events of type C are regularly followed by events of type E.

There is a Humean tradition of regularity theories in which the regularities in (5) are merely de facto regularities without the aspect of necessity often associated with our idea of causation.
Another kind of regularity theory is the nomic subsumption view, which says that when \( c \) is a cause of \( e \) this is so in virtue of a relation of necessity or a strict law between \( C \)-type events and \( E \)-type events. So (5) would not be merely de facto regularity but nomic necessity. The difference between a de facto regularity theory and the nomic subsumption theory raises the question of how to account for the aspect of necessity that makes the difference. For this we need a theory of the nature of laws. Below in Section 2.3 we discuss Armstrong's Non-Humean account and Lewis’ Humean account of causal laws.

2.2.2 Mackie’s INUS Conditions

Above we have sometimes spoken of “the cause.” But do not most events have multiple factors that bring them about? Once you start thinking of causes – plural – instead of the cause, you see there is a set of causes and conditions that were jointly sufficient but not necessary for the effect. For example, a fire can be started by a short circuit but only in the presence of oxygen. Together those two states amount to a sufficient but not necessary condition for a fire. That is, they are enough to cause a fire, but the fire could be caused in another way, say by striking a match. With respect to the cause of the fire Mackie (197x) offered a proposal he called INUS conditions:

In this case, then, the so-called cause is... an insufficient but necessary part of a condition which is itself unnecessary but sufficient for the result (414).

So the short circuit is itself a necessary part of the set of conditions which were sufficient but not necessary for the fire. Also the short circuit is with respect to the set of conditions of which it is a part, insufficient alone for the fire. It needs the rest of the set of conditions. So the cause, the short-circuit, is an insufficient but necessary part of a set of conditions, the set itself being unnecessary but sufficient for the result – INUS.
A naïve regularity theory might take causes to be necessary and sufficient for their effects. Mackie's view is therefore a sophisticated regularity view. With all its sophistication it does not appear to be the case that the problem of mental causation would not arise on this theory. That is, Kim would probably say that even given Mackie’s INUS conditions we would still get a conflict between causes, mental and physical, where the physical excludes the mental. Or Kim might say that the physical property meets the INUS conditions and the mental property does not.

2.2.3 Counterfactual Theories of Causation

Counterfactual theories of causation abound today. But there was a time when they were not in favor. Regularity theories were favored for a variety of reasons, including the remaining influence of logical positivism. But, additionally, counterfactual accounts of causation had been unpopular because such a program looked like explaining a difficult concept with another difficult concept since it was controversial just what made counterfactuals true. Counterfactuals were things needing explanation, not things to appeal to in an explanation. Lewis’ possible world semantics for counterfactuals, however, gave one more or less plausible account of the truth conditions of counterfactuals. Indeed, Lewis’ counterfactual theory was intended to unseat regularity theories, whose “prospects look dark,” according to Lewis (1967: 160).

The regularity theory says, “A cause is defined (roughly) as any member of any minimal set of actual conditions that are jointly sufficient, given the laws, for the existence of the effect,” (Lewis 1967: 159). He goes on to note problems with the regularity view. Lewis notes that “c might rather be an effect of e: one which could not, given the laws and some of the actual circumstances, have occurred otherwise than by being caused by e. Or c might be
an epiphenomenon of e: a more or less inefficacious effect of some genuine cause of e,” (xyx).

Lewis took his counterfactual account as both building on a part of what Hume said and improving on regularity theories. Lewis argued that chains of counterfactually dependent events are sufficient for causality. Counterfactual dependence of e on c is the thought that had c not occurred, e would not have occurred. As Lewis says (as noted above), “We think of a cause as something that makes a difference, and the difference it makes must be a difference from what would have happened without it,” (557). So a cause, c, of an event, e, makes a difference to the event, e, coming to be. That is, without c we would not get e. We can say with Lewis, “e counterfactually depends on c iff, if c had not been, e never existed,” (xy). The counter to fact situation “if c had not been, e never existed,” needs to be evaluated. According to Lewis, “If P were the case, Q would be the case” is true just in case Q is true in the closest P-world. In other words, just in case Q is true in the world in which P is true and that, apart from P’s being true there, is as much like the actual world as possible.

Lewis represents counterfactual dependence as ~c □→ ~e and asserts it as Hume’s meaning in his (Hume’s) second “definition” of causation: “Where if the first object had not been, the second never existed,” as we noted above. For an example at this stage of Lewis’s theory, consider a fully functional barometer that accurately measures atmospheric pressure. If the barometer is working correctly then the reading we make depends counterfactually on the atmospheric pressure. That is, if the pressure goes up, then the barometer’s reading goes up; and had the pressure not gone up, then the barometer’s reading would not have gone up. Having eliminated the possibility of any common cause, we would say that counterfactual dependence of this kind bespeaks casual dependence of the barometer reading on the atmospheric pressure.
It has been widely noticed, however, that such kind of counterfactual dependence is not necessary for causal dependence. That is, there could be causally related events that were not counterfactually dependent. This is so in the famous so-called potential pre-emption cases.42 A potential but pre-empted cause c* is an event that would have led to an effect e, but it is such that its occurrence is blocked or pre-empted by the occurrence of another event c, which nonetheless causes e. So, for instance, suppose that two men Mr White and Mr Pink, independently of each other, are set on killing Osama bin Laden (Psillos 2002: 97). The case is such that they both make similar arrangements to kill Osama:

Mr White fires his rifle; the bullet takes its course and strikes Osama in the head. Osama dies. Mr Pink was ready to fire his rifle, and had he fired it, given his position, his shooting skills, and so on, the bullet would have also struck Osama in the head, leading to his death. But Mr White’s shot scares off Mr Pink, who then flees the scene. Mr Pink’s shot is a potential alternative cause of Osama’s death: it was pre-empted by Mr White’s shot, but had it not been pre-empted, it would have caused Mr Smith’s death.

Cases such as these demonstrate that there can be causation without counterfactual dependence because Osama’s death is not counterfactually dependent on Mr White’s shot. It’s not because if Mr White had not fired his shot then Mr Pink would have fired his own and Osama would have died anyway.

The Lewisian notion of *chains* of actual counterfactually dependent events deals with cases of pre-emption like this. There is, if we introduce plausible intermediate events, such a chain between Mr White’s shot and Osama’s death, while there is not such a chain between Mr Pink’s possible shot and Osama’s death.

---

42 We are only discussing cases of early, asymmetrical pre-emption and will not be covering the more recondite cases of late pre-emption and symmetrical overdetermination.
Take an intermediate event d (e.g., that the bullet passed in between two trees) between Mr White’s shot c and Osama’s death e. Then e is causally dependent on d and d on c. And <c,d,e> is a causal chain and it is in virtue of this chain that c caused e.

There is no such chain between Mr Pink’s shot and Osama’s death.

There is a problem here however. One may doubt that the effect e counterfactually depends on the intermediate event d. That is, one might think this way: if d had been absent, c would also have been absent, since d is there because c caused it to be; but then c* would have occurred and caused e. In order to block this, Lewis denies the counterfactual, “if d had been absent, c would also have been absent.” He calls such a thing a “backtracking counterfactual,” because it would make a temporally prior event be counterfactually dependent on a temporally posterior event.

The counterfactual approach to causation may hold promise as a solution to the problem of mental causation. Some authors have suggested that overdetermination (see Section 3.4.4 below) is an option provided one endorses a Humean, counterfactual theory of causation (see Bennett 2003, Loewer 2007, and Harbecke 2008). Esfeld (2010) argues that “whatever stance one takes in the metaphysics of causation, one faces the problem of mental causation,” (Section 1). I compare the counterfactual account and the nomological account in Section 2.3.1 below.

### 2.2.4 Probabilistic Causation

The program of probabilistic causation is intended to characterize causes as events that change the probability of their effects. Here is Reichenbach’s early formulation:

\[
C \text{ is a cause of } E \text{ if } P(E | C) > P(E | \neg C)
\]

43 In Section 3.8.2 I consider Marras’s (1993) counterfactual account of causal relevance.
In other words, C is a cause of E if the probability of the effect given the cause is greater than the probability of the effect absent the cause.

As we have seen, it is important that any account of causation allow us to differentiate a purported cause that is merely correlated with an effect from an actual cause of the effect. The probabilistic account can make this differentiation in its own terms. Spurious correlations are discovered to be spurious when the purported cause is “screened off” by the common cause. D screens C off from E just in case:

i. \( P(E|DC) = P(E|D\sim C) \)

ii. \( P(E|\sim DC) = P(E|\sim D\sim C) \)

This means that the probability of the effect given the screener off and the purported cause is equal to the probability of the effect given the screener off without the purported cause. Plus the probability of the effect given the purported cause absent the screener off is equal to the probability of the effect given neither the purported cause nor the screener off. And these two jointly sufficient conditions for screening off strike us as intuitively for the following reasons: if a purported cause is screened off then its purported effect would be unchanged from a situation in which the cause did not exist and the screener off did. Plus, if a cause is screened off then it makes sense that effect would be the same if there were neither screener off nor cause – the same as if there was the cause without the screener off.

Consider an example. In the barometer case, the low atmospheric pressure D screens off the barometer C from the storm E i.e., makes C statistically irrelevant to E. We can understand this as follows. The probability of the storm is the same in the case in which we have the pressure with the normal barometer reading as it is in the case in which we have the pressure but the barometer reading fails and reads a false value. Plus, another condition is
met when the pressure $D$ screens off the barometer $C$. The probability of the storm were there, counter to fact, a high atmospheric pressure with the barometer still reading low is the same as the probability of the storm were the atmospheric pressure high while the barometer reads high too. We see thereby that spurious correlations are those mere correlations in which the purported cause is screened off by the common cause.

We may keep the discussion of probabilistic causation short because it is not likely to be of much use in the mental causation debate unless there were knowable probabilities associating reasons with and actions. Hempel (1942) does make an attempt to provide a framework for using probabilities in historical explanation, but the project seems like a long shot in history as well as in action.

### 2.2.5 Process Theories of Causation

Hume challenges us to give a philosophical account of any connection between cause and effect. In order to meet Hume’s challenge, process theories introduce considerable conceptual novelty. For instance, Salmon (1974, etc) talks about so-called “processes” as primitives and dispenses with events. A process is anything that has a structure and maintains a unity even as it undergoes slight changes. It is continuous: a process cannot be represented as a series of discrete events. The continuity of the process ultimately links up cause and effect (Salmon 1984: 156-7) in the way Hume thought impossible. Whereas events are localized in space and time, processes “have much greater temporal duration,” (1984: 139). In the language of special relativity, a process is represented by a world line in a Minkowski diagram, while an event is represented as a point. Processes “are the mechanisms that propagate structure and transmit causal influence in this dynamic and changing world… they provide the ties among the various spatiotemporal parts of our universe,” (1997: 66). Objects
that persist through time, therefore, are processes. So is a propagating wave. A rolling ball. Turning gears.

We were just now discussing causal processes. There are also such things as pseudo (causal) processes. Differentiating the two is key. A causal process is one that transmits a mark and a pseudo causal process is one that does not. A mark is a modification to the structure of a process. For example, a moving car is a causal process while the shadow the car makes is not. Let’s say a process P is that which would remain uniform with respect to characteristic Q in the absence of other processes; and a process P is that for which Q would be manifest over an interval spanning both A and B (A not equal to B). Then a mark would consist of a modification from Q to Q* which has been introduced into process P by means of a single local interaction at point A and which is transmitted to B, (see 1984: 148). For example, a rotating beacon casts a white spot that moves around a circular wall. The spot is marked by interposing a red filter near the wall. This process is not causal. But we can put a red lens at the beacon which would be an interaction on our part which transmits its structure. As a result of this, the white spot turns red, by means of a single local interaction and remains so while it moves around the wall.

For another example of a causal interaction, consider the classic case of a collision between two billiard balls. Two processes, P1 with characteristic Q and P2 with characteristic R, interact at point S. It is a causal interaction and each are causal processes if P1 exhibits a modified characteristic Q* throughout an interval after S and P2 exhibits a modified characteristic R* throughout an interval after S. Salmon writes: “if two processes intersect in a manner that qualifies as a causal interaction, we may conclude that both processes are causal, for both processes have been marked (i.e., modified) in the intersection with the other and each process transmits the mark beyond the point of intersection,” (1984: 174).
It has been noticed that process theories assume physicalism. They explore, after all, the physical mechanism by which things are brought about. Process theories are therefore not useful in the mental causation debate because they’ve already decided the question about any competition between mental and physical for causal efficacy. It’s been decided in favor of whatever physical realization neuroscientists can tell us is responsible for thinking and mental experience.

2.2.6 Manipulationist and Interventionist Accounts of Causation
Once unpopular and much maligned, manipulationist accounts have recently been improved. Now interventionist accounts offer an exciting and promising account of causal claims. At the core of all manipulationist (including interventionist) accounts is the thought that “if c is a cause of e then if I can manipulate c in the right way this should be a way of manipulating or changing e,” (Woodward 2009: 2). This thought is common to the early manipulationists including Gasking (1955), Collingwood (1940) and von Wright (1971); to recent versions, including Menzies and Price (1993); and to recent interventionist formulations, including Woodward (2003), Pearl (2000) and Spirtes, Glymour and Scheines (1993).

Woodward (2003) argues that manipulationist notions of causation abound in experimental design and the sciences generally. For example, Cook and Campbell (1979) say in their highly influential book on experimental design:

*The paradigmatic assertion in causal relationships is that manipulation of a cause will result in the manipulation of an effect…. Causation implies that by varying one factor I can make another vary, (36, emphasis in the original).*

Consider, as well, the fact that the statistician Holland (1986) expressed his view in the slogan “no causation without manipulation.” From the field of economics, Hoover (1988) says that the following “definition of cause is widely acknowledged”:
A causes B if control of A renders B controllable. A causal relation, then, is one that is invariant to interventions in A in the sense that if someone or something can alter the value of A the change in B follows in a predictable fashion (173, all quotations quoted in Woodward 2003:25).

But manipulationist accounts have not fared well in the field of philosophy, until recently. Of the early manipulationists, Woodward (2009) writes:

… their strategy has been to take as primitive the notion of manipulation (or some related notion like agency or bringing about an outcome as a result of free action), to argue that this notion is not itself causal (or at least does not presuppose all of the features of causality the investigator is trying to analyze), and to then attempt to use this notion to construct a non-circular reductive definition of what it is for a relationship to be causal, (4).

So the idea was to stipulate as primitive some shared experience of human agency and reduce causation to some notion related to it, which itself would not be causal. However, early philosophical manipulationist accounts faced a variety of problems including circularity and anthropocentrism. For example, we worry about circularity when von Wright (1971) writes:

… to think of a relation between events as causal is to think of it under the aspect of (possible) action. It is therefore true, but at the same time a little misleading to say that if \( p \) is a (sufficient) cause of \( q \), then I could produce \( p \) I could bring about \( q \). For that \( p \) is the case of \( q \), I have endeavored to say here, means that I could bring about \( q \), if I could do (so that) \( p \) (74).

We may object that “producing” and “doing” are already causal notions and so should not be appealed to in a reductive account of causation. In order to dodge the circularity, von Wright may reply that he was referring to a primitive, non-causal notion of experience with human agency. But then just what causes are is unhappily married to the anthropocentric notion of our experience of agency, which is itself arguably causal anyway.

The recent interventionist program, for instance in Woodward (2003), does not depend on human agency; nor does it even attempt the reduction of causation to non-causal
notions. A theory does not have to be reductive to be illuminating with respect to “how causal concepts are interconnected,” Woodward (27) says. Additionally, Woodward adds a counterfactual aspect that the early manipulationists did not properly countenance.

The basic idea can be put this way:

A causes B if and only if B would change if an appropriate manipulation on A were to be carried out.

Consider the familiar example in which atmospheric pressure, A, is a common cause of B, the reading of a barometer, and S, a variable representing the occurrence or non-occurrence of a storm.

![Diagram](image)

In this diagram, B is represented as not being a cause of S. The interventionist account should be able to account for this in its own terms. And it does. Manipulate B directly and S does not change. Manipulate A and S indeed does change, which is the right answer.

Woodward’s account can also handle familiar potential pre-emption cases. Consider again the gunmen case. “Gunman one shoots (c1) victim, causing his death, d, while gunman

---

44 Woodward has the virtue of isolating the concepts of a total, direct and contributing cause. However, there is not the space to reconstruct the discussion.
two does not shoot but would have shot (c2) also causing d, if c1 had not occurred,”
(Woodward 2009: 25).

If we fix (via an intervention) the behavior of gunman two at its actual value (he does not shoot), then an independent intervention that alters whether gunman one shoots will alter whether victim dies, thus identifying c1 as the actual cause of e, despite the absence of counterfactual dependence (of the usual sort) between d and c1,” (25-26).

Woodward’s basic idea is that the claim “X causes Y” means that at least for some individuals (tokens) there is some manipulation of or intervention on the value of the variable X that they possess such that in the right conditions the value of the variable Y changes for the individual possessing it. Obviously, much depends on getting “conditions” of intervention right. We need to specify what other variables, if any, are held fixed when X is manipulated.  

Just what counts as an intervention is “a matter of some delicacy,” (Woodward 200x: yz). We may think of an intervention on X with respect to Y as a causal process that changes X in such a way and under conditions such that any change in Y occurs only in virtue of Y’s relationship to X.

Heuristically, we may think of the allowable changes to X (interventions, as we have been calling them) as processes that satisfy whatever conditions must be met in an ideal experiment designed to determine whether X causes Y, (Woodward 2003: 46).

Consider the barometer example again. We can randomly change the value of B by directly manipulating the dial from either high to low. “[T]he intervention ‘breaks’ the previously existing endogenous causal relationship between A and B,” (46). But this example can now be used to illustrate more about interventions per se.

45 Different possibilities give different concepts of causation, including total, direct and contributing cause.
This example illustrates the idea that interventions involve exogenous changes in the variable intervened on. When an intervention occurs on B, the value of B is determined entirely by the intervention, in a way that is … independent of the value of A.

In other words, “In this sense, the interventions need to be surgical” which means the intervention does not have untoward causal consequences, i.e., “that no other causal relationships in the system are changed,” (Woodward 2009: 16)

An intervention \( I \) on a variable \( X \) is always “defined with respect to a second variable \( Y \) (the intent being to use the notion of an intervention on \( X \) with respect to \( Y \) to characterize what it is for \( X \) to cause \( Y \),” (citation). Such an intervention must completely disrupt the causal history of \( X \), as we said in the barometer example. Also \( I \) must not itself directly cause \( Y \) unless through a route that includes \( X \); and \( I \) should not be caused by any cause that affects \( Y \) via a route that does not go through \( X \). Woodward (2003) lays out other conditions in greater detail.

Woodward is primed to offer an attempt at solving the problem of mental causation. In fact at the end of his book (2003) he makes an advertisement to that effect: “Even a very casual reader of work in philosophy of psychology, for example, will be struck by the extent to which current discussions of everything from reduction to mental content are still hostage to DN-inspired ideas. It is natural to wonder how these problems would look in other frameworks for thinking about cause and explanation,” (Woodward 2003: 375).

\[\text{The interventionist theory only works on non-closed systems such that interventions can be made exogenously, i.e., from outside the system, so to speak. This has the result that the interventionist account cannot be used to model causal relationships within the entirety of the universe. Woodward replies that at the scale of the entire universe, the concept of something causing something else falls away as his theory would predict.}\]
However, to my mind, Woodward remains one who gives an account of causal explanation – indeed of a very general, non discipline-specific causal explanation. But this is not causation as I am treating it here. An account of causation needs to keep causation as extensional and usually it will say what causal relations reduce to, whether it be laws or truth conditions of counterfactuals, etc. It will say what in virtue of which a cause is a cause. Woodward does not answer this question.

§

2.3 Adopting a View

There is a good amount of plausibility to the nomological account of causation. In any case, I adopt it for the sake of argument, since it is arguably Davidson’s position and that of his closest critics. Davidson’s view works with strict, exceptionless laws which make the effect necessary given the cause. A de facto regularity theory lacks the salient notion of necessity. In order to understand the difference between de facto regularities and the kind of laws Davidson appeals to we need a theory of causal laws and the necessity therein.

I immediately below consider just one reason to favor the nomological account to the counterfactual account of causation. This is not a knock down argument. It’s just one relevant consideration in favor of the nomological theory.

[47] I am going to show in later sections that Davidson’s view, properly understood, does not suffer from the exclusion problem and also does not suffer from the qua problem. I will disambiguate two “in virtue of” questions and show that on either question Catastrophe is not entailed. Then in Chapter Five I am going to switch gears and deny the implied directionality of what’s in virtue of what in the metaphysical “in virtue of” question. This may not be an answer to the question but it is a response to it.
2.3.1 The Nomological Approach Versus the Counterfactual Approach

A counterfactual theory of causation is a promising attempt. There may be a worry that it presupposes a nomological account. I wish to hold the nomological account for the sake of exploring the qua problem. But perhaps a counterfactual account is just as good. Immediately below, we will discover that there are good reasons to believe that the counterfactual approach itself requires reference to laws, which is one prima facie reason to not give up the nomological approach for the counterfactual approach.

The counterfactual approach to causation has intuitive plausibility. The space heater’s short-circuiting caused the fire. Why do we say that? Because if the space heater had not short-circuited, the fire would not have occurred. What is the basis of saying that c caused e? It is that had c not occurred, e would not have occurred. The same hold for situations involving intentional action. On what basis do we think that George’s thirst caused him to go to the fridge? On the basis that we believe that if George had not been thirsty, then George would not have gone to the fridge. As Kim writes, “In confidently making these ordinary causal assertions or counterfactual claims, we seem entirely unconcerned about the question whether there are laws about [being thirsty and going to the fridge.]”

However, just having counterfactuals at hand is not enough to build a theory on. That “c caused e” cannot be grounded on the fact that “if c had not occurred, e would not have occurred” because they are two different ways of saying the same thing, Kim argues (2006: 190). “Neither can ground the other; that is, neither could be offered as an explanation of how the other could be true,” (190). But perhaps an account of mental causation can be built in terms of what makes the mental-physical counterfactuals true. The semantics of counterfactuals – the conditions under which counterfactuals can be evaluated as true or false
– come in two main varieties (neither is without controversy). There is (1) the nomic-derivational approach and (2) the possible worlds approach.

On the nomic-derivational approach, the counterfactual conditional “If P were the case, Q would be the case” is true just in case the consequent, Q, of the conditional can be logically derived from its antecedent, P, taken together with the laws and the conditions holding of the situation. Consider an example: “If this match had been struck, it would have lighted.” This counterfactual is true because its consequent “The match lighted,” can be derived from its antecedent, “The match was struck,” in conjunction with the law “Whenever a dry match is struck in the presence of oxygen it lights,” plus the auxiliary premises “The match was dry” and “There was oxygen present,” (See Kim 2006: 191-192). The relative plausibility of this approach does not constitute a reason to give up the nomological approach because it too makes essential reference to laws. So let’s consider the possible worlds approach.

According to the possible worlds approach to the semantics of counterfactuals, the counterfactual “If P were the case, Q would be the case” is true just in case Q is true in the world in which P is true and that, apart from P’s being true there, is as much like the actual world as possible. (To put it another way: Q is true in the closest P-world.) In other words, we go through the following steps in order to see whether or not a counterfactual is true. Since we are dealing with something counter-to-fact, the antecedent, P, is false in the actual world. Then, as Kim relates it:

We must go to a possible world in which P is true and see whether Q is also true there. But there are many worlds in which P is true – that is, there are many P-worlds – and in some of these Q is true and in other false. So which P-world should we pick in which to check on Q? The answer: Pick the P-world that is the most similar, or the closest, to the actual world.
So in other words, the counterfactual “If P were true, Q would be true,” is true if Q is true in the closest P-world; it’s false otherwise. But which is the world most similar to the actual world?

Again, consider the example of striking a match. The counterfactual is “If this match had been struck, it would have lighted.” So in the actual world the match was not struck. We need to imagine or suppose a world in which the match was struck while keeping other conditions the same as the actual world as much as possible. As Kim reminds us, “Certain other conditions must also be altered under the counterfactual supposition that the match was struck: For example, in the actual world the match lay motionless in the matchbox and there was no disturbance of the air in its vicinity, so these conditions would have to be changed to keep the world consistent as a whole,” (192). But in the world we have picked we should not suppose that the match is wet or that there is a deficit of oxygen. So in the world we are thinking of, the match is moved and air disturbed and the match is struck while the match is dry and there is sufficient oxygen. The question then is, does the match light? In asking this question we are asking which of the following two worlds is closer to the actual world:

\[ W_1: \text{The match was struck; it was dry; oxygen was present; the match lighted.} \]
\[ W_2: \text{The match was struck; it was dry; oxygen was present; the match did not light.} \]

It seems that \( W_1 \) is closer to the actual world thereby making the counterfactual true. But why do we judge \( W_1 \) as more similar to the actual world? Kim’s answer: “Because we believe that in the actual world there is a lawful regularity to the effect that when a dry match is struck in the presence of oxygen it ignites, and \( W_1 \), but not \( W_2 \), respects this regularity,” (192). So in
judging that W1 is closer to the actual world than W2 is we are making use of a lawful regularity.

So once again we lack a good reason to leave behind the nomological account of causation. One thing we have learned, however, is that the possible worlds approach to the semantics of counterfactuals allows us to account for causation with reference to non-strict laws. They are non-strict laws that we appeal to when making similarity judgments between worlds. But this is not a decisive reason to leave behind our nomological account of causation which refers to strict laws. “For,” as Kim says, “it may well be that these nonstrict laws are possible only if strict laws are possible and that where there are no underlying strict laws that can explain them or otherwise ground them, they remain only rough, fortuitous correlations. It may well be that their lawlike appearance is illusory and that this makes them incapable of grounding causal relations,” (194). We cannot know that nonstrict laws are any better than correlations or cases of common cause unless the nonstrict laws are backed by strict laws which rule out such eventualities.

§

2.4 On Causal Laws

The questions “What is causation?” and “What is a law of nature?” are separate questions even if the answer to the first question is that causation reduces to something about laws of nature. Having adopted the nomological account of causation according to which causation is related to laws do I need to come to a position on the nature of laws? Not necessarily. However, as I said above, Davidson’s view works with strict, exceptionless laws which make the effect necessary given the cause. A de facto regularity theory lacks the salient notion of
necessity. In order to understand the difference between de facto regularities and the kind of laws Davidson appeals to we need a theory of causal laws and the necessity therein.

So, I discuss causal laws here in advance of what I say in Chapter Five about the bruteness of causal relations with respect to the causal laws supposedly backing them.

2.4.1 Introduction

We are familiar with the widespread and relatively plausible idea that causal relations are backed by regularities of some kind. Token causal interactions are backed by regularities. Philosophers take these regularities to be either de facto regularities or lawful regularities that necessitate e given c.

Here is Beebee (2004) summarizing a key trend in the literature on laws:

There are two main camps in the debate about the metaphysics of laws of nature.\(^{48}\) In one corner, there is the anti-Humean view of David Armstrong: laws are relations of necessity between universals. And in the other corner, there is the Ramsey-Lewis view: laws are generalizations which figure in the most economical true axiomization of all the particular matters of fact that obtain… The debate between the rival camps can be read as a debate about whether or not supervenience holds for laws of nature: whether or not nomic facts supervene on non nomic facts, [i.e.,] to put it in Lewis-esque terms, whether or not laws supervene on the overall distribution of particular matters of fact (250).

Whether or not supervenience holds is a matter of what depends on what. The Humeans affirm supervenience of laws on non-nomic “local matters of particular fact” and the non-Humeans deny the same while affirming the supervenience of causal facts on laws. In a slogan, Humeans say two worlds cannot differ with respect to laws without differing with respect to local matters of particular fact. And non-Humeans say two worlds cannot differ with respect to local matters of particular fact without differing with respect to laws, in the

\(^{48}\) There are actually more camps then she countenances, but contrasting these two is important.
sense that the supervening items are dependent on the subvening base. Let’s turn to the Humean position first.

2.4.2 Humeanism and Regularities

There is Humeanism about causation and there is Humeanism about laws. They are related but it’s important to keep them distinct. According to Psillos’s (2002) treatment, the “regularity view of causation,” or RVC, is a good general Humean view on causation.

RVC:
(a) causation is a species of regularity
(b) the species of regularity that causation reduces to are laws of nature.

Now, Humeanism about laws, i.e., the regularity view of laws, RVL, goes as follows:

RVL:
(c) laws of nature are regularities.

It is important to note that RVL does not entail RVC. You can hold (c) and deny (a) as well as (b). However, RVC does trivially entail RVL. If you say causes depend on laws, you had better have a view on laws.

Let’s begin to present the general characteristics of a Humean view on laws by starting with a naïve regularity view of laws. According to such a regularity view of laws, laws are regularities or generalizations connecting types of events. How does the naïve RVL address the canonical issue of differentiating between generalizations that are merely accidental and those that are actually laws? There is a clear difference, after all, between the regularity that all apples in the bowl are red and the regularity that all metals expand when heated. Both are regularities but the latter strikes us as a candidate for lawhood while the former does not.
The Humean tradition says there are no necessary connections between events; it “bans objective necessity from nature,” (Psillos 2002, p 139). So, not being able to say the laws are the necessarily true generalizations (and the accidental generalizations accidental), the Humean tradition has an especially hard time with this issue of differentiating accidents (as I’ll call them for short) and laws.

According to the Humean view, when it’s said it’s a law that metals expand when heated, what’s meant is that there is in nature a regularity according to which when metal gets heated, it expands. There is no necessity in the regularity because it is logically possible that a metal is heated and yet does not expand. But accidents are regularities as well and we still need something differentiating them; laws and accidents need differing treatments.

How can the naïve RVL explain the differences? Well, only by no longer being a naïve view. The more sophisticated view will say that laws are regularities plus something else. I’ll call it “an extra bit.” And the sophistication will come from saying what this something else or “extra bit” is.

To formalize, let’s say the naïve RVL holds:

It is a law that all Fs are Gs if and only if all Fs are Gs.

Some preliminary sophistication that can be added to the naïve view includes holding that a statement, L, is a statement of a law of nature if and only if:

- L is universally quantified
- L is true everywhere and across time
- L contains only natural-kind predicates, besides connectives and quantifiers.

This is a good start, but it is still unable to deal with Reichenbach’s now classic accidental generalization: “All gold cubes are smaller than one cubic mile.” This statement has all the
required features just listed, but it is still arguably different than the following law, which is not merely accidental: “All uranium cubes are smaller than one cubic mile.” It’s only a contingent lack of interest and resources that prevents us from constructing a cube of gold a mile wide, high and deep. But we could not possibly create a cube of uranium those dimensions even if we had the resources and interest. That much uranium would be over the element’s critical mass and it would explode long before arriving at those dimensions.

The upshot here is that an accident lacks the “extra bit” a law has. A law doesn’t merely state that things are such and such a way. A law doesn’t merely say that all Fs are Gs; rather, a law also says that were a non-F thing to become an F thing then it would also become G, or, in any case, and less contentiously, a law makes a claim about counterfactuals.

Among the sophisticated Humean views, David Lewis’s Best System Account is, among other things, a powerful attempt to account for the laws’ counterfactual support that accidents lack.\(^{49}\)

### 2.4.3 Lewis’ Best Systems Account

Lewis’ view has an ancestry in Mill’s approach. In thinking about how to “ascertain the laws of nature,” Mill said:

> According to one mode of expression, the question, What are the laws of nature? may be stated thus: What are the fewest and simplest assumptions, which being granted, the whole existing order of nature would result? Another mode of stating it would be thus: What are the fewest general propositions from which all the uniformities which exist in the universe might be deductively inferred? (1911: 207)

\(^{49}\) The idea of laws as the “best” axioms of a deductive system goes back to Mill (1843) and was first defended in contemporary times by Ramsey (1928). It was made popular by Lewis (1973, 1983, 1994).
Lewis’ version or “mode of expression,” goes something like this. Laws are those regularities (the regularities that aren’t laws are accidents) that are “members of coherent system of regularities, in particular, a system that can be represented as a deductive axiomatic system”\textsuperscript{50} striking the best balance between simplicity and strength, where the deductive system “results in” or “deductively explains” the total universe of local, particular matters of fact. This implies that “no regularity taken in isolation can be characterized as a law. Lawlikeness cannot be ascribed to a regularity in isolation from other regularities,” (Psillos 2002: 149) that play a part in the explanation. Further, simplicity and strength are virtues of explanatory, theoretical systems and they pull in different directions. A system should be as informative as possible (have the highest degree of strength) but simultaneously be as simple as that strength will allow. As Lewis elegantly puts it:

\begin{quote}
The virtues of simplicity and strength tend to conflict. Simplicity without strength can be had from pure logic, strength without simplicity from (the deductive closure of) an almanac… What we value in a deductive system is a properly balanced combination of simplicity and strength – as much as truth and our way of balancing will permit (1973: 73).
\end{quote}

Lewis’ account holds, more formally:

\begin{quote}
It is a law that all Fs are Gs if and only if (i) all Fs are Gs, and (ii) that all Fs are Gs is an axiom or theorem in the best (balanced for simplicity and strength) deductive system.\textsuperscript{51}
\end{quote}

The best system view is in keeping with what’s called “Humean supervenience,”\textsuperscript{52} which is “the doctrine that all there is to the world is a vast mosaic of local matters of

\textsuperscript{50} This is not actually Lewis’ own mode of expression but that of Psillos (2002).

\textsuperscript{51} Or, if there is no unique best deductive system, then that it is an axiom in all deductive systems that tie in terms of simplicity and strength.
particular fact, just one little thing and then another (Lewis 1986, p. ix). As the Beebee quotation above indicated the Humeanism of the best systems view has it that the laws supervene on “local matters of particular fact.” In a slogan, the regularities depend on the local matters of particular fact and their lawfulness is dependent on their having a role in the best deductive system explaining all those local matters of particular fact.

Lewis’ account makes a good case for how to differentiate between accidents and laws. Those true regularities that are nevertheless accidents get eliminated by the strength and simplicity requirement. “All apples in this bowl are red,” and “All gold cubes are less than 1 cubic mile,” might be true regularities, never falsified in the life of the universe, but they won’t play a role in the simplest deductive system that would explain all local, particular matters of fact.

Lewis’ account answers in a unique way the question of what makes a law the law that it is. The view says a law is a regularity that plays a role in the network of laws that makes up the deductive system. But the simplicity and strength requirement has moved some critics to accuse Lewis’ best system account of mind-dependence.

Psillos says there are two questions here: “The first is whether the web of laws approach makes laws mind-dependent. The second is whether there is a worldly feature that makes some regularities laws.” Loewer (1996) answers that the lawfulness of the regularities might be mind-dependent but the regularities themselves (which are the same as the laws missing the “extra bit”) are not mind-dependent. The regularities we write down when we collect patterns do form a system. That is, the objects and properties in the world have an objective nomological structure. The patterns they make can be said to depend on the way

---

52 There are many different formulations of the general idea of Humean supervenience. The literature on “Humean Supervenience” is quite large and diverse. But see Earman (1984) and Loewer (2004).
the world is. The regularities are the regularities that they are in virtue of the (brute) way the world is. And: “These relations can be captured by relations of deductive entailment” (p. 154) in an ideal deductive system of our (future) knowledge of the world, balanced in the right way with the right theoretical virtues of simplicity and strength. So in the last analysis, a regularity is regularity in virtue of the world but a regularity is nomological, i.e., is a law, in virtue of being a part of this network of regularities (themselves thereby laws) in the best system.

2.4.4 The Necessitarians

However, the non-Humean critic replies that to treat laws this way leaves out a lot of what we usually understand by “laws of nature.” As Carroll (1994) says, just intuitively speaking we usually conceive of laws as guiding or “governing” the relations between events that populate the universe. Laws are seen as responsible for the patterns or regularities not merely identical to them. To take laws as the more basic, “more primitive” entity which determines the local particular matters of fact among events is the non-Humean strategy of Armstrong (1983), Dretske (1977) and Tooley (1984). On accounts like this, there is a relation of universal necessity between types or universals.

Suppose it to be a law that Fs are Gs. F-ness and G-ness are taken to be universals. A certain relation, a relation of non-logical or contingent necessitation, holds between F-ness and G-ness. This state of affairs may be symbolized as ‘N(F,G)’ (1983: 85).

Notice N is not just a generalization holding between F and G, it is a relation holding between two things other than F and G, namely F-ness and G-ness. As Carroll (2010) says, “… a law is not just a universal generalization, but is an entirely different creature — a

53 As the Beebee citation above says, the non-Humeans and the Humeans can be usefully described as disagreeing on whether or not nomic facts supervene on non-nomic facts.
relation holding between two other universals.” Armstrong’s framework is also consistent with lawhood not supervening on local matters of particular fact. This is a denial of Humean supervenience, which often goes with the universals approach.

2.4.5 Evaluating the Humean and Non-Humean Theories

Let’s compare these two. Against BSA it is argued that BSA injects a sort of mind-dependence that is undesirable. We will return to that. Against the non-Humean necessitarians, there is leveled the “identification problem” and the “inference problem.” Basically, just what the lawmaking relation is needs to be specified (the identification problem). Then, it still remains to be seen if it is suited to the task (the inference problem).

Does N’s holding between F and G entail that Fs are Gs?

In this regard, Lewis writes the following against the necessitarians:

“Whatever N may be, I cannot see how it could be absolutely impossible to have N(F,G) and Fa without Ga. (Unless N just is constant conjunction, or constant conjunction plus something else, in which case Armstrong’s theory turns into a form of the regularity theory he rejects.) The mystery is somewhat hidden by Armstrong’s terminology. He uses ‘necessitates’ as a name for the lawmaking universal N; and who would be surprised to hear that if F ‘necessitates’ G and a has F, then a must have G? But I say that N deserves the name of ‘necessitation’ only if, somehow, it really can enter into the requisite necessary connections. It can't enter into them just by bearing a name, any more than one can have mighty biceps just by being called ‘Armstrong’” (1983: 366).

Here Lewis makes what I take to be a classically Humean move. He denies that we can see or experience necessity. He also nicely exposes the fact that Armstrong can name his relation “necessitation” but that alone will not make it a necessary relation.

Another dimension on which theories differ is the question of what determines what. The necessitarians deny that laws supervene on local matters of fact. Instead, of course, the
necessitarians want to say that laws determine causes and are metaphysically responsible for making causes what they are.

The Humean takes it the other way around and with a twist. Laws are determined by causes (taken as regularities in matters of fact) in this sense that the BSA details. However, aside from mind dependence Humeans have a challenge in the cases brought against them that turn on the intuition about laws determining local matters of fact.

There are some important cases that appear to show that local matters of fact do not determine the laws. Tooley’s (1977) “not yet (or never) interacting particles” example:

The interaction of X and Y particles have not been studied because conditions are such that they never will interact. Nevertheless, it seems that it might be a law that, when X particles and Y particles interact, P occurs. Similarly it might be a law that when X and Y particles interact, Q occurs. There seems to be nothing about the local matters of particular fact in this world that fixes which of these generalizations is a law.

Carroll’s case features two possible worlds, U1 and U2 which he says are the same with respect to local matters of fact:

\[ L1, \text{ the generalization that all } X \text{ particles subject to a } Y \text{ field have spin up, could be a law of } U1. \]

What is new about U2 is that when b enters the Y field at time t it does not acquire spin up. … Of course, there must be at least one more difference between these two worlds. Though L1 could be a law in U1, L1 could not be a law of U2; L1 is not true in U2. There is nothing particularly remarkable about either U1 or U2 — nothing to make a Humean suspicious. But here is the catch: It is natural to think that L1’s status as a law in U1 does not depend on the fact that b entered that Y field at time t.

Both of these thought experiment are aimed at the Humean. In Tooley’s it appears intuitively that the local particular matters of fact do not determine what the laws are. These two
particles have never interacted, so there is no fact about how they interact, yet it remains intuitively plausible that there is a law governing what they would do were they to interact.

I do not intend to make any knockdown arguments here against any view in the law of nature literature. But I do have one thought about Tooley’s case for the Humean. Tooley fails to emphasize that we are talking about laws and local particular matters of fact over all of time – the whole time scale. To do so insures that we are not talking metaphysically about how we discover laws nor about how we explain events by appeal to laws. Rather we are talking about laws and patterns in their most complete sense. With this in mind then the stipulation that two particles have never collided means they *never will collide*. And if they never will collide it seems intuitive, at least to me, that there would be no law describing their colliding-behavior.

§

2.5 Conclusion

In this chapter, we have discussed causation per se, causal explanation, scientific explanation and the literature on the laws of nature. We return throughout the dissertation to the notions laid out here.
CHAPTER THREE

THE METAPHYSICS OF THE EXCLUSION PROBLEM

3.1 Introduction

In the following section, I lay out the exclusion problem for an anonymous or general non-reductive physicalist. I am not below describing anomalous monism; that will happen in Section 3.9.5 below. Furthermore, I do not endorse the premises of the exclusion argument except for the sake of getting Kim’s position out on the table.

I give a brief formulation here, then consider the history of the mind-body problem. And then return to consider only a few pros and cons for each premise in Kim’s argument.

3.2 Brief Formulation of the Exclusion Problem

In the philosophy of mind literature, the exclusion problem arises from relatively plausible theses that nevertheless seem to conflict with one another.

*Mental Causation:* Mental events cause physical events.

*Physical Causal Closure:* Each physical event, insofar as it has a cause, has a sufficient physical cause.

*Non-Reduction:* Mental events neither reduce to physical events nor are identical with them.

*Non-Overdetermination:* Physical events are not pervasively overdetermined, i.e., they are not subject to multiple sufficient causes.
We are concerned with a problem that arises for those who – call them non-reductive physicalists – would affirm these four theses.\textsuperscript{54} We are not concerned at the moment with why they would affirm them, although they are all prima facie plausible and in any case widely held. Briefly, \textit{Mental Causation} is meant to be the phenomenon needing to be explained. If it is denied, then \textit{Catastrophe} follows. \textit{Non-Reduction} is desirable, as well as plausible, according to many theorists. It is desirable as one way to protect the autonomy of explanations that refer to mental properties, or mental events, under mental descriptions; and it is plausible according to philosophical arguments such as the multiple realizability argument. \textit{Physical Causal Closure} is a widely held and plausible commitment for physicalists. It says, roughly, for any event (of any type) that has a cause, that cause is a sufficient physical cause.\textsuperscript{55}

Why is there the \textit{Non-Overdetermination} thesis? \textit{Physical Causal Closure} did not rule out \textit{multiple} sufficient causes. However, many philosophers have been quick to note that multiple sufficient causes is implausible. To this end we assert the \textit{Non-Overdetermination} thesis. It says that events are not subject to widespread overdetermination, meaning very rarely is any event caused by two sufficient causes. These each will be discussed in greater detail below in Section 3.9.4 below.

\§

3.3 A Brief History of the Mind Body Problem

3.3.1 Descartes, Elizabeth and Leibniz

\textsuperscript{54} In this formulation I leave any distinction between token and type ambiguous because Kim does. Part of the reason anomalous monism dodges the exclusion problem, discussed below, involves properly countenancing the token-type distinction so it would not do to introduce it too early.

\textsuperscript{55} There are a variety of more or less sophisticated formulations of the causal closure of the physical, but this one is sufficient for our purposes. See Papineau (2002).
Princess Elizabeth forced Descartes to attempt an account of psychophysical interaction. The two corresponded while Descartes was writing his *Principles*, which is dedicated to her. He wrote the *Passions* in large part as a response to her probing questions. Most significantly Elizabeth asked Descartes: “Given that the soul of a human being is only a thinking substance, how can it affect the bodily spirits, in order to bring about voluntary actions?” The question was pertinent because Descartes had described the mind as exclusive of extension, yet held that how a thing moves depends on “how much it is pushed, the manner in which it is pushed or the surface-texture or shape of the thing that pushes it,” which require contact or touch – impossible for an immaterial substance. Descartes made a response that left many unconvinced and the mind-body problem had started.

### 3.3.2 The 1950s and Behaviorism

Cartesianism, in one form or another, ruled the day until the 1950s when behaviorism put forth a challenge that ended in a dialectical stalemate. As Lycan (2008) writes:

---

56 See Smith (2010).

57 Any attempt to somehow have the mind move matter in the realm of bodies would break the law of conservation of quantity of motion. Descartes was not without resources to respond, however. Leibniz made a friendly reconstruction of Descartes’ argument such that indeed the physical system was closed with respect to conservation of quantity of motion. The affect that the mind had was on the directional vector of the unit of quantity of motion. Thus, Leibniz argued, the mind’s causal power was constituted by this special kind of influence on the physical world: “Descartes knew that minds could not at all give force onto bodies because there is always the same quantity of force in matter. Therefore, he held that the mind could change the direction of bodies. But this is because in his time they did not know of the law of nature that conserves the total direction of matter [momentum].” Translation by D.F.; the original passage is: “Descartes a reconnu que les âmes ne peuvent point donner de la force aux corps parce qu’il y a toujours la même quantité de force dans la matière. Cependant il a cru que l’âme pouvait changer la direction des corps. Mais c’est parce qu’on n’a point su de son temps la loi de la nature qui porte encore la conservation de la même direction totale dans la matière.”

58 Whom I follow closely here.
Until the 1950s… the philosophy of mind was dominated by Descartes’s “first-person” perspective, our view of ourselves from the inside. With few exceptions, philosophers had accepted the following claims: (1) that one’s own mind is better known than one’s body, (2) that the mind is metaphysically in the body’s driver’s seat, and (3) that there is at least a theoretical problem of how we human intelligences can know that “external,” everyday objects exist at all, even if there are tenable solutions to that problem.

Lycan says that for a number of reasons substance dualism of the Cartesian variety lost some favor:

The first reason was the accumulated impact of logical positivism and the verification theory of meaning. Intersubjective verifiability or testability became the criterion both of scientific probity and of linguistic meaning itself. If the mind, in particular, was to be respected either scientifically or even as meaningfully describable in the first place, mental ascriptions would have to be pegged to publicly, physically testable verification conditions. Science takes an intersubjective, third-person perspective on everything; the traditional first-person perspective had to be abandoned for scientific purposes and, it was felt, for serious metaphysical purposes also.

Behaviorism (see, for example Ryle [1949]) was the first main competitor against dualism. According to behaviorism, mental state ascriptions simply mean something about dispositions to behave in a certain way in response to environmental stimuli. So, for example, “He is in pain,” just means that he is disposed to grimace-behavior, and the like. In this way, no reference is made to some insubstantial, non-observable, immaterial, unscientific Cartesian soul.

This kind of “analytical behaviorism” guarded its territory aggressively against “soft-headed” types who would countenance an immaterial mind. However, some criticisms stuck. For one, it was argued that we pre-theoretically can introspect actual inner mental states which are not tied to any behavior or even disposition to behave. Place (1956) speaks of an
“intractable residue” of conscious experience that is unrelated to behavior in any way.\textsuperscript{59}

Secondly, the inverted qualia argument asserts that it is entirely possible for two persons to differ mentally, psychologically, or experientially while being the same in terms of behavior and dispositions to hypothetical behavior. It might be, after all, that when I see a red object I have the sort of color experience that you have when you see a green object, but we both call the object by the same name and interact with it in entirely the same ways. Finally, behavioral analyses of mental state ascriptions were shown to be circular or incomplete for tacitly relying on mental terms within the ostensible behavioral analysis. For example, as Lycan writes, “if Leo believes that parsnips are dangerous and he is offered parsnips, he would shun them but only if he does not want to die.” The “wanting” there is unanalyzed.

Lycan writes that there was a dialectical stalemate between dualism and behaviorism until the identity theory of Place (1956) and Smart (1959) offered an irenic solution and moved the discussion forward.

3.3.3 Place and Smart

According to Place, the “intractable residue” of consciousness is to be identified with neurophysiological states of their bearer’s central nervous system. However, Place remained a behaviorist about what we would call intentional states. “In the cognitive concepts like ‘knowing,’ ‘believing,’ ‘understanding,’ ‘remembering,’ and volitional concepts like ‘wanting,’ and ‘intending,’ there can be little doubt, I think, that an analysis in terms of dispositions to behave (Wittgenstein 1953, Ryle 1949) is fundamentally sound,” (44). But the phenomenal quality of experience (perhaps including experience of intentional states) included an “intractable residue” that Place held could not be analyzed behavioristically. While

\textsuperscript{59}Although, as we will see, Place is a behaviorist about intentional states.
behaviorism had some limited application, “On the other hand, there would seem to be an intractable residue of concepts clustering around the notions of consciousness, experience, sensation, and mental imagery, where some sort of inner process is unavoidable,” (44). These non-behaviouristic inner processes were to be identified with neurophysiological states, for instance c-fiber firings. Later Armstrong (1968) would identify intentional states as well with neurophysiological states.60

On the identity account, then, to be in pain is to have your c-fibers firing and “to believe that broccoli will kill you is to have your Bₔfibers firing,” writes Lycan. It is an irenic account because it countenances genuine inner, episodic mental states like the dualists, but unlike the dualists these states are identical to some neurophysiological state. And like the behaviorist the mental states are scientifically respectable for being so identified. But unlike the behaviorist, mental states are not at all associated with behaviors, nor with dispositions to behave.

As Lycan writes: “By making the mental entirely physical, this identity theory of the mind shared the behaviorist advantage of avoiding the objections to dualism. But it also brilliantly accommodated the inner and the episodic as behaviorism did not.”61

3.3.4 Functionalism and Multiple Realizability

Quite soon it was realized that the identity theory was inadequate as a type-identity theory. If being in pain means to have c-fibers firing, then organisms with other biology and anatomy could not properly be said to be in pain. Putnam (1960, 1967a, 1967b) and Fodor (1968b)

60 See also David Lewis (1966, 1972).

61 The identity theory dodged the problems associated with both dualism and behaviorism. Plus Lewis and Armstrong independently offered, as well, a deductive argument for the identity theory. (See Lycan 2008 for details.)
argue for the appropriate fix. They argued that pain and other mental states were multiply realizable, that is, they could be realized by a wide variety of biological or physiological realizers. What mattered was not the realizer but the causal role or function being realized—the role of pain or other mental states.

Functionalism argues that pain “tokens” are identical to physiological “tokens,” but the property of pain, or any other mental state types, are identical only to the role that pain or these other states play in the mental and behavioral economy of the agent. So functionalism asserted token identity but denied type identity on the grounds that mental types are multiply realizable.62

Around the same time as functionalism, or actually 5-10 years later, another token identity view that denies type identities arose—namely, Davidson’s anomalous monism. The particulars of anomalous monism (as well as its troubles with the qua problem and the identification problem) are discussed below in Chapters Three and Four so we may skip over the details for now. Suffice it to say that anomalous monism signaled a shift in the state of the dialectic in the mind body problem. In time, non-reductive physicalist views (type-nonidentity views that nevertheless claimed physicalist bona fides) became the received wisdom.

§

3.4 Formulation of the Exclusion Problem

As promised we will now layout a more detailed formulation of Kim’s exclusion problem.

62 Lycan (2008) gives further details on the various kinds of functionalism. Lewis (1972 and 1994) and Jackson, Partridge and Prior (1982) argue from functionalism to type identity, for one instance. Also see Block (1980).
The problem is the four are inconsistent together, yet individually plausible. Anyone who holds these four theses – namely non-reductive physicalists – will have the exclusion problem. Let’s now take a less cursory look at each premise individually. I am assuming these propositions for the sake of argument. I’ll give a few considerations for or against each one. However, I am *assuming* them in the end in order to represent Kim’s formulation.

### 3.4.1 First Premise

The first premise is

(1. MC) Some mental events cause physical events.

There are considerable pressures to avoid dropping this premise in any attempt at consistency. The causal efficacy of the mental is after all the phenomenon we have tasked ourselves with accounting for. Let’s review a few different but interrelated reasons to hold this premise.

Some will say that the causal efficacy of mentality, mental events and/or mental properties, is non-negotiable. We can see this by observing the extent to which it is fatal to a theory if it entails mental epiphenomenalism.

---

63 There is evidently counterfactual dependence often enough between agents’ reasons and their actions. But the multiple reasons argument of Chapter One requires that the reason be a cause of the action, not merely correlated or related by counterfactual dependency.
But we should also ask a further question. Namely, why is it so important that causal efficacy be explained in our account and not eliminated? Why is it non-negotiable? I believe the answer lies in the practical significance of our reason explanation practices, which themselves refer to mental events as causes.

Indeed, the project in the mental causation debate is to vindicate the legitimate and autonomous explanatory status of reason explanations or explanations of intentional actions. This, admittedly, is only one part of scientific psychology and even only one part of folk psychology. But it is the recalcitrant part, the part which has resisted a naturalistic accounting.

Kim captures the non-negotiable status of this premise, that some mental events cause physical events, in his distinction between how and whether questions. He says the question of whether or not there is mental causation must be answered in the affirmative. Our task is to answer the question about how that actuality is possible.

One final way in which this premise has been argued for has been to invoke G.E. Moore (for example in Bontly 2005). Mental causation, the fact of the causal efficacy of our mental events, is a “Moorean fact,” meaning it has more in its favor than anything you could say from philosophy suggesting its non-existence. As when Moore held up his hands and said “Here is a hand and here is another,” in his argument against external world skepticism, so too can participants reply to any skeptic about mental causation, “Here is some mental causation: I am making my finger move with my mind,” or some such. Mental causation is everywhere and is basic.

I am, as I said, assuming the premise that mental events are causally efficacious in bringing about physical events. The debate treats it as non-negotiable. The proposition: “Some mental events cause physical events,” cannot be dropped. However, as I should hasten to add now, this does not mean that our interpretation of MC might not change.
slightly. In what follows, it will be suggested, among other things, that MC be read as saying that mental event particulars but not the mental properties of those particulars are causally related to physical event particulars. This seems natural in a philosophical discussion. Of course, MC would change its meaning from one theory to another if those theories differed with respect to their theories of the causal relata.

Furthermore, we might also add similar but slightly different premises and propositions. In the reconstruction of the qua problem it is often put forth that the premise we cannot intuitively do without is this:

(MC*) Mental events qua mental cause physical events.

We will return to MC* in Chapter Four.

3.4.2 Second Premise

The second premise is:

(CP) Each physical event, insofar as it has a cause, has a sufficient physical cause.

This proposition – called “the causal closure of the physical” – entails that in principle an explanation of a physical event could be accomplished without reference to any non-physical cause. It has been argued that causal closure is a key tenet of any physicalist theory. One might argue that if any biological or psychological events are to have physical effects then they themselves must be physical events or constituted by physical events, causal closure implies.64

64 The inference however does not logically follow. “If physical then causal” is fine. But there may be other ways of being causal. However, it is not my intention to challenge this premise of
For a long while, causal closure functioned in important arguments in the philosophy of mind without being explicitly stated. According to Papineau (2002) we need only:

… consider J.J.C. Smart's (1958) thought that we should identify mental states with brain states, for otherwise those mental states would be "nomological danglers" which play no role in the explanation of behaviour. Or take David Lewis's (1966) and David Armstrong's (1968) argument that, since mental states are picked out by their causal roles, and since we know that physical states play these roles, mental states must be identical with those physical states.

Smart’s, Lewis’s and Armstrong’s arguments assume something like causal closure. If something is not physical then it’s a dangler without causal efficacy. This is because the only way to have effects is to be physical.65

Causal Closure can be resisted, but it is not my intention to attempt a direct refutation. For instance, there might be multiple sufficient causes for all Causal Closure says. Kim’s non-overdetermination is discussed below.

3.4.3 The Third Premise

The third premise is:

\[(NI) \quad \text{Mental events do not reduce to or are not identical with physical events.}\]

As discussed above, there are at least two main reasons philosophers have adopted nonidentity with respect to mentality and physicality. However, those two reasons, multiple realization and anomalism of the mental, lead to denying only type-identity. Token identity

---

65 Again, this does not actually follow from the way Causal Closure is formulated.
can remain. In the due course of our discussion below, (NR) will have to be disambiguated as to the token-type distinction with respect to events.

3.4.4 The Fourth Premise

The fourth premise is generated after contemplation of the premise (CP), Causal Closure.

The fourth premise is, of course, non-overdetermination:

(NO) Physical events are not pervasively subject to multiple sufficient causes, or overdetermination.

Causal closure, as it is written, does not rule out multiple sufficient causes. It does not exclude the effect also having a sufficient mental cause. If it also had a mental cause then we would have an event that was “overdetermined.” This would be analogous to some oft-cited examples. As Kim says, “A man is shot dead by two assassins whose bullets hit him at the same time; or a building catches fire because of a short circuit in the faulty wiring and a bolt of lightning that hits the building at the same instant,” (Kim 1993: 252). These are perhaps legitimate examples of overdetermination of an effect; but it is dissatisfying to imagine that a situation so statistically unlikely happens in every last case of mental causation. Every time I willed my arm to move and it moved, there would be a grand cosmic coincidence such that a neurophysiological state caused my arm to move, just at the same time as my willing caused my arm to move. Kim finds this unacceptable, and, so to disallow the possibility of overdetermination, he institutes his non-overdetermination principle.
There are a number of moves available to a non-reductive physicalist that could embarrass the proponents of the non-overdetermination thesis. That is, overdetermination may not be a crazy view. But I am not pursuing the matter here.

### 3.4.5 The Logic of the Exclusion Problem

The four plausible theses are mutually inconsistent. And that is the problem. We could have a consistent account if we relinquished any one of these, but no such relinquishing is without cost. For the sake of thoroughness, here is the logic that supports the inconsistency claim.

\[(1.\text{MC}) & (2.\text{CP}) & (3.\text{NR}) \Rightarrow \neg(4.\text{NO})\]

That is, affirming (1), (2) and (3) seems to entail denying (4), which amounts to an implausible commitment to widespread overdetermination of all mentally caused events.

\[(1.\text{MC}) & (2.\text{CP}) & (4.\text{NO}) \Rightarrow \neg(3.\text{NR})\]

Affirming (1), (2) and (4) seems to entail denying (3), which amounts to mental type reduction or identity. However, as we’ve discussed, there are good reasons to deny type reduction or identity.

\[(1.\text{MC}) & (3.\text{NR}) & (4.\text{NO}) \Rightarrow \neg(2.\text{CP})\]

Affirming (1), (3) and (4) seems to entail denying (2), which denies a key tenet of physicalism.

\[(3.\text{NR}) & (2.\text{CP}) & (4.\text{NO}) \Rightarrow \neg(1.\text{MC})\]

---

66 See Sider (2003), Bernstein (manuscript), Bennett (2003), and Carey (2011) for discussions of the viability of overdeterminationism.
Affirming (2), (3) and (4) seems to entail denying (1), which denies the phenomenon to be captured.

Kim’s version of the exclusion problem concludes that no mental event ever causes any physical event. But in other places Kim has put forth the same argument plus a supervenience thesis. With the addition of a supervenience claim Kim concludes that no mental events cause any mental events.

§

3.5 What is Supervenience?

Supervenience is a philosophical notion that is meant to capture a certain metaphysical relation between properties or kinds of properties.\(^67\) It is most often supposed to be a determination or dependence relation, such that a variation in the supervening properties cannot occur (i.e., is impossible) without a variation in the subvenient or base properties. That is to say, the sets of properties must exhibit a pattern-like covariation. It was first proposed, with respect to the philosophy of mind, by Davidson (1970).\(^68\) As Horgan (1993) writes in surveying these early days, philosophers took a great interest in the notion and developed all sorts of varieties of supervenience.\(^69\)

\(^67\) For example, pain is a token of the type of property called a mental property. Pain may be said to supervene on its subvenient base, or mental properties in general may be said to supervene on physical properties in general.

\(^68\) See Hare (1952) and Moore (1922).

\(^69\) “Davidson's invocation of supervenience in connection with the mind/body problem resonated strongly among philosophers working in philosophy of mind and metaphysics; there commenced a rapid and fairly widespread appropriation of supervenience into these branches of philosophy,” (Horgan 1993: 564).
As a preliminary definition we might say: A-properties supervene on B-properties “if and only if two objects cannot differ with respect to their A-properties without also differing in their B-properties,” (Horgan 2002: 150). The B-properties fix the A-properties. In a slogan, no change in the A-properties without change in the B-properties.

So, for example, “being healthy, a property instantiable by humans, plausibly is supervenient on physical features [of humans] such as percentage of body fat, level of cholesterol in the bloodstream, absence of cancerous tissue, and the like: i.e., if one human being is healthy and another is not they must also differ in some of these physical features,” (150). Your being unhealthy depends on your possessing one of the physical characteristics underlying ill health. Looked at the other way, if you possess any of the physical characteristics underlying ill health, then it is determined that you have the property of being unhealthy.

The impossibility implied by the “cannot” in the above rough definition of the supervenience relation (“two objects cannot differ in mental properties without differing in physical properties”) invites a definition in modal terms of necessity and impossibility. Modality is often expressed in possible worlds talk such that we might say: A (short for A-properties) supervenes on B if and only if in any possible world where A changes, there is a change in B. Or, put negatively, there is no possible world in which A changes, and B does not change. To say, “in any possible world,” is to say, “necessarily.” So we can rewrite the positive characterization of supervenience this way: “A supervenes on B if and only if necessarily where A changes, there is a change in B.” Now, this possible worlds – or
necessity – claim can be cashed out either in metaphysical possibility, logical possibility, or nomological possibility.\textsuperscript{70}

We are now prepared to notice that multiple realizability of A by B is supposed to be allowed in cases where A supervenes on B. In other words, the determination is supposed to be one-way or asymmetric. An unhealthy human may differ from a healthy one by differing in any one of the physical features we mentioned. If someone possesses the property of being unhealthy it is not determined which subvenient physical base property is also possessed. But, if someone has the subvenient physical base property of, say, cancerous tissue, then it is determined that this person also has the property of being unhealthy. In possible worlds talk, imagine an object x and an object y in one world. Object x and object y could be indiscernible (i.e., no different) in terms of their A-properties while being discernible (i.e., different) in terms of their B-properties. This is so because A is multiply realizable by B. (They may be both unhealthy but for different reasons.) But, because B determines A, if object x and object y are indiscernible in terms of B-properties, then they also are indiscernible in terms of A-properties. (If they both have high cholesterol, then both are unhealthy.)

Now it is time to catch up this discussion with recent work. Recently some have argued that supervenience is more a statement of a problem, than it is a solution. What we need to know is what explains the supervenience, i.e., we need to know the superd supervenience relation. Indeed there has been a disillusionment about the concept’s usefulness. After a series of influential articles hit the scene (see Horgan 1993, Wilson 1999), philosophers became more and more convinced that supervenience lacks some important

\textsuperscript{70}Nomological possibility means what is possible in terms not of what is conceivable according to our concepts (that’s logical possibility), but what is possible according to the laws of physics of our world. Because of what inheres in the concept of “bachelor,” it is logically impossible to be a bachelor and yet be a married male. In this sense, the property of being a bachelor logically supervenes on being a married male.
desiderata. For just one example consider the summary of the situation by Lynch & Glasgow (2003):

… a growing consensus amongst writers on the topic indicates that nonreductive materialists face a special explanatory burden when invoking supervenience. Unless supervenience itself is materialistically respectable, it can hardly be taken to contribute towards the respectability of an upper-level phenomenon. As Terence Horgan and others have noted, materialistic respectability requires more than a supervenience relation in which the physical facts are “ontically basic” by virtue of fixing all other facts (see Wilson, 1999; Kim, 1990; Schiffer, 1987). The example of Moore makes the point: he held that non-natural moral facts supervene on the physical facts. Since these Moorean facts and the accompanying Moorean supervenience relation are not materialistically respectable, it is clear that, as John Heil puts it, we need to “move beyond formal characterizations of supervenience” (1998, p. 150). Consequently, Horgan has argued that nonreductive materialists must appeal to an upgraded “superdupervenience,” if supervenience is to do any work for their view.

One can try to defend non-reductive physicalism by coming up with a viable account of how mentality supervenes on, but does not reduce to, physicality. However I leave that project to others. Now I will reconstruct Kim’s use of supervenience in his criticism of non-reductive physicalists.

§

3.6 Kim’s Exclusion + Supervenience Argument

Kim’s version of the exclusion problem concludes that no mental event ever causes any physical event. But in other places Kim has put forth the same argument plus a supervenience thesis. With the addition of a supervenience claim Kim concludes that no mental events cause any mental events.

Kim’s overarching claim in this argument seems to be this: given supervenience and minimal physicalism – that is, given a metaphysical necessitation relation between M (mental properties) and P (physical properties), and given the completeness and closure of physical
causation – we can give an explanatory account for any selected event (mental or physical) without appealing to any other mental event. We can make the account in terms of P alone, and if we mention an M it will be merely as a by-product, an epiphenomenon of some P.

To prove his point, Kim asks us to imagine a case in which “an instance of mental property M causes another mental property M* to be instantiated,” (Kim 1998: 41). For example, we might think of a case in which the experience of pain (M) causes a desire for aspirin (M*). That is relation (1) in Figure 5: one mental experience causing another mental experience.

![Figure 5](image_url)

Kim sees a tension between M causing M* and a commitment to supervenience:

Under the assumption of mind-body supervenience, M* occurs because its supervenience base [let’s call it] P* occurs, and as long as P* occurs, M* must occur no matter what other events preceded this instance of M* – in particular, regardless of whether or not an instance of M preceded it. This puts the claim of M to be a cause of M* in jeopardy (Kim 2002: 176).

That is, according to supervenience M* occurs whenever P* occurs. So if P* occurs then M* necessarily must even if not preceded by M. So the fact that M* supervenes on P* seems for Kim to undermine, or put in “jeopardy,” M’s claim to be a cause of M*. 
But, of course, it might still be argued against Kim that M caused P*, which then is responsible for M*. This is represented by the unnumbered, dashed arrow in Figure 1. Kim considers this possibility, which he has elsewhere called “downward causation.” Kim says that it may be a plausible general principle that to cause a supervenient property to be instantiated, you must cause its base property to be instantiated. “To relieve a headache, you take aspirin: that is, you causally intervene in the brain processes on which the headaches supervenes. That’s the only way we can do anything about our headaches,” (2002: 176). So now we find ourselves trying to defend mental to physical causation. We need only recall, however, that M also has its subvenient base P; and “we must compare M and P in regard to their causal status with respect to P*,” (176). Again, “we begin to see reasons for taking P as preempting the claim of M as a cause of P*,” (176). Those reasons namely are physical causal closure and so-called principle of non-overdetermination. I will examine these reasons below. In any case, for those reasons, Kim says that P preempts M as a cause of P*. This allows Kim to conclude that “the most natural way of viewing the situation [our whole diagram] is this: P caused P* and M supervenes on P and M* supervenes on P*,” (177) and the M to M* (and M to P*) causal relations are only apparent, arising epiphenomenally out of the genuine causal process between P and P*.

This argument might seem to go through. But let’s examine as promised the reasons why, if P is the cause of P*, then M cannot be the cause of P*. The principle of physical causal closure, according to Kim, holds that for every physical event \( e \), some physical event \( c \)

---

71 For a discussion see Kim (1999).

72 We won’t bother with “the power of positive thinking,” or “trying to relax,” as headache relief because the problem of mental to mental causation is taken up in just a moment.

73 Again it is not to my purpose to point out the deficiencies in Kim’s formulation of the problem. For present purposes, I grant what he says.
is causally sufficient for $e$.\textsuperscript{74} “If we trace the causal ancestry of a physical event,” Kim says, “we need never go outside the physical domain.” Kim insists that if the non-reductive physicalist is to be a physicalist she must endorse the principle of physical causal closure.

What physicalism amounts to, Kim recognizes, is contentious. But to be a minimal physicalist requires, Kim argues, the rejection of Cartesian souls and nonphysical causally active entities.

As Kim says:

> To deny this assumption [physical causal closure] is to accept the Cartesian idea that some physical events need nonphysical causes, and if this is true there can in principle be no complete and self-sufficient physical theory of the physical domain. If the causal closure of the physical failed, our physics would need to refer in an essential way to nonphysical causal agents, perhaps Cartesian souls and their psychic properties, if it is to give a complete account of the physical world. I think most physicalists would find that picture unacceptable, (Kim 1993: 280).

This amounts to advocating the conjunction of materialist ontology and physicalist epistemology. Kim’s minimal physicalism assumes 1) that the universe is made up of physical stuff and only physical stuff; and 2) that the completed science of physics in some kind of future state will completely explain all the physical phenomena.\textsuperscript{75}

Let’s summarize the discussion of physical causal closure and in so doing make a transition to a discussion of the principle of non-overdetermination. Under the auspices of physical causal closure, mental events are locked out of the closed and sufficient system of causes in every legitimate causal history. Mental causes need never be appealed to, because for every event we observe we may assume that there was a sufficient cause which will always

\textsuperscript{74} Here I am following Kim. But note according to Baker, the causal closure of the physical holds, roughly that “every physical property-instantiation that has a cause at $t$ has a complete physical cause at $t$.” See Baker (1993: 78).

\textsuperscript{75} Recently some philosophers have begun to deny causal closure. It is not my strategy to argue directly against the principle of physical causal closure.
have been a physical cause. Remember that, in the dialectic, we are discussing whether M or P has causal relevance in relation to P* (which is the subvenient base of M* which we are ultimately interested in causally explaining). Notice that the only way there could be a legitimate mental cause (and, therefore, the only way M could legitimately cause P*) would be if it were possible to have two sufficient causes acting simultaneously. That is, we could avoid Kim’s conclusion that M doesn’t cause P* were we to allow that both M and P caused P*. This is still allowable under physical causal closure, which requires only that every event has a sufficient physical cause. It does not exclude it also having a sufficient mental cause. If it also had a mental cause then we would have an event that was “overdetermined.”

But, as we’ve already said, Kim argues that this would be analogous to some oft-cited examples. As Kim says, “A man is shot dead by two assassins whose bullets hit him at the same time; or a building catches fire because of a short circuit in the faulty wiring and a bolt of lightning that hits the building at the same instant,” (Kim 1993: 252). As noted earlier, it is dissatisfying to imagine that a situation so statistically unlikely happens in every last case of mental causation. Kim finds this unacceptable, and he institutes his non-overdetermination principle: If an event $c$ is causally sufficient for an event $e$, then no event $c^*$ distinct from $c$ is causally relevant to $e$. In combination with physical causal closure (which, again, says “for every physical event $e$, some physical event $c$ is causally sufficient for $e$”), the non-overdetermination principle guarantees that in a causal competition between M and P for causal potency in regard to P*, M will lose. M will be screened off.

So summarizing Kim’s argument in a more simple form, we get the following:

(1) The principle of non-reduction: For every physical event P and mental event M, P is distinct from M. (This represents a reasonable distinctness thesis for the non-reductive physicalist.)
(2) The physical causal closure principle: For every physical event $e$, some physical event $c$ is causally sufficient for $e$. 
The principle of non-overdetermination: If an event \( e \) is causally sufficient for an event \( e^* \) distinct from \( e \) is causally sufficient for \( e \).

(Therefore) Epiphenomenalism is true: For every physical event \( P \), no mental event \( M \) is causally relevant to \( P \).

The way this argument is formulated so far only concludes that there is no \( M \) to \( P \) causation. It says only that any time a \( P \) is caused to come into existence it was another \( P \) that is responsible (and responsibility is not shared with – or overdetermined by – a mental event). \( M \) may be distinct from \( P \) ((1) may stand), but \( M \) also never causes any \( P \). However, if you add the supervenience of \( M \) on \( P \), then you get the result that there is no \( M \) to \( M \) causation either.

The supervenience of \( M \) on \( P \): Every mental event \( M \) is metaphysically necessitated by some underlying physical event \( P \), whose causally sufficient antecedents (\( P' \)) are presumably sufficient for \( M \) as well.

Then, by the principle of non-overdetermination \( M \)'s mental antecedents are irrelevant to \( P \)'s occurrence. As Yablo says, “Here the mystery is how mental events, desires for example, can be making a causal difference when their unsupplemented neurophysiological underpinnings are already sufficient to the task at hand,” (Yablo 2002: 180). Kim’s claim is that the non-reductive physicalists cannot get what they want; their commitments are inconsistent and to regain consistency they must give up the distinctness or nonreductive thesis regarding mental and physical properties. He seems to suggest that they must give up Mental Causation in the earlier inconsistent tetrad.

It is interesting to note that Kim’s argument does not appeal to any special characteristic of mentality – such as the mental’s inherent normative or intentional characteristics – in order to show that mentality is not causal. He need only appeal to mentality’s supervenience on physicality and on certain characteristics of the physical realm.
This means that the problem could be generalized to other special sciences besides psychology. For example, if geological properties, which are appealed to in geological explanations, supervene on physical properties, which we assume they do, then according to Kim’s argument those geological properties are not really causally relevant in the explanations, only the basic physical properties are. To the non-reductive physicalist, who holds dear the autonomy of the special sciences, including psychology, this seems like a conclusion worthy of resisting.\textsuperscript{76}

\section*{3.7 An Initial Pragmatist Rejoinder: Baker and Burge}

I want to look at one way of responding to the worries which Kim’s arguments present to non-reductive physicalists in philosophy of mind. This way of responding consists of making a certain objection from the point-of-view of explanatory practice. The idea seems to be to claim that if your metaphysical commitments make mental causation appear never to happen, then there is something wrong with your metaphysical commitments. Kim has provided a preemptive defense against this style of argument (Kim 1998: 57-72). Kim’s main foil is an argument he sees being put forth by Lynne Baker and Tyler Burge.

Kim quotes Burge who says that the worries about mental causation “are symptomatic of a mistaken set of philosophical priorities. Materialist metaphysics has been given more weight than it deserves. Reflection on explanatory practice has been given too little,” (Burge 1993: 97). The objection, as Kim understands it, seems to be “that we should

\textsuperscript{76} Kim claims that his own position allows for the autonomy of the special sciences via a particular notion of reduction and a discussion of preemptive moves (about part-whole relationships) against the generalizing of his supervenience argument. We will not argue the point here.
look to explanations and explanatory practice, not to metaphysics, for guidance on the matter of mental causation,” (Kim 1998: 59).

The explanatory practices of ordinary life, and of psychology, commonly make attribution to intentional mental events. Our explanatory practice assumes “that intentional mental events are often causes and that psychological explanation is often a form of causal explanation,” (Burge 1993: 118). According to Burge, we have rather strong grounds on which to reject a metaphysics that says mentality is not causal. Kim finds that Burge is echoed by Baker, who suggests we “take as our philosophical starting point, not a metaphysical doctrine about the nature of causation or of reality, but a range of explanations that we have found worthy of our acceptance…If we reverse the priority of explanation and causation that is favored by the metaphysician, the problem of mental causation just melts away,” (Baker 1993: 92-93).

In response, Kim says that Burge and Baker are making a fundamental error, which amounts to eschewing their responsibility as philosophers to think about philosophical questions. Kim admits that, “as Burge says, our confidence in the truth of familiar intentional explanations does exceed our commitment to any recondite metaphysical principles.” However, Kim adds that he has never contravened the explanatory relevance of mentality. He says, “I doubt that very many of us who have worried about mental causation have actually been concerned about the possibility that our thoughts and desires might turn out to have no power to move our limbs,” (Kim 1998: 61). The worry about epiphenomenalism is not the worry that our minds cannot move our bodies; it’s the worry about how to create an

77 This is certainly debatable. I, for one, think that his argument generalizes such that if mental events are excluded so too are any properties supervening on properties of basic physics. However, as mentioned before, Kim thinks he can dodge the generalizing worry himself even if non-reductive physicalists cannot. I can disagree with Kim’s reasons for thinking mental causation is a metaphysical issue while still thinking that mental causation is a metaphysical issue.
account of the fact that our minds move our bodies. Kim says, “Our worries are not evidential or epistemological worries…the problem of mental causation is primarily a metaphysical problem. It is the problem of showing how mental causation is possible, not whether it is possible,” (Kim 1998: 61). Answering how mental properties manage to play the role in causal explanations that they do play is an exercise in metaphysics. Kim says, “The issue is how to make our metaphysics consistent with mental causation and the choice that we need to make is between various metaphysical alternatives, not between some recondite metaphysical principle on the one hand and some cherished epistemological practice or principle on the other,” (1998: 62).

As the entry on Mental Causation in the online Stanford Encyclopedia of Philosophy says, “The explanatory strategy [of Baker and Burge] would at best seem to be addressing only the ‘whether’ question, not the ‘how’ question.” The strategy seems to rest “on a conflation of what appears to be an epistemological notion (explanation) with metaphysical notions (causation and causal relevance),” (Robb & Heil 2003: Section 7.5). (Below I argue that in the case of his criticism of anomalous monism, it is Kim who does not take metaphysics seriously enough. Or, in any case, I argue he conflates a metaphysical understanding of an “in virtue of” question with an epistemic one. The theme of this metaphysics–epistemology distinction can be found throughout this dissertation.)

Kim thinks that Burge and Baker are not taking the problem of mental causation seriously with their deflationary suggestion that it will “melt away,” (Baker 1993: 23) given “inexpensive repairs,” (Kim 1998: 59). He says that they seem to want a “free lunch” when in fact “we need to make fairly drastic [metaphysical] adjustments if we are serious about coming to terms with the problem,” (Kim 1998: 59).
For my purposes here, I agree that the problem ought to be addressed as a metaphysical problem, as Kim says, although I am also sympathetic to Baker and Burge’s worries. In the end, my view, proposed in Chapters Four and Five below, will be a metaphysical solution but one that gives considerable importance (a kind of priority) to our epistemological practices. As such, I believe my view improves on Kim, as well as on Baker and Burge. The latter say, “we should look to explanations and explanatory practice, not to metaphysics, for guidance on the matter of mental causation.” But what that guidance would be is left vague. My view gives the details regarding the guidance. Still, my view must wait until later chapters. We should now return to the literature on the mental causation debate and consider a few more sub-debates so I can establish my attitude toward them.

§

3.8 Canonical Solutions

In this section, we evaluate some of the canonical solutions to the exclusion problem.

3.8.1 Eliminativist Reductive Physicalism

My Catastrophe thesis, if true, is not catastrophic merely in virtue of its name. That is, it is logically possible to embrace the Catastrophe thesis verbatim. One could affirm: “Mentality plays no causal role in the causal explanation of intentional action.” It does not become a liability because it is called Catastrophe; it is a liability because the proposition has a number of unhappy entailments. If mental properties cannot be appealed to in the explanation of intentional action, then every time someone gives a reason explanation, they are either speaking nonsense or saying something false. As discussed earlier (see Section 1.1.1),
eliminativism seems to entail *Catastrophe* which itself has unhappy entailment such the undermining of our folk explanatory practices and our self-understanding.

### 3.8.2 The Dual Explananda Response

The dual explananda response seems close to my view and has some merit. It is close in that it insists on a distinction among things-to-be-explained; and the critics violate the distinction. The main idea is that if there are two things to be explained, then we need to be more careful than Kim is with respect to the details of just what we are asking to be explained. Kim’s exclusion problem could be seen as arguing that by non-reductive physicalism’s lights physical properties are causally relevant to behavior and action while mental properties are not.

Ausonio Marras argues that there are two things to be explained (that is, two things to be relevant to) and that if there are two things to be explained, then we need to be careful in forming the problem. According to Marras’ solution, mental properties (and not neurological properties) will be causally relevant to intentional action and its mental properties, while neurological properties (and not mental properties) may be causally relevant to bodily movement physically described. Let’s look at Marras’s argument.

The causal relevance idiom says that events cause their effects in virtue of certain of their properties. Cause-events cause in virtue of a certain property of the cause-events, or in virtue of which kinds they fall under, or types they typify.

The idiom of causal relevance prefers to present matters in this way: The idiom does not say ‘c caused e’, it says

---

78 It remains to give an account of how principally to individuate the explananda or the levels of explanation.
c’s being F caused e’s being G.

But this is more helpfully expressed as follows:

c caused e, and this was so in virtue of c’s being F and e’s being G.\(^{79}\)

This formulation is a giant step forward for the dual-explananda approach because not only do you discuss the cause-event’s properties (as causally relevant) but you also discuss what property in particular of the effect-event you are interested in having explained. The non-reductive physicalist has often argued that the same token event may be described as either an intentional action or as physical behavior; that is, one effect-event can typify physical as well as mental properties. And so it may be the case that the physical properties of the cause-event are not relevant to the action-theoretic properties of the effect-event. To decide that question you need a theory of causal relevance. After all, a property F of c may be relevant to causing e to be G and irrelevant to causing e to be G* or I or many other properties of the effect-event. But with the explanandum individuated finely enough so we know we are asking for an explanation of an action described in action-theoretic terms, then it will be difficult for friends of Kim to insist that physical properties are relevant.

In this light, we can see that it is perhaps not coincidental that Kim, as it must be noted, often fails to provide the effect-event with the same structure as the competing cause-events. And some have noted that it is his failure to individuate the explanandum properly that allows the problem of exclusion to rear its head.

\(^{79}\) The “in virtue” idiom is more correct because after all we are saying that the properties are causally relevant not causes per se.
Here’s Kim’s own example. We can see that he embraces the causal relevance idiom, but only in relation to causes. He says:

\[
\text{c's being an event of the kind N (or having property N) caused B} \]

It is quite remarkable that the effect-event is not properly represented here as having the same structure as the cause-event. Kim says, “B, too, can be thought of as having a similar structure of an event being a certain kind; but for now this won’t be necessary,” (1990: 39).

The example being symbolized here is one in which George rises from the couch (B). This example of Kim’s does not conveniently fall into both an intentional description and a physical one. It is biased, perhaps, toward a physical reading. But a different example works better without changing what is at stake. Imagine instead that George is hailing a cab which could also be physically described as George’s arm rising to a 45 degree angle from the ground. According to Kim, the cause of B is either an event having a certain neural property or an event (Kim allows that it could be the same token event) having a certain belief-desire-intentional (BDI) property, such as wanting beer, etc. and therefore getting off the couch, or in our example wanting to go downtown in a cab to see the sights. Kim thinks that one competing explanation (say that of BDI psychology) will appeal to a cause-event which causes B in virtue of the cause-event’s being a psychological kind (S). Another competing explanation (that of neurology, say) will appeal to a cause-event which causes B in virtue of the cause-event’s being a neurological kind (N). This may be symbolized as such for explanation 1 and 2:

\[
\begin{align*}
\text{E1)} & \quad \text{c’s being an event of the type S caused B.} \\
\text{E2)} & \quad \text{c’s being an event of the type N caused B.}
\end{align*}
\]

Now it becomes a matter of deciding which property is relevant to bringing about B. For this you need a theory of causal relevance. But, before presenting a sketch of a theory of causal relevance, we need to translate E1 and E2 to express two concerns. First, we need to translate it into the “in virtue” idiom; and, second, we need to give the effect-event B the same structure as the cause-event, that of being of one kind or another. To do this we will write “B” as “e,” as is a standard for symbolic representation of an effect-event. That event which is extensionally equivalent to George’s hailing a cab, we said could be described in action-theoretic terms (let’s call that as falling under kind A) or in merely physical terms as a movement in space (let’s call that P). Then each explanation, E1 and E2, can be rephrased in two ways, such that:

- S=Psychological
- N=Neurological
- A=Action theoretic
- P=Physical

E1(a) c caused e, and this was so in virtue of c’s being S and e’s being A.
E1(p) c caused e, and this was so in virtue of c’s being S and e’s being P.
E2(a) c caused e, and this was so in virtue of c’s being N and e’s being A.
E2(p) c caused e, and this was so in virtue of c’s being N and e’s being P.

Now, which of these explanations is correct can only be answered with a theory of causal relevance. There is a lot to be said here, but a standard theory of causal relevance is formulated in terms of counterfactuals. Marras says:

Where c causes e, and where c is F and e is G, c’s being F is causally relevant to e’s being G only if the counterfactual ‘¬Fc → ¬Ge’ holds.\(^{81}\)

---

\(^{81}\) Marras (1998: 448). I am following Marras very closely here.
So we must now evaluate the counterfactuals entailed by each explanation E1a, E1p, E2a and E2p. They are as follows:

(1a) \( \sim Sc \rightarrow \sim Ae \)
(1p) \( \sim Sc \rightarrow \sim Pe \)
(2a) \( \sim Nc \rightarrow \sim Ae \)
(2p) \( \sim Nc \rightarrow \sim Pe \)

There is a prima facie way of deciding the truth of these counterfactual such that (1a) and (2p) are true and (1p) and (2a) are false.

(1a) \( \sim Sc \rightarrow \sim Ae \) T Comment: Would not have hailed cab if not desired to go.
(1p) \( \sim Sc \rightarrow \sim Pe \) F Comment: Hand would have risen to 45 degrees for any number of reasons.
(2a) \( \sim Nc \rightarrow \sim Ae \) F Comment: George’s desire is multiply physically realizable. Without N, A can still be the case.
(2p) \( \sim Nc \rightarrow \sim Pe \) T Comment: In principle, science could tell us which exact neurological configurations produce which exact bodily movements.

Thus, it is true to say along with (1a) that George would not have hailed a cab had he not wanted to go downtown in a manner that provided him a view of the sights. But it is false to say as (1p) does that his hand would not have raised to a 45 degree angle had he not intended to go downtown with a view of the sights, for his hand might take that position for a number of intentionally described reasons including blessing, saluting, etc. Furthermore, it is false to say as (2a) does that George would not have hailed a cab had his neurological state not have been exactly as it was, because by the generally accepted principle of ‘multiple realizability’ there is another neurological state which can realize the intention to hail a cab. Finally,
although it gives a lot of credence to the physicalist’s Whiggish hope for a perfect science of neurology, (2p) can be imagined to be true. That is to say, George’s arm would not have moved exactly as it did had he not had the exact neurological and muscular (physical) configuration that he did have.

While I believe that there is something right about Marras’s criticism of Kim, it is also true we could just complain that Marras has remained in the epistemic realm and has not yet dived into “real metaphysics.” What an objection like that would mean is this: It may be that mental properties are explanatorily relevant today (in our scientific climate) to intentional action and neurological properties are today in our scientific climate not explanatorily relevant to action. However, Kim could say, when we do metaphysics we may assume a future idealized science. This is a good objection to Marras’s argument in the terms in which it is put forth.\footnote{This is an important consideration for a theory like mine too. But we will not be dealing with it here.}

\subsection*{3.8.3 Yablo’s Proportionality Constraint}

Yablo (1992) offers a unique and clever solution to the exclusion problem. He says that causation is subject to a proportionality constraint, which is a constraint that mental properties can often satisfy while their realizers cannot. However, I would argue that Yablo is conflating causation with an intuitive notion of causal explanation. In articulating this criticism I follow closely Bontly (2005).

Yablo writes that “causes should incorporate a good deal of causally important material but not too much that is causally unimportant,” (1992, p. 274). This is his proportionality constraint. Bontly argues, against Yablo, that proportionality is not really a
constraint on causation but is rather “a pragmatic feature of our use of causal language,”
(332) and cannot save nonreductive physicalism from the exclusion problem. I think this
shows that Yablo is working with a non-extensional, intensional notion of causation, which is
a liability. Yablo has given us an account of general (non discipline-specific) causal
explanation, but not an account of causation.

Yablo works in the familiar framework of nonreductive physicalists who are
confronted by the exclusion problem. He embraces these three standard theses:

*The Causal Completeness of the Physical* (CCP): every physical event is causally determined
by prior physical events.

*The Irreducibility of the Mental* (IM): mental properties and events cannot be reduced to
or identified with physical properties and events. Instead, the relation between the
mental and physical is one of supervenience or multiple realization.

*The Causal Efficacy of the Mental* (CEM): mental events and properties can bring about
physical effects.\(^{83}\)

Now, these seem inconsistent with a plausible principle, which we’ll call *The Exclusion
Principle*:

*The Exclusion Principle* (ExP): if E is causally guaranteed by C, then no C* distinct from
C is causally relevant to E (Yablo 1997: 255).

Or as Kim (1993: 360) puts it: “a sufficient cause of an event excludes the claim of any other
distinct synchronous event to be a cause of that same event.”

To understand how Yablo thinks his proportionality constraint helps dodge the
exclusion problem we need to start at the beginning with the Goldilocks Principle.

As Bontly relates the fairytale:

When Goldilocks came to the house of the three bears, she found the house
abandoned. But she was hungry and tired, so she went inside, where three bowls of

\(^{83}\) Slightly altered from Bontly for clarity in this context.
porridge were laid upon the table. The first bowl was too hot, and the second too cold, but the third bowl was just right, and she proceeded to wolf it down.

Yablo’s proportionality constraint has it that causation must also be “just right.” As Yablo puts it: “causes should incorporate a good deal of causally important material but not too much that is causally unimportant,” (1992: 274). That’s the analogy. But to illustrate Yablo’s idea we do not actually need all three bears’ bowls. Suppose Goldilocks only burns her mouth on Papa Bear’s porridge. As Bontly says, “There are several porridge-eating events to which this effect might be attributed (or, alternatively, several different ways of describing one porridge-eating event),” (332).

On the one hand, we could say that eating porridge caused her mouth to be burned, but that seems insufficiently specific. […] So perhaps we should cite the exact circumstances of her porridge-eating and say that it was eating porridge with a temperature of precisely 205.63 degrees on a Tuesday in November that caused the burn to her mouth.

Merely citing her eating of porridge is not explanatory because porridge is often found at temperatures that do not burn mouths. But then citing details about the day and time is irrelevant; the exact temperature is overkill because there are other temperatures nearby which would also burn her mouth. “So perhaps we should say that the burn to her mouth was caused by eating scalding hot porridge, and that description is just right,” (Bontly 2005: 333).

What Yablo is proposing is that causation should follow the Goldilocks Principle and be “just right” between too much and too little information. More precisely, Yablo’s notion involves two counterfactual conditions, involving the concepts of “determinables” and “determinates”:

**Requiredness:** C is required for E iff C has no determinable C- such that if C- had occurred without C, then E would still have occurred.
Enoughness: C is enough for E iff for all determinates C+ of C, if C had occurred without C+, E would have still occurred.

Requiredness and Enoughness are together necessary but not sufficient for causation. “Red” is a determinate of the determinable concept “color.” And “Scarlet,” “Crimson,” “Fire Engine Red,” are determinates of the determinable “Red.” Thus, determinable and determinates work at one level or another and are level-relative. At one level “Red” is a determinate and at another level “Red” is a determinable. So let’s say that a certain bull will only charge at red capes. He will not charge at any color cape, only red. But he will charge at any red color cape, including scarlet, crimson and fire engine red capes. Intuitively then red should satisfy Requiredness while its determinable (color) does not; and red should satisfy Enoughness while its determinates (scarlet, etc) do as well. Intuitively red is both required and enough to cause the bull to charge.

Now the relation between determinables and determinates, Yablo has noted, is very similar to the relation between mental events and their realizers.

If something has a determinable like coloredness, then it must have some determinate form of coloredness like redness. Having redness entails having coloredness, furthermore, but of course the converse does not hold. Similarly, if mental properties are physically realized, then having a given mental property M requires having some physical property P that necessitates, or determines, the having of M, though again the converse needn’t hold, (Bontly 2005: 333).

In other words, being red entails being colored as being a determinate entails being some determinable. But being colored does not entail being red, for there are other colors to serve as the determinate of the determinable color.

How Yablo solves the exclusion problem with these tools is as follows.

Suppose I feel a pain in my hand (M) and move my hand away from the source of irritation (E). If we contrast M with one of its higher determinables – like feeling
something – we see that M is required for E. If I had felt something else, like a tickle, without feeling pain, then E may well not have occurred. If, on the other hand, we contrast M with its determinate realizer P, we see that P is not required for E and M is enough. For had M been differently realized, say by P*, E would quite probably still have ensued, or so we shall assume. Thus Goldilocks yields the conclusion that M caused E and P did not, even though P was (by hypothesis) sufficient to guarantee that E would occur.

In other words, the pain (M) in my hand causes me to move (E) my hand. M satisfies Requiredness and Enoughness. This is so because the pain and not any old feeling would cause my hand to move; and, going the other way, P (one determinate realizer) is not required for E.

For a different realizer P* would (counterfactually) also have caused E.

3.8.3.1 Problems for Yablo’s Proportionality Constraint

Consider the following anecdote. There is lesson that has application to my distinction between causation and causal explanation.

Suppose Socrates drinks the hemlock and dies. As a matter of fact, he guzzled the hemlock, but let’s say he would still have died had he instead sipped the hemlock. Yablo’s proportionality constraint would hold that his guzzling was not required and so not proportional to his death. But does it follow that his guzzling did not cause his death? To answer this question, suppose that guzzling the hemlock was proportional to, and caused, his dying quickly and painlessly, Bontly (2005) says. He would have died a lingering and unpleasant death had he sipped it. So, his guzzling the hemlock caused him to die in a certain way. But if guzzling caused him to die in a certain way then it is hard to avoid the intuition that his guzzling caused him to die simpliciter. Or again:

Suppose I speed through a radar trap at precisely 41 mph over the limit and get a ticket for $250. If I had sped through at merely 26 over, the ticket would merely have been $175. So my speeding through at 41 over was proportionate to, and thus caused,
my getting the particular ticket I did. But if it caused me to get a particular ticket, then *a fortiori* it caused me to get a ticket *punkt* (340).

Yablo’s proportionality constraint would have the guzzling *not* be a cause of Socrates’ death, since guzzling does not meet the *Requiredness* criterion. But we have just seen that there is an intuition that the guzzling was a cause of his dying a certain way and thereby a cause of his dying. Bontly diagnoses what’s going in the following manner:

… if Socrates’ drinking the hemlock was enough for his dying, then we would not in many contexts *describe* [my emphasis] his guzzling it as the *cause* of his dying. That, however, is consistent with the thesis that his guzzling it was a *cause* of his dying, which it does seem to be, since replacing the guzzling with some other determinate may well affect how the death comes about. It’s just that we wouldn’t cite guzzling as the *cause* of death, because there is another event (or if you prefer another way of describing the same event) which is preferable on pragmatic grounds – that is in virtue of our interests and purposes in giving and receiving causal explanations.

Yablo’s argument for the non-proportionality of Socrates’ guzzling the hemlock and therefore for his notion that guzzling did not cause Socrates’ dying is pumped by the pragmatics of the particular context or level of explanation at which the question is pitched. Drinking the hemlock is all it would take to causally explain Socrates’ death in this relatively neutral, undifferentiated context where a contrast space has not been specified. But the importance of context should not be underestimated, Bontly says. It is easy to come up with a different context surrounding Socrates’ death where a determinate is a cause even where not required.

Suppose (as is in fact the case) that hemlock comes in two varieties: ‘western water hemlock’ (*Cicuta douglasi*) and its less toxic cousin ‘poison hemlock (*Conium maculatum*). In sufficient quantities, either one is sufficient to cause death, but let us suppose that it is the latter that Socrates actually drank. Now since *Cicuta* would have done just as well, his actually drinking *Conium* was evidently not required (meaning that the event of his drinking the plain ole (disjunctive) hemlock was enough (Bontly 2005: 341, altered slightly for clarity).
So his drinking *Conium* is in the same position isomorphically, but one level down deeper, as his guzzling was above. That is, the guzzling was not proportional but still seemed to be a cause. *Conium* is here not required and therefore not proportional, but if it caused him to die in a certain way (perhaps faster for being more poisonous) then it caused him to die *simpliciter*. But to show that this all has to do with the pragmatics of the context, let’s change the context and see if we can get the *Conium* to seem required and thereby proportional.

Certainly a coroners report would say *Conium* and not *Circuta*. Rewriting history a bit, imagine Socrates to have fallen victim to foul play; the police have two suspects, only one of whom had access to extract of *Conium*. In that case, knowing which type of hemlock was involved may well prove crucial to a conviction, in which case it would be quite natural to say that it was his ingesting *Conium*, and not merely his ingesting hemlock, that caused his death (341-342, altered).

That shows, I believe, that we are dealing in causal explanation. Let’s pursue a brief reverie. One event and not another (or one way of describing one event and not another way of describing it) will seem to be causal with respect to an effect depending on the context in which we seek a causal explanation. As we push through ever deeper explanations we push through levels of causal explanation. When do we reach causation? Perhaps we don’t. Any empirical explanation has the structure of explanatory relevance as we’ve been discussing in regards to the pragmatics of contexts of causal explanation. There is always the possibility of a deeper empirical explanation and the prospect of a context in which we would be interested in citing that feature as the cause. But one does not actually “arrive” at causation. Instead perhaps when I say causation is metaphysical and not epistemic I mean that causation is a posit made in the act of offering a causal explanation.
3.8.4 Horgan’s “Quausation”

Horgan (1989) introduces a 4-place relation he calls “quausation”; he then tries to motivate the “problem of mental quausation.” The quausation relation is expressed by the locution “c qua F causes e qua G” and is explicated as follows:

For any two events c and e and any two properties F and G, c qua F causes e qua G iff:

(i) c causes e;
(ii) c instantiates F;
(iii) e instantiates G; and
(iv) the fact that c instantiates F is explanatorily relevant to the fact that e occurs and instantiates G.

In order to have a full understanding of what is meant by quausation, we would clearly have to understand what is meant by explanatory relevance. Horgan offers a fairly standard counterfactual account, stipulated to avoid vacuity and back-tracking counterfactuals.

The interest for our purposes is not in Horgan’s account of explanatory relevance but in the fact that his attempt to motivate the quausation problem reveals how he misunderstands the dialectic of the mental causation debate itself.

Horgan says the problem of mental causation is that c may cause e but it may not be the case that c qua F causes e qua G. To illuminate why this is a problem, Horgan makes reference to two now-famous intuition pumps: Dretske’s soprano and Sosa’s loud gun shot.

Meaningful sounds, if they occur at the right pitch and amplitude, can shatter glass, but the fact that these sounds have a meaning is irrelevant to their having this effect. The glass would shatter if the sounds meant something completely different, of if they meant nothing at all. This doesn’t imply that the sounds don’t have a meaning, but it does imply that their having their meaning doesn’t help explain their effects on the glass. To know why the glass shattered you have to know something about the amplitude and frequency of these sounds, properties of the sound that are relevant involved in its effect on the glass (Dretske 1989).

A gun goes off, a shot is fired, and it kills someone. The loud noise is the shot... in a certain sense the victim is killed by the loud noise. But not by the loud noise as a loud
noise, but only by the loud noise as a shot, or the like… The loudness of the shot has no causal relevance to the death of the victim. Had the gun been equipped with a silencer the shot would have killed the victim just the same (Sosa 1989).

These two cases are meant to pump the intuition that “Sometimes the cause of a given effect,” will have “a certain property which is not appropriate to cite in a causal explanation of that effect,” (49). This means that something might be a cause without its being a cause in virtue of some property, X.

Horgan explains his point this way. In the following sentence the first “and” needs to be turned into a “because.”

(1) He exercised and he wanted to reduce and thought exercise would do it.

Proposition (1) says only that the agent possesses a “primary reason” that rationalizes his exercising. Horgan then argues that Davidson (1963) holds that all we need to add is proposition (2) to get what we want.

(2) His exercising was caused by his desire to reduce and his belief that exercise would do it.\(^{85}\)

Horgan counters, however, that (2) is not sufficient for (3):

(3) He exercised \textit{because} he wanted to reduce and thought exercise would do it.

---

\(^{84}\) See Chapter One.

\(^{85}\) Horgan adds an “in an appropriate way” clause to block deviant causal chains. I’ve elided it for simplicity’s sake.
What Horgan is saying is that it is possible that a primary reason be a cause of an action without it being possible for us to say that the agent took the action *because* of the primary reason. Maybe it is a cause qua neurophysiological state of affairs. That is, it might be a cause but it may not yet be a cause qua rationalizing reason. That’s the problem of mental quausation.

However, the way Horgan has set the problem up indicates he misunderstands the real dialectic of the mental causation debate and the difference between causation and causal explanation. Let me start as follows.

The multiple reasons argument says that it is possible that an agent have an appropriately rationalizing reason for an action and yet when the agent takes that action that reason did not produce, bring about or otherwise cause the action. This situation is not the one Horgan calls the problem of mental quausation. Horgan is highlighting a situation in which a reason that rationalizes an action is in fact a cause of an action but is not a cause qua reason. However, why should that bother us? Horgan says that we expect it but do we? He says, that is, “We believe not merely that reasons *are* causes, but also that people act the way they do *because* they have reasons which rationalize the actions,” (51). But there is something wrong about what he is saying.

The first problem is one I can best address after the apparatus of Chapter Four is on the table. But suffice it here to say the following. Horgan has not specified clearly enough the sense of an “in virtue of” question he means with his “because.” If he means that reasons need to be able to be referred to in causal explanations, then the question he asks is easily answered. But if he means that reasons need to be the causes that they are in virtue of being a reason I submit he does not know what he is talking about.
The second problem is a problem with the implications of what Horgan is saying.

Consider the soprano example again. Yet this time let’s actually spell out the analogy to the mental case.

Meaningful sounds, if they occur at the right pitch and amplitude, can shatter glass, but the fact that these sounds have a meaning is irrelevant to their having this effect.

The analogy would be as follows:

Rationalizing reasons, if they are instantiated by the appropriate neurophysiological underpinnings, can cause actions, but their being appropriate rationalizations is irrelevant to their having this effect.

By “can cause actions” does he mean that rationalizing reasons can be referred to in causal explanations or does he mean that rationalizing reasons can be related by the causal relation to actions? He does not say. But neither way creates a problem. (See Chapter Four.)

For instance, the loudness of the shot and the meaning of the notes the soprano sang were not causally explanatory. That is all the intuition pumps show. Reference to those properties is not explanatory while reference to some other properties would be. Here the analogy to the mental case finally breaks because reference to a rationalizing reason for an action (as a rationalizing reason) that actually brought the action about is causally explanatory, contra to Horgan’s misplaced worry. So there’s no problem of mental quausation there.

Now with respect to causation, Horgan readily admits that the rationalizing reason is a cause of the action. That’s clause (i). So there’s not a problem of mental quausation there.

The last thing that Horgan could mean is this: the rationalizing reason which is a cause of the action needs to be the cause that it is in virtue of being a rationalizing reason. But as I will argue in Chapter Four, why does this question need to be answered in mental terms? It seems like it will be answered in physical terms and that this is to be expected. It has no
epiphenomenalist ramifications. As I show in Chapter Four, the question of in virtue of what properties a cause is the cause that it is is a different question than the question of which properties are those properties reference to which is causally explanatory with respect to the effect coming to have the properties it does in the explanandum.

Let’s consider yet another issue to which Horgan does not speak clearly enough. Consider: If c qua F causes e qua G does that mean that the relata of the causal relation are “c qua F” and “e qua G” or is the relata of the causal relation still c and e? Horgan doesn’t say. On a reasonable understanding of causation and causal explanation there is no problem of mental quausation. On my understanding of causation and causal explanation there is no problem of mental quausation. But Horgan has not specified exactly what he means by the relation of causation, nor has he discussed its relationship to causal explanation. But, by my lights, whichever he might mean is not problematic.

§

3.9 Anomalous Monism

“Non-reductive physicalism” is one name for a contemporary philosophical solution to the problems of mental causation. Its goal is to account for the natural world of causes without implausibly reducing explanations of human action to deterministic causal laws, or the like.86 Perhaps the urtext of non-reductive physicalism is Davidson’s anomalous monism (1963,

86 There are a great variety of non-reductive physicalism theories in the literature. Certain functionalist theories count as versions of non-reductive physicalism (others do not). The role-functionalist theories of Putnam (1967, 1975) and Fodor (1974) are examples of non-reductive functionalists theories. A priori reductive functionalist theories include Armstrong (1968) and Lewis (1973). A posteriori reductive (non-role, filler-) functionalism is defended by Kim (1998).
1967, 1970, 1993). However, anomalous monism is actually importantly different from the
generic non-reductive physicalism that suffers the exclusion problem.

The three premises of anomalous monism are relatively plausible notions, at least prima
facie. In fact, Davidson originally assumes them without argument.

*Interaction*: Some mental events cause some physical events.

*Anomalism*: There are no strict laws connecting events under mental descriptions.

There are no strict laws connecting mental events with physical
events, such that mental events could be predicted or explained by
physical events.

*Cause-Law*: When events are related by the causal relation there is a strict law of
necessity connecting the events at least under some (physical)
description.

*Token Identity*: Every mental event is numerically identical to some physical

If understood appropriately, the first three lead deductively to the fourth proposition, *Token
Identity*.

The way the argument works is like this. *Interaction* says that some mental events cause
some physical events. Given the cause-law thesis this means that there are strict laws covering
such (singular) causation between events. But isn’t this denied by *Anomalism*? No. By properly
countenancing the token-type distinction there is no inconsistency.  

---

87 It is relatively common in the early reception of Davidson for commentators to take the three
premises of anomalous monism to be prima facie inconsistent. Honderich argues that the
conjunction of the first and second “denies” (1982: 23) the third. “Taking the strict law principle
seriously makes it seem as though the initial premise of the argument was simply false and mental
events cannot causally interact with physical events at all. In the author’s [Davidson’s] view,
however, the conflict is only apparent,” (Harbecke 2008: 119).
physical event and when a mental event is causally related to a physical event it is their physical properties which instantiate a strict law as per the cause-law thesis. So there’s a law connecting cause and effect in physical terms; but it is ruled out that there is any law connecting them in mental terms. Since the mental event is causally related (by Interaction) there must be a law couched in physical terms (by Cause-Law and Anomalism) which means the mental event must have physical properties, i.e., Token Identity. The argument is often called “groundbreaking,” (Sandis 2009: 1) and “ingenious” (Yalowitz 2005, Hutto 1998, Blackburn 2005, Smart 2007) in part because it argues deductively from plausible premises to substantive conclusion in line with naturalistic physicalism. Anomalous monism caught the interest of a generation of philosophers.

The argument goes through and Anomalous Monism is often considered as a system of four propositions. Let’s consider each proposition in turn and note any auxiliary tenets of anomalous monism such as the extensionality of causation, and its unique take on the causal relata. We have discussed the exclusion problem which looks a lot like Davidson’s argument to token identity, so a lot of this looks familiar. But in fact anomalous monism is not the same as the generic non-reductive physicalist we were working with in setting up the exclusion problem. Only the first premise is the same, so we can forego discussion of Interaction.88

3.9.5.1 The Cause-Law Thesis

According to this premise – also called The Principle of the Nomological Character of Causality – singular causal interactions are covered by strict (i.e., exceptionless) laws. That is, if an event c

88 Some of the premises in the argument of anomalous monism, I take as commitments only because Davidson does. Others I have my own reasons for adopting.
causes an event e, then there are properties H and J such that c instantiates H and e instantiates J; and all H instantiations nomologically necessitate J instantiations.

First, a few minor notes that may or may not be obvious. Anomalous monism asserts that there is, metaphysically speaking, some contingently necessary law covering every causal interaction among singular events. However, this is consistent with there being, epistemically speaking, ceteris paribus laws in the special sciences. Also, “strict” is not equivalent to deterministic. Indeterministic laws are consistent with the cause-law thesis because probabilistic versions of strict laws are possible as well, Davidson (1970: 219) affirms. What distinguishes a strict law from other kinds of laws (e.g., ceteris paribus laws) “is not so much the guaranteeing of the effect by satisfaction of the antecedent as the inclusion, in the antecedent, of all conditions and effects which could possibly prevent the occurrence of the effect,” (Yalowitz 2006: Section 3.1).

Where are we to find strict laws? Much of the secondary literature takes it that only physics issues in strict laws. Thus the properties, H and J, are properties of physics. However, Davidson is more chary: “It is plausible that there is a set of concepts (perhaps there are many such sets) which lend themselves to the formulation of” strict laws. Let us call these the concepts of physics,” (1995: xy). Davidson here is not yet talking about the social practice of whatever is done in physics departments; he is not talking about an idealized physics either; he has only used certain letters “P,” “H,” “Y,” “S,” “I,” “C,” “S” to refer to his notion that there is a set of concepts which lend themselves to the formulation of strict laws. I’d argue that we may as well call it “schmysics.” It need not be “physics,” neither contemporary nor idealized.

It has sometimes been noted, in any case, that ceteris paribus laws might need underwriting by strict laws. See Kim (2006).
All anomalous monism asserts is that there is a strict law connecting causal events. It holds only that there is such a law; it need make no commitment to our coming to know the law. The cause-law thesis is after all a metaphysical claim not an epistemological one.

3.9.5.2 Anomalism of the Mental

I am going to assume this premise. I am not going to reconstruct Davidson’s arguments for the anomalism of the mental in their entirety. I just want to describe what Davidson holds, which includes his epistemic arguments and his metaphysical conclusions.

The anomalism premise has two parts: there are no psychological laws – left to right, so to speak – and no psychophysical laws – neither up to down nor down to up. First, Davidson holds that the intentional action explanations that make essential reference to mental properties (reasons) do not invoke psychological laws, but instead are broadly “hermeneutic.” (See Chapter One.) Actions are explained by being made “intelligible,” an endeavor which falls under *Verstehen* and not *Erklären* (see Chapter One). We explain an agent’s action by reference to what would be rational for the agent to do given the beliefs and desires we attribute to the agent.

What grounds might there be for saying that a person going downstairs is off to fetch the dinner wine if, for example, we know they believe that the wine is upstairs, or if we know they want beer with dinner? And, supposing they do go fetch the wine, what ground could there be for saying that they desired to fetch the beer, unless we also hold that they believed that the wine they fetched is after all beer? The contents of the mental states it makes sense to ascribe to agents in order to explain their actions depend upon their interrelating holistically so as to render the overall pattern of their mental states and actions roughly rational, that is, roughly so that they meet normative standards according to which their actions are appropriate in light of their beliefs and desires (Manning 2006: 472).

Second, Davidson holds that there are no strict laws connecting mental with physical events – none by virtue of which mental events could be predicted or explained, even from full
knowledge of the physics of the relevant physical event. This means to be a denial of bridge laws. It cannot be a denial of laws covering events in succession because the cause-law thesis affirms that there are such laws, albeit only in the physical vocabulary and albeit quite unknown to us. I'll explain.

According to the cause-law thesis, there is some law connecting each mental event that is a cause of a physical event to that physical event. Because of token identity, the cause-law thesis connects (in succession, left to right) events of both types to events of both types. After all, a mental event is also a physical event. They are lawfully connected qua physical event but they are connected. There are no mentally-formulated laws connecting events of both types whereas there are (in principle) physically-formulated laws connecting events of both types, but these latter we may not necessarily come to know.

Now, Davidson gave at least three reasons for the anomalism premise; each reason worked for both parts, left to right and up and down, any place mentality shows up. They are “holism with respect to particular attributions, indeterminancy with respect to systematic interpretative frameworks and the responsiveness of mental ascription to an ideal of rationality,” (Yalowitz 2005). “According to holism, particular mental states can be cited in explanation of behavior only in the context of other mental states, whose ascription in turn depends on others” (27). This is supposed to make psychophysical laws impossible.

The thesis of indeterminacy of translation says that “there are empirically adequate but non-equivalent complete frameworks for assigning linguistic meanings and mental states to a person on the basis of his behavior, and that there is no fact of the matter that determines that one but not other such frameworks is correct,” (27-28). “For if all the

---

90 We discuss below the problem of which physical event is relevant to the given mental event. This is the identification problem: Anomalous monism needs a principled way of picking out the physical event to which the given mental event is identical.
physical facts are consistent with different psychological assignments, then it seems that knowing all the physical facts could not tell us whether some mental states were true of some person… just as mental anomalism maintains,” (28).

3.9.5.3 Token Identity

What do the three premises we have just discussed entail? As discussed above, in the early days of reaction to Davidson (1970), critics took the three premises to be inconsistent. We now understand that what makes the argument go through is a proper countenancing of the token-type distinction and an understanding that the causal relation’s relata are events (property bearers) not properties.

Once again the argument works like this: there must be a law connecting the causally related events M1 and P2 but there is no strict law of the form, M1 → P2, so there must be some other property of M, namely P1 which is covered by a strict law with P2. M1 is the same event as P1. The token identity is a metaphysical claim. It says that there is such an identity and, further, that there must be. But do we know which physical event the reason is identical to? (Problems surrounding answers to this question are discussed below in the section on the so-called “identification problem,” which is about determining what’s identical to what in token-identity theories.)

Token identity says that every mental event is numerically identical to (or just is) some physical event, although not every physical event is a mental event. This gives us an understanding of how a kind of mental causation is possible. The given mental event (the reason) is shown to be identical to a physical event which thereby falls under a law and thereby can be a cause. So the reason is identical to the thing that counts as a cause. (Challenges to this aspect of the view are discussed below in the section on the qua problem.)
One problem to briefly consider is the challenge brought up by Leder (1995). He argues that to find the token identities it would require a considerable amount of type-identities. “The only empirical evidence for specific token-identities could be type-identities between other mental and physical properties,” (p.42). But this is to confuse, again, epistemology with metaphysics. That there is a token identity is known a priori. It would be an epistemological endeavor to attempt to find what’s identical to what.\(^91\)

### 3.9.6 Anomalous Monism Dodges the Exclusion Problem

The basic thrust of the exclusion problem is that one event out competes another for causal efficacy. But if Davidson’s argument achieves token identity then the two events are one and the same. And it’s hard to conceive how an event can compete with itself. (And if it does, it’s hard to see how it could lose.) So, the anomalous monism framework rules out the exclusion problem.

Instead of the exclusion problem, anomalous monism falls foul of the “identification problem” and “the qua problem,” which are quite different from the exclusion problem. Critics of anomalous monism in fact press the qua problem not the exclusion problem when they press “problem of mental causation” against anomalous monism. But we should take this opportunity to note again the differences between the exclusion and qua problems. First, the exclusion problem makes use of *Causal Closure* while the qua problem for anomalous monism uses the *Cause-Law Thesis*.\(^92\) Causal Closure is an existential claim that a cause exists. The *Cause-Law Thesis* is a claim about the nature of causation. The main question in the qua

---

\(^91\) Certain identities or neural correlates may be suggested or “discovered” but psychophysical laws allowing for the exceptionless prediction and explanation of action is ruled out by *Anomalism*.

\(^92\) Also called the *Principle of the Nomological Character of Causation*. 

126
problem framework – in virtue of what is c a cause of e? – does not arise in the exclusion problem framework until a theory of the nature of causation is adopted.

§

3.10 Anomalous Monism and the Identification Problem

Before pursuing the qua problem at length we should consider briefly the identification problem.

3.10.1 The Identification Problem

Anomalous monism asserts a physical-mental token identity and denies that there are psychophysical laws. Proponents of the identification problem argue that “if there are no psychophysical laws connecting mental and physical descriptions of events it must be a brute fact or coincidence that physical events have the properties they do,” (Campbell 2008: 31, italics added). That is, there is nothing in Davidson’s account – apparently – that would allow us to provide a principled relationship between the physical and the mental; that is, nothing “that would allow us to understand how it is that a particular event with certain causal powers is also the mental event it is,” (31-32). In other words, why does this one event have both the physical and mental properties that it does have? If given P there seems no principled way to pick out the relevant M; if given M there seems no principled way to pick out the relevant P. If there is no way to move from the physical to the mental, then the mental properties that physical events do indeed have will seem unfortunately arbitrary. There seems to be no way to understand, in Davidsonian terms, why this physical event here, so to speak, is identical to this mental event here. The connection is unexplained, the identification problem says. “That
any event should be a token of both types (on an occasion) is beyond explanation,” Stoutland (1980: 358) says, by which he means brute or unexplained. “For if it is a brute fact without explanation that the reason that explains an action just happens to be identical to the physical event that causes it, we seem to have fallen short of giving reasons genuine causal efficacy,” (32).

This problem is in fact one of the earliest brought to bear as a challenge to anomalous monism (Honderich 1982, Smith 1983). It has not been addressed much since, as a massive discussion grew up around problems of mental causation, in particular the exclusion problem and the qua problem. The identification problem will have to be met in order for anomalous monism to give an account of genuine mental causation.

Let’s begin by asking the anomalous monist how we are to pick out the physical event which is identical to a given mental event. If this is suggesting that one could pick out the physical event without a mental event already “in hand” so to speak, then it is not only violating the distinction between metaphysics and epistemology but it is violating what I take to be the order of explanation between the two, which we established in the discussion of “the priority of the phenomenon” thesis. This is what certainly seems to be happening in Yalowitz (2009): “If there are no explanatory relations between mental and physical properties, how is it possible that the psychological generalization that some individual will (given that he has certain reasons) open an umbrella on a certain occasion, predict an event that is also (under a different description) predicted by the physical laws?” (Yalowitz 2009: 48) This latter clause says “predict,” which is an epistemic notion.

According to the Priority of the Phenomenon Thesis of Chapter One, first our epistemology of the reason delivers its result. Then and only then do we try to give a
metaphysical accounting to understand how that reason could be a part of the successful explanation that it actually is a part of. Epistemological success first, then metaphysics.

To further explain what I mean by the order of explanation between epistemology and metaphysics, consider Jackson and Pettit’s (1990) doctrine of “program explanation.” Jackson and Pettit make a relevantly similar point. According to their project, a property can be causally relevant to an effect without being itself causally efficacious with respect to that effect. The causally efficacious properties are those, Jackson and Pettit say, that are “lower down,” so to speak, most likely physical properties; while causally relevant properties are those that feature essentially in a so-called program explanation, which is really just any adequate and true higher order explanation. For example, consider two explanations operating at different levels or orders of explanation: (A) Because some of its atoms decayed, versus (B) Because these particular atoms decayed. A and B are both explanations of the same phenomenon, let’s assume; A is a higher order explanation containing an existential quantifier and it “programs” for the particular causally efficacious properties featured in the explanation operating at a lower order or level. Jackson and Pettit ask: “How are we to describe the relationship between such a [causally relevant but causally inefficacious] property and an effect?” (1990, p.114). Their answer:

The realization of the property ensures – it would be enough to have made it suitably probable – that a crucial productive property is realized and, in the circumstances, that the event, under a certain description, occurs. The property-instance does not figure in the productive process leading to the event but it more or less ensures that a property-instance which is required for that process does figure. A useful metaphor for describing the role of the property is to say that its realization programs for the appearance of the productive property and, under a certain description, for the event produced.

The fact that epistemology comes “first” before metaphysics works as a response to the identification problem. Where Jackson & Pettit say “programs for” I could say “posits.” It
doesn’t matter what I say instead of “programs for.” What I find useful in Jackson and Pettit is the headspace it gets one into. On my view it would be epistemology “programming for” metaphysical commitments. The successful explanation (in the epistemological domain) imputes to us certain metaphysical commitments with respect to causation.

In a slogan, “epistemology posits metaphysics” or epistemological success imputes to us certain metaphysical commitments. This is a “kantian” – small “k,” big “ian” – attitude towards metaphysics insofar as the position it gives to epistemology is prior to metaphysics. That is, there is no point to speculative metaphysics – no point to trying to decide how many angels can dance on the head of a pin – aside from the purpose metaphysics plays in grounding success in epistemology. I return to this discussion in the Conclusion to this dissertation.

§

3.11 Conclusion

In this chapter we have surveyed the mind-body problem and the mental causation debate. We focused on the popular exclusion problem championed by Kim. We also conducted a small literature review of the mental causation literature with respect to some canonical attempted solutions to the exclusion problem.

In the next chapter, we discuss what has been the central critical issue with respect to anomalous monism – namely, the qua problem.
CHAPTER FOUR

ANOMALOUS MONISM AND THE QUA PROBLEM

4.1 Introduction

Criticism of anomalous monism is widespread and extremely influential.\footnote{In my view, the non-reductive physicalist’s trouble with the “exclusion problem” for events and/or properties traces back to this criticism of Davidson. As Bennett (2006) writes: “[I]t makes good sense to talk about the causal efficacy of properties, to say that \(c\) (your thirst, say) caused \(e\) (your movement towards the kitchen) in virtue of or qua some but not other of its \([c’s]\) properties. This idea – and the worry that establishing the causal efficacy of particular (token) mental events… does not guarantee the efficacy of event-types or properties – attracted attention in the wake of Davidson’s defense of anomalous monism,” (p. 34).} Kim (1993b) writes that philosophers believe with “an impressive if unsurprising unanimity” (20) that anomalous monism entails epiphenomenalism about mental properties.\footnote{Earliest examples include Stoutland (1980), Honderich (1982), Sosa (1984), and Dretske (1989). Recent recapitulations of the criticism occur in Harbecke (2008) and Campbell (2008). Defenses of Davidson against type-epiphenomenalism include Gibb (2006), which asserts confusion about properties and predicates on the part of the critics, and Campbell (1997, 2003b), which pivots on causation as an extensional relation. My tack is different from both insofar as I allow the critics the entailment of type-epiphenomenalism, but argue it is no liability given proper understanding of other auxiliary propositions.} Davidson puts forth, as we have noted, three theses, roughly: (i) mental events cause physical events (ii) causal relations are backed by strict laws and (iii) there are no strict laws of psychology. These may look inconsistent. The way Kim and McLaughlin (1993) present the objection, however, is, not as a matter of inconsistency. Instead, they propose to allow Davidson all of his claims – including auxiliary theses such as token identity and extensional causation – but argue that his final position unhappily entails mental type epiphenomenalism. This being the case, the...
way for me to adjudicate the matter best would be to present what all parties agree to and then to consider what is in fact entailed. I must also adjudicate whether this entailment is a liability or not.

In the end, my conclusion in the following sections will be that were mental type epiphenomenalism entailed by anomalous monism in exactly the way the critics formulate it, then this would nevertheless not be any liability for anomalous monism. The structure of Kim’s actual argument seems to be that anomalous monism entails type-epiphenomenalism which itself entails an additional thesis, namely that mental properties play no role in causal explanations of action, i.e., Catastrophe. But Kim is wrong about the second entailment: even if anomalous monism entails type-epiphenomenalism, it is not entailed that mental properties play no causal role in action explanation.

So my position amounts to saying that the critics misunderstand what their objection actually amounts to. The qua problem qua qua problem is really about the metaphysical grounding of the causal relation which is a matter separate from the role of mental properties in folk psychological explanation. In order to show this, I analyze the representation of causal explanations that all parties to the debate agree upon, e’s being F caused e’s being G, and all that follows from it, which all parties are committed to. In the next part of this chapter, I articulate and defend an account of the metaphysical grounding of the causal relation that does not appear in the mental causation literature. But let me begin with a more complete reconstruction of the debate.

§
4.2 The Criticism

The widely accepted objection against Davidson’s anomalous monism is that the candidate
token event may cause the effect but it does not cause it qua mental event, or, in other
words, not in virtue of its mental properties. This allegedly unhappy result is
straightforwardly entailed by the tenets of anomalous monism, say the critics. Recall that
Davidson is committed to the following:

1. Mental Causation: some mental events cause physical events;
2. Principle of the Nomological Character of Causation: When two events are related as
cause and effect they have descriptions under which they instantiate a strict law;
3. Mental Anomalism and Irreducibility: There are no strict laws of psychology and no
strict psycho-physical laws, by which mental events could be predicted or
explained with physical terms.⁹⁵

The second premise tells us what it is to be a cause. To be a cause is to be causally related to
an effect, of course; but, according to (2), the causal relation is backed by strict laws. So what
it is to be a cause is to be related to an event (the effect) by a strict law. But the only strict
laws available are laws instantiated by physical properties; and these might be available only
in principle, in some idealized future completed physics.⁹⁶ McLaughlin expresses what
anomalous monism entails this way:

4. Type-Epiphemomenalism: (a) Events cause other events in virtue of falling under
physical types, but (b) no event can cause anything in virtue of falling under a
mental type.

⁹⁵ In fact, in order for the argument to token identity to go through (3) must also claim that the
only strict laws available are laws instantiated by physical properties. It is plausible that Davidson
meant to claim this and let’s assume it here.

⁹⁶ Subtleties of this idealization not broached by Kim or McLaughlin.
A large number of intuition pumps sharing the same form are brought to bear by many critics, represented here by McLaughlin and Kim. For one example, a silver and square paperweight dropped on wax will produce an indentation in the wax. But it is the paperweight’s mass and square shape, not its silver color, that seem causally relevant to the impression made (Robb & Heil 2000). What’s pumped is the intuition that causes cause their effects in virtue of some properties and not others. Anomalous monism’s answer is counterintuitive. Mental events intuitively cause behavior in virtue of their mental properties, but anomalous monism goes in for physical properties.

As I said, this is supposed to be an unhappy result. Interestingly, nowhere in McLaughlin (1993) can I find an evaluative expression with respect to this entailment being a liability or a virtue. McLaughlin merely argues that Type-E is entailed. Kim (1993) argues that it is entailed and that this is a liability. Here is Kim:

“On Davidson’s account there is no causal role for mental properties in psychological explanation,” (19).


“… anomalous properties are causally and explanatorily impotent, and it is doubtful that they can have any useful role at all,” (1989/1993: 271).

“… [F]or if something that purports to be a theory of mental causation assigns no causal role to mental properties – if it has nothing to say about the causal powers of mental properties while having plenty to say about those of physical properties – the theory can, it seems to me, reasonably be said to be epiphenomenalistic with regard to mental properties, (1993: 20-21).

“[T]his means that individual events can enter into causal relations only because they possess physical properties that figure in laws. Consider an example: Your desire for a drink of water causes you to turn on the tap. On Davidson’s nomological conception of causation, this requires a law that subsumes the two events, your desiring a drink of water and your turning on the tap. However, psychophysical anomalism says that this law must be a physical law, since there are no laws connecting mental-event kinds with physical-event kinds…. [T]he fact that it is an event of this mental kind… apparently has no bearing on its causation of your turning on the tap…. Mental events play no role in making mental events either
causes or effects... We must therefore conclude that Davidson’s anomalous monism fails to pass the test of mental causation; by failing to account for the causal efficacy and relevance of mental properties, it fails to explain the possibility of mental causation, (2006: 188-189).

The reason Type-E seems devastating is because the salience of our practice of reason explanation commits us very deeply to a causal explanatory role for mental properties. As we’ve said, if they are not causally explanatory then that important practice falls apart. So, I would assert that a hidden part of Kim’s argument against Davidson says that Type-E entails or at least leads to what we have been calling the Catastrophe Thesis. Again:

5. Catastrophe: Mentality plays no role in the causal explanation of intentional action.

In the next section, I begin my criticism of the critics, by first laying out the shared assumptions about causal explanations, the deeper implications of which I later argue the critics fail to understand.

§

4.3 The Canonical Formula For Causal Explanations

When we ask for an explanation it is because we find ourselves in an epistemic deficit. There was an event, \( e \), or, more precisely, a particular aspect of it, \( G \), which now stands in need of explanation. So we ask:

\[ e, G \]

---

97 Before moving on though, I should say that, to my mind, the problem between Davidson and McLaughlin/Kim cannot be satisfactorily solved by noting how regularly Davidson spoke of predicates or descriptions and not properties. See footnote xy above. Davidson himself and others have certainly tried this defense. But it would isolate anomalous monism from the contemporary discussion, which is usually conducted in terms of properties and not predicates. And in any case, McLaughlin and Kim take Davidson to be talking about properties of things, not just linguistic ways of approaching things.
6.) Why is $e$ G? (Why did $e$ become G?)

That is, we might ask:

Why did this building collapse?
Why did this wine glass shatter?
Why did Jill return to the bank?

Let’s call “$e$’s being G” the primary explanandum. In response to questions of this form there comes an explanation, for example:

Hurricane Katrina!
The wine glass shattered because the soprano emitted a sound wave with acoustic properties P.
Because she left her glasses there.

These explanations can take considerably different surface forms and are undoubtedly enthymemetic. For example, uttering “Hurricane Katrina!” is a causal explanation of why the given building collapsed because it offers the event described as Hurricane Katrina as the cause of the event described as the building collapsing. In this debate, the descriptions are reformulated into a gerundial form of “to be,” thus: the event’s being Hurricane Katrina caused the event’s being a building collapse. Indeed, all parties to the debate I am discussing are in agreement with respect to representing the canonical formulation of causal explanations in this manner, like this:

7. Form of Causal Explanation, CE): $e$’s being F caused $e$’s being G

---

98 Even more precisely: Rather than the event or the property, it is actually “$e$ coming to be (or becoming) G” that stands in need of explanation.
where, \( c \) and \( e \) are particular token events, and \( F \) and \( G \) are descriptions, predicates, types or properties of \( c \) and \( e \) respectively.\(^{99}\) Including its token-type structure, this is indeed the formula all of the participants use.\(^{100}\)

Nevertheless, it would behoove us to review why we ought to think it is the correct formula.

\[ \§ \]

### 4.4 Epistemically Adequate Explanatory Generalizations

One of the strongest considerations in favor of CE is the observation that to identify the cause of an effect (or to state that a causal relation holds between the primary explanandum and one’s candidate cause) is not yet to explain that effect. This is because it is possible that there be a difference in explanatory adequacy between two true singular causal claims that refer to the same two events. Compare, for example, the two following singular causal claims:

8a.) “*The hurricane* caused the building to collapse” and
8b.) “*The event reported on page one of the newspaper* caused the building to collapse.”\(^{101}\)

\(^{99}\) It is standard in this debate to speak of particular events which token a variety of different types. This is the Davidsonian notion of events as opposed to the Kimian notion of events — the latter takes the token event to be the instantiation (intuitively an event) of a specific type at a time, \( t \), where *what* instantiates the type is the thing that has that type or property. It would be unfortunate if Kim’s criticism of Davidson turned on their talking past each other because of different notions of “events.” Kim insists this is not the case.

\(^{100}\) Kim himself says, “… it makes sense to ask questions of the form ‘What is it about events \( c \) and \( e \) that makes it the case that \( c \) is a cause of \( e \)’ and be able to answer them, intelligibly and informatively, by saying something like ‘Because \( c \) is an event of kind \( F \) and \( e \) is one of kind \( G \),” (22).
These two sentences are, let’s say, extensionally equivalent (they are referring to the same events) and (let us suppose) both true. But even when they are both true, the second sentence is not explanatory because… well, it is difficult to say. In our example, an event’s being a hurricane does seem explanatorily relevant to another event’s being a building collapsing. This may be because we are familiar with the former type of thing typically causing the latter type of thing. We are not familiar with any general tendency for events reported in the newspaper causing buildings to collapse. But some generalization relates the other types, like this: “Hurricanes make buildings collapse.” Maybe that particular generalization relating those types is not explanatory enough for some purposes. In that case, to be explanatory F and G would have to express a relation between, maybe, wind speeds and yield strength of the building materials. In any case, participants to this debate all agree that in order to be explanatory F and G will instantiate a generalization relating them as types. Not “the hurricane caused the building to collapse” (a singular causal statement) but “winds of 200 mph tend to cause wood to reach its yield point,” a statement connecting all instances of 200 mph winds to wood yielding, i.e., a generalization. What more can be said about generalizations between F and G that are actually explanatory?

I discuss what’s required of the relation between F and G in order for explanation to succeed in more detail below. For now let’s just note the second sentence about the hurricane, (b), lacks whatever it takes to achieve epistemic satisfaction and therefore lacks what

---

101 I have italicized the candidate cause (appearing under different descriptions) in both sentences. This example is of course adapted from Davidson (1963, p. 16) and may seem overused, but it remains misunderstood.
I will call “explanatory adequacy.” The point is that because (8a) is explanatorily adequate and (8b) is not – yet both are true – the truth of the causal statement, we can see, does not depend on how the relata are described. That stands in favor of revising CE into CE*, which better represents the fact the causal relation and causal explanation can be pulled apart in the way just described:

9. Intensional Causal Explanation [CE*]: $c$ caused $e$ and $c$’s being $F$ causally explains $e$’s being $G$.

Now, what does it takes for a relation between $F$ and $G$ to be explanatory? We earlier (Section 2.1.3) examined the literature on scientific explanation which has moved from Hempel’s deductive-nomological account of scientific explanation to widespread dissatisfaction with that account. There is a long list of alternative accounts, including Salmon’s statistical-relevance account, Friedman and Kitcher’s unificationist accounts, David Lewis’s account of causal explanation as providing information about the explanandum’s causal history, transfer of energy accounts, and Woodward’s interventionist account. My

102 Throughout, whenever I say “causal explanation” I intend an epistemic notion. There may be such a thing as metaphysical explanation, but it should be clear that I am talking about epistemological matters when I talk about adequate causal explanations of events here. On the other hand, the metaphysical, extensional, causal relation is not intended to be epistemic but metaphysical. (The epistemic appeal to it, though, while itself not being metaphysical, grounds the causal explanation in factive events.) As usual, just what “metaphysical” means is difficult to say, but here I suppose I am using it to mean something like “beyond our best epistemologies,” which is a “kantian” sense, small “k,” big “ian.”

103 We can see that singular causally explanatory claims exhibit intensionality, such that explanatory adequacy is not preserved upon substitution of co-referential terms. In other words, “The hurricane” and “The event reported on page one of the newspaper” can be descriptions that refer to the same event, but when we substitute one for the other, the status of the explanatory adequacy of the true singular causal claim changes. Antony (1989) misunderstands the significance of the hurricane case. She says “While it could be perfectly obvious that the event reported on page 5 of Tuesday’s Times caused the event reported on page 13 of Wednesday’s Tribune, it would be ludicrous to look for a law relating events of these kinds” (p. 164). She mistakenly takes (8b) to be explanatory, which entirely misses Davidson’s point.
purpose here is not to characterize scientific explanation \textit{in toto} but just to survey the variety of ways a causal \textit{generalization} might be conceived to achieve adequacy.

I imagine a continuum from very minimal requirements to requirements that are robust to the point of being unattainable. On the end of the continuum with the most minimal requirements, our first stab at the explanatory relation between $F$ and $G$, we could call the “causal lore” view. It is just \textit{whatever minimal connection between those two descriptions that remedies the epistemic deficit in regards to the primary explanandum}. Importantly, such a remedy of the contextual epistemic lack may or may not be accompanied by a mere psychological buzz or “feel-good.”\textsuperscript{104}

So on the causal lore view of the relation between $F$ and $G$ that is required for explanatory adequacy, we would say:

10. \textit{Causal Lore}: $c$ caused $e$ and $c$'s being $F$ and $e$'s being $G$ is some lore that has the minimal virtue of remedying the epistemic deficit in regards to the primary explanandum.

So now on a slightly more demanding view of what’s required for $F$ and $G$ we might say:

11. \textit{More or Less Strict}: $c$ caused $e$ and $c$'s being $F$ and $e$'s being $G$ instantiate a more or less strict law or ceteris paribus generalization.

Or, in the specific domain of reason explanation:

12. \textit{Constitutive ideal of rationality}: $c$ caused $e$ and some relevant platitude of folk psychology is invoked.

\textsuperscript{104} Rather than allowing the mere feel-good buzz of explanation to count as adequately explanatory, we should demand more be true than some psychological fact about the hearer. On the other hand, when an explanation is explanatory, we do empirically often get “epistemic satisfaction” (some take it to be a kind of buzz or intuitive feel that explanation has been accomplished). As Keil & Wilson (2000) say, “There is a sense both that a given explanation satisfies a cognitive need, and that a questionable or dubious one does not.”
Now let’s imagine the most demanding view on causal explanatory adequacy.

13. Procrustean: c caused e and c’s being F and e’s being G instantiate a strict law of an ideal physics.

How shall we choose which is the correct account of the required relation between F and G in an explanation to achieve genuine explanatory adequacy? For my part, I think I can mostly stay out of the fight. The participants in the debate between Davidson and Kim do not feel the need. All we need to do is take it as given that some reason explanations somewhere are adequate without requiring they be Procrustean explanations. That is to say, for my purposes, any position from Causal Lore on down will work as long as it is not as demanding as the Procrustean account. The Procrustean view, if correct, would have untoward ramifications for anomalous monism, as I have laid it out. It would amount to a denial of my central claim that identifying a causal relation is not the same as offering an adequate explanation. But I would argue it is a mistake to affirm the Procrustean view of causal explanatory adequacy, and McLaughlin and Kim do not explicitly endorse it in the articles I’m reviewing.¹⁰⁵

I call the most demanding view “Procrustean,” because instantiation of strict laws is a bad fit for epistemic adequacy. This view would force our epistemic notions about causal explanation into a Procrustean bed of metaphysical notions about the relation between causes and what grounds them – often laws – more about which later. It would eviscerate the explanatory adequacy of every special science and even of any non-ideal physics.

¹⁰⁵ Although Kim seems to endorse it elsewhere, namely, in his discussion of “explanatory exclusion,” (1989).
Much less than what is demanded by the Procrustean view will provide epistemic adequacy. Many philosophers, when discussing properties and necessary and sufficient conditions for explanatory adequacy, talk about the nomic subsumption of properties. But they are making a metaphysical point, I believe. As far as the necessary requirement on the relation between F and G for an explanation to be adequate, we are asking about epistemic matters. And with respect to epistemic causal explanation, “Ignorance of competent predictive laws does not inhibit valid causal explanation, or few causal explanations could be made” (Davidson 1963: 16). So unless we want to eviscerate the explanatory adequacy of every discipline of knowledge except fundamental physics, we need to deny the Procrustean view of the second (epistemic) necessary condition on an explanation’s explanatory adequacy.\footnote{As is well known, in order to be explanatory these generalizations cannot be accidental, they must be law-like; but this is a difficult distinction to delineate. The necessity distinct from accident in these generalizations is discussed in Section 4.6 below.}

§

4.5 Two Conditions On Causal Explanations’ Adequacy

The arguments I gave above for adopting CE* can be recast as suggesting two necessary and jointly sufficient conditions on the explanatory adequacy of causal explanations. We say a causal explanation asserts a causal relation between two events; that is natural enough. But that the asserted causal relation hold is not sufficient for the adequacy of the explanation that asserts it. The hurricane example shows this. But we can also see that this is a necessary condition. That the asserted causal relation hold is necessary because it grounds the
explanation insofar as causally explaining one thing by appealing to another thing would not be possible if there were no causal relationship between the two things.

So, it is a necessary but not sufficient condition on an adequate causal explanation that its asserted causal relation be true. (Let’s call this condition the “metaphysical” or “groundedness” condition on adequate causal explanatory statements.) Now when we ask, “What more is needed in a causal explanation, besides a true causal relation, to make the causal explanation adequately explanatory?” a ready answer is that the grounded cause be described in an epistemically informative way relative to the effect-event. As the hurricane example hints, the candidate cause must be described in an informative way relative to the description of the given-event in the primary explanandum. That is, F and G must be related by some epistemically informative generalization connecting their relevant types. This condition of informativeness, we can call the “epistemic” condition on adequate causal explanatory statements. In any case, the epistemic condition is a second necessary condition, which, jointly with the necessary fact of the causal relation, might be considered a sufficient condition for adequate causal explanations.

We should now consider a different example in order to see how the two conditions hang together. First note that the hurricane example is a case where the epistemic condition is not met while the metaphysical or “groundedness” one is. We now need the case where the “groundedness” condition is clearly not met, while the epistemic one arguably is. Reverend Pat Robertson may say that flooding in New Orleans was caused by God’s displeasure at the city’s moral corruption; and that may be explanatorily satisfying (psychologically-speaking) to people for whom there is an explanatory generalization between God’s anger and natural disasters. That is, it may meet the necessary condition that the descriptions of the events advert to an explanatory generalization. But Robertson’s
explanation is false or non-explanatory because it does not meet the necessary condition that
there actually, truly, be an extensional causal relation between the candidate cause and the
event in primary explanandum.

It should now be clear that Davidson holds (and his critics wish to grant him) that
the causal relation is a relation between token particular\textsuperscript{107} events (not between types) and
that various types, descriptions, predicates and properties can be more or less informative
ways of describing those particular token events, so as to achieve epistemic satisfaction and
explanatory adequacy. So \(c\) and \(e\) are related as particulars by the causal relation, and \(F\) and \(G\)
are related by some explanatory generalization.

We may therefore parse Davidson and the critics’ shared representation of causal
explanations this way:

14.) \(c\) and \(e\) are related by the causal relation; and \(c\)'s being \(F\) causally explains \(e\)'s being
\(G\), because \(F\) and \(G\) are types related by an explanatory generalization.

\[\§\]

\textbf{4.6 What Metaphysically Undergirds The Causal Relation?}

According to the previous section, referring to the fact of \(c\)'s being \(F\) is causally explanatory
with respect to \(e\)'s being \(G\) because of, or indeed, \textit{in virtue of} \(F\) and \(G\) being related by some
explanatory generalization. With that \textit{in virtue of} question on our lips, we may ask further what
in virtue of which are \(c\) and \(e\) related by the causal relation. That is we may ask, in virtue of
what are \(c\) and \(e\) related by the relation of causation? (I put it carefully so as to preclude an

\textsuperscript{107} The critics also allow Davidson his point that causation is a two-place relation not a multi-
place relation with some place for relevancy and the like. It is not “cause as” as Davidson insisted
it was not. See (McLaughlin 1993).
epistemic understanding.) They, $c$ and $e$, will be so related in virtue of some of their properties instantiating a relation that backs or grounds the causal relation, to put it tautologically. But the question is what is the relation that backs or grounds the causal relation?

In the context of this debate, what backs the causal relation is taken to be some necessary relation. This is what most readers take Davidson to have meant by a “strict law” of an idealized physics. And whether one prefers a nomic subsumption view like that endorsed by Kim and McLaughlin in Davidson’s stead, or a counterfactual account without nomic necessities, there is still an idea of necessity involved in our idea of causation. As Kim writes in his textbook,

“A principle connecting laws and causation that is widely, if not universally, accepted, is this: *Causally connected events must instantiate, or be subsumed under, a law*. If heating a metallic rod causes its length to increase, there must be a general law connecting events of the first type and events of the second type,” (2006: 188, italics in the original).

It might be enough for the epistemological case of explaining some particular fact that the law be only ceteris paribus. However, as far as what undergirds the causal relation as a metaphysical matter, it should be clear we need strict laws which express relations of necessity between the cause and effect.\textsuperscript{109}

Now, even for those who do not take the nomological approach, necessity remains relevant. The counterfactual approach, as Kim notes, seems not to require laws, but does not

---

\textsuperscript{108} I speak this way now, although in Chapter 5 I consider that the “in virtue of” order of explanation might go the other way around.

\textsuperscript{109} For a discussion of ceteris paribus laws see Lange (2002) and Fodor (1991), which argues that Davidson’s strict laws are a special case of ceteris paribus laws, namely when the requirement that “all things be equal” is met “vacuously.”
deny the kind of necessity under discussion presently. According to the counterfactual approach, to say that event $c$ caused event $e$ is to say that if $c$ had not occurred, $e$ would not have occurred.\textsuperscript{110} As Kim says: “The basic idea is that a cause is the \textit{sine qua non} condition, or necessary condition, of its effect…” (189).

This idea has intuitive plausibility. “The overturned space heater caused the house fire. Why do we say that? Because if the space heater had not overturned, the fire would not have occurred,” Kim says.\textsuperscript{111} But there’s more. The necessity appearing in both the counterfactual and nomological approach is needed to block mere correlations. To see this, consider another example.

Imagine every time I light a fire in my fireplace an owl hoots outside. It happens a whole lot, so I am beginning to want to say that I am causing the owl to hoot by lighting my fire. But I am not allowed to conclude there is a causal relation between my lighting the fire and the hooting, or the explanation for owl hooting won’t be believed at any rate. It remains a coincidence, a mere correlation between the hooting and my lighting the fire.

If I look into the chimney and see an owl in there I may want to go ahead and assert more than a mere correlation. If I do go ahead and claim that my fire caused the owl to hoot, basically what I am saying is that this is more than an accident, more than constant conjunction, more than correlation. It is the kind of thing that can be predicted and intervened upon. When I make a causal statement, in fact, I offer a prediction that certain counterfactuals are the case. That is, I say, "\textit{were} I to light a fire now the owl would hoot." So to assert a causal claim is to assert what would necessarily be the case in another possible world.

\textsuperscript{110} It is also a requirement that there be a chain of “counterfactual dependencies” connecting the cause to the effect, and the necessity must be mitigated for cases of pre-emption. But the special considerations are not relevant to our discussion. See Lewis (1973).

\textsuperscript{111} There is, of course, Mackie’s INUS condition theory.
world. As Scruton, from whom I adapted the owl example, says "every time we assert the existence of a causal connection our thought reaches beyond the actual to embrace the possible as well" (77). When we posit a cause we also posit and commit ourselves to there being a necessary relation between the posited cause and the given effect.

The properties in virtue of which \( c \) and \( e \) are causally related are, therefore, the properties that relate \( c \) and \( e \) necessarily.\(^{112}\) We can add this to our schema thus:

15. Form of Causal Explanation, Disambiguated or CE**): \( c \) and \( e \) are related by the causal relation in virtue of \( \epsilon \)’s being \( H \) and \( \epsilon \)’s being \( J \) where \( H \) and \( J \) instantiate a necessary relation; and \( \epsilon \)’s being \( F \) causally explains \( \epsilon \)’s being \( G \), in virtue of \( F \) and \( G \) being related by an explanatory generalization.

§

4.7 Two Distinct \textit{In Virtue Of} Questions

We can now see that there are at least two different ways properties are involved with various causal explanatory talk. There is the role properties play in describing the particular token events (the cause and the effect) such that reference to the cause under that description explains the effect under the description given in the primary explanandum; and there is the role of the properties of \( c \) and \( e \) in virtue of which \( c \) and \( e \) are related by the metaphysical causal relation. \( H \) and \( J \) are whatever properties in virtue of which two causally related events are in fact causally related – whatever properties ground the causal relation, whatever properties instantiate the relation of necessity we’ve said backs causation.

\(^{112}\) One last thought, the relation of necessity between \( c \) and \( e \) can itself be contingent. It is \textit{nomologically} necessary, according the laws of nature in the actual world, that if \( c \), then \( e \). But in other worlds \( c \) and \( e \) may not be related by necessity. This is sometimes called contingent necessity.
Behind the fact that talk of properties has two roles to play is the fact that there are two *in virtue of* questions to be asked of any causal explanation, one epistemic and one metaphysical. About any normal causal explanation we can pose an epistemic question in which, basically, we ask for a more informative description of the candidate cause relative to the effect given in the primary explanandum. But we can also pose a metaphysical (and non-epistemic) question about what undergirds or grounds the causal relation. Importantly, these two *in virtue of* questions are two distinct questions. Answers to these two *in virtue of* questions are as logically independent as the hurricane example shows epistemic causal explanations and metaphysical causal relations are.\textsuperscript{113}

A friend of the critics might object by denying that the two questions are numerically distinct. One could say that explanation is not achieved until the necessity is demonstrated, but that would amount to saying that F is identical to H and G to J, which would involve a commitment to what we have called Procrustean notion of epistemic causal explanation.

We are now finally in a position to see what is wrong with the critics’ program.

§

\textbf{4.8 My Criticism Of The Critics’ Objection}

It follows from the original argument for anomalous monism that there are two *in virtue of* questions with respect to causal explanations; but this has been not been recognized by the

\textsuperscript{113} As above I sometimes refer to the metaphysical *in virtue of* question as simply the “second” *in virtue of* question and the epistemic as the “first” because an inquiry does not become metaphysical just by calling it metaphysical. We have seen this before: the \textit{Catastrophe} thesis is not catastrophic in virtue of its name and a relation called the “necessitation” relation does not thereby necessitate anything. Many of my interlocutors conflate what I call metaphysics and epistemology; I wonder if the word metaphysics continues to remind them of what I mean by it (namely what makes a cause be the cause it is., exclusive of our epistemic intuitions.)
critics. When asking an *in virtue of* question in natural language, there can be ambiguity between whether you mean a possible metaphysical *in virtue of* question or a possible epistemic *in virtue of* question. The properties that would answer the *in virtue of* questions come in two flavors, as do the questions. The epistemic *in virtue of* question does not pose a threat to anomalous monism. Why is the epistemic *in virtue of* question not a threat to anomalous monism? In this framework, a merely epistemic question will be either the first request for an explanation of the explanandum or it will be a question put to a purported explanation that does not achieve adequacy. Provided we do not accept the Procrustean requirement on explanatory adequacy, anomalous monism is not threatened by persistent epistemic *in virtue of* questions.

We are interested in the metaphysical *in virtue of* question. However, the key question to which the critics accuse Davidson of offering a counterintuitive answer is actually ambiguous. Here’s a disambiguation:

16. *Type-E, Metaphysical Disambiguation*: Events are causes of other events in virtue of falling under physical types, but no event is a cause of another event in virtue of falling under mental types.

To say, as *Type-E, Metaphysical Disambiguation* does, that “events are causes… in virtue of… Q” is to say that a cause and its effect are causally related in virtue of… Q, in this case, physical types. But *Type-E, Metaphysical Disambiguation* does not entail *Catastrophe*. Here, in one long sentence, is why. That an event token (i.e., a cause) is related by the causal relation to

---

114 To the charge of equivocating over an ambiguity, the critics will instantly insist that they are indeed asking a metaphysical question. However, unless more is said, all they have asserted is that they label their ambiguous *in virtue of* question with the same word as I label the *in virtue of* question about what grounds causation (not what is explanatory). They would still fail to prevent the equivocation I am talking about unless they can show the two questions are not numerical distinct, instead of merely showing that these two distinct questions are sometimes called by the same name.
another event token (i.e., its effect) in virtue of the physical properties of both does not at all conflict with the positing of the cause event under a certain (mental) description serving to explain the effect under the description in which it appeared in the primary explanandum. This is demonstrated via the hurricane example. Mental properties absolutely play an important role in causal explanation. (This was the phenomenon we meant to capture in the first place.) Are mental properties the properties in virtue of which $c$ and $e$ are metaphysically related by the causal relation? No. But to want that that be the case is a mistake. We do not need mental properties to undergird the metaphysical causal relation. In fact, we do not want that. We want physical properties to undergird the metaphysical causal relation, if we are to be physicalists.

We want mental events to be causally explanatory. That would thankfully deny *Catastrophe*. But we do not need those events to have the causality they have *in virtue of* their mentality. They need only be actually causally related (and referred to in an appropriate way in an explanation).\(^{115}\) In an important sense, then, the power of the *in virtue of* criticism relies on an equivocation. It asks what in virtue of which are $c$ and $e$ related, requesting the metaphysical answer. When it hears that answer (“in virtue of something physical”) it charges that it is incorrect epistemically—speaking because in intentional explanations of action we say that reasons cause actions in virtue of being reasons, or in virtue of their mentality. But that is an equivocation. We know it is an equivocation if only because we know that to detail what in virtue of which a candidate cause-event is causally related to the given effect-event is not sufficient to provide an adequate causal explanation of $e$’s being G.

\(^{115}\) Anomalous monism is given by its critics that the purported token cause event is indeed a cause via the original argument from anomalism to token identity. “The very mental event is token-identical with a physical event, and thus causally efficacious insofar as the physical event with which it is identical is causally efficacious,” (Yoo 2007).
Another way to understand what goes wrong with the charge of type-epiphenomenalism would be as follows. The critics think they have token identity, type nonidentity views on the ropes if those theories cannot seem to answer the in virtue of question intuitively. But in fact, the critics do not see an ambiguity in what their question asks. The answer from views like Davidson’s to the in virtue of question exhibits the suppressed equivocation. The question, again, is: c causes e but in virtue of what is c a cause of e? To this I say, explanations that appeal to c’s being F to causally explain e’s being G are adequate in virtue of something about the properties of c and e, namely an epistemically informative relation between F and G; but that c can be a cause of e is another question, a question of a metaphysical nature. An answer to how c can be a cause of e is not going to suggest epiphenomenalism one way or the other about causally explanatory statements… about the causally explanatory statements that describe the causal relation between c and e in an informative way, such that it is clear how c’s being F causally explains e’s being G.

In the end, many of the critics’ claims are unclear. I quoted Kim at the beginning as saying, “Mental events play no role in making mental events either causes or effects” (2006: 188-189). If he means something more epistemic, which he denies by the way, then he’s just wrong. Mental properties do play a role in showing how George’s going to the fridge is caused by his desire for beer. But if Kim means by “making mental events causes of effects,” – what I’ve called “grounding” the causal relation between causes and effects – then his mistake is to think that that is a challenge to the status of reason explanations. What makes c and e related by the causal relation, in this debate, (namely, a law of nature) is much different than why appeal to c under a description F causally explains e’s being G.

For one more nail in the coffin of this criticism, consider that Kim has says:
All that is necessary to see the problem for Davidson is the recognition that it makes sense to ask questions of the form “What is it about events \( c \) and \( e \) that makes it the case that \( c \) is a cause of \( e \)” and be able to answer them, intelligibility and informatively, by saying something like “Because \( c \) is an event of kind \( F \) and \( e \) is one of a kind \( G \) (and, you may add if you favor a nomic conception of causality, there is a law of an appropriate form connecting \( F \)-events with \( G \)-events)”. This is only to acknowledge that the causal relation obtains between a pair of events because they are events of a certain kinds, or have certain properties (22).

However, as I’ve shown, the properties under which \( c \) and \( e \) are nomically subsumed (i.e., those properties in virtue of which \( c \) and \( e \) are related by the metaphysical causal relation) are not the properties we refer to in epistemic causal explanation. So thinking it is sensible to ask Kim’s question does not entail that *Catastrophe* follows for anomalous monism.\(^{116}\)

§

4.9 The Shattering Glass Objection

One might object to what I’ve put forth here with the same intuition pumps used at the beginning, during the early criticism of anomalous monism. Objectors to my view could say that it is a liability to say that the weight of the rock doesn’t *cause* but only *causally explains* the window’s breaking. We say that properties are causal not causally explanatory, the objector might offer. Critics have offered various intuition pumps. To repeat our earlier example, a silver, square paperweight dropped on wax will produce an indentation in the wax. But it is the paperweight’s mass and square shape, not its silver color, that seem causally relevant to

\(^{116}\) If he means the metaphysical question (i.e., if he means to ask for those properties in virtue of which \( c \) is related by a relation of necessity to \( e \)), then it’s not a liability. If he means it epistemically, then it is like the case we discussed above where a purported explanation somehow fails to achieve explanatory adequacy and the person in an epistemic deficit about the primary explanandum asks for another explanation or a deeper explanation; they ask for a generalization between some \( F \) and \( G \) that is satisfying.
the impression made (Robb and Heil 2000). More famously, an opera singer may shatter a
glass by singing the word “Shatter!” Yet, when we explain the event we say that the word
caused the glass to shatter in virtue not of its semantic properties but in virtue of, say, its
acoustic properties (Dretske 1989, Braun 1995). Intuitively, it is a French pear’s weight and
not its color, nor its “Frenchness,” that makes the scale depress and present its reading
(Honderich 1982).

But the fact that intuition pumps of this form were marshaled against Davidson’s
anomalous monism is one of the ironies of a philosophical criticism working with dense and
difficult texts. The intuition pump shares the same form as the one Davidson offered to
illustrate his account, namely the “hurricane and page one of The Times” case. These pumps,
clearly, are meant to show that some properties (or types) are relevant and some are not; and
that, for a given explanandum, we have an intuitive sense of which is which.

Those who offer these pumps as considerations against anomalous monism fail to
notice that the cases they are describing remain, in Davidson’s terms, epistemic causal
explanations. The properties appearing in those cases of explanations are not the properties
in virtue of which the cause and the effect are necessarily related. What the critics mean by
“causal relevance” Davidson means by “explanatory relevance” or what I’ve called
“epistemic informativeness.” It’s true that a causal explanation requires that the cause actually
be related to the effect by the metaphysical causal relation, but causal explanation does not
require that we know the properties in virtue of which they are so related (necessity in many
accounts), only that it is a fact that they are.
4.10 LePore and Loewer on Two Types of Causal Relevance

LePore and Loewer (1987) draw a distinction between what they call “causal relevance\textsubscript{1}” and “causal relevance\textsubscript{2}.” Like what I’ve said above, their distinction allows them to show how anomalous monism is able to dodge the qua problem. On their account as well, it is the critics who equivocate between the two possible meanings of causal talk and thereby violate an important distinction. “The confusion is between two ways in which properties of an event c may be said to be causally relevant or irrelevant,” to an event e. Consider the following locutions:

(a) Properties W and X are causally relevant\textsubscript{1} to making it the case that c causes e, and
(b) c’s possessing property Y is causally relevant\textsubscript{2} to e’s possessing property Z.

The distinction that Lepore and Loewer highlight seems rather similar to the distinction that I have highlighted. But there are differences. Causal relevance\textsubscript{1} is the way properties can be relevant to two particular events being related by the causal relation. LePore and Loewer, furthermore, hold that W and X need to be connected by a strict law to actually be causally relevant\textsubscript{1}. Causal relevance\textsubscript{2} describes what I would call explanatory relevance. In our everyday causal explanations, you want to know what property of the cause is causally relevant\textsubscript{2} to the given property of the effect in the explanandum. When you hear the explanation “Because c is Y,” you understand why e became Z. But you may understand such without direct appeal to a strict law. “The heart of our response to the claim that anomalous monism is committed to epiphenomenalism is this: anomalous monism entails that mental features are causally irrelevant\textsubscript{1}, but does not entail that they are causally irrelevant\textsubscript{2},” (LePore and Loewer 1989, adjusted slightly).

Although they are different distinctions, the one I make and that of LePore and Loewer seem to share the same form. However, LePore and Loewer do not make the exact
The distinction between metaphysics and epistemology that I make here. They do not understand the claim that “there is” a strict law as a metaphysical claim and not epistemological. Further, they do not offer much of an answer to just what it is that makes it the case that c is a cause of e when c is in fact a cause of e. Or, if they do, they say “strict law” without any elaboration. I consider possible elaborations below.

§

4.11 Davidson on “Cause” and “Explain”

Davidson strikes a distinction between “causal explanation” and “causation.” When people use the words “causes,” “cause,” or “caused,” and even “cause in virtue of” they most of the time mean causal explanation, Davidson (1967) says. Causal explanation is intensional, mind-dependent, pragmatic, contextual and so on. The causal relation, Davidson says, is extensional, mind-independent, non-pragmatic, noncontextual and so on. More specifically, Davidson held that events simpliciter, just the bearers of properties, are the relata of the causal relation, not properties. Properties, for their part have their role in causal explanations. The causal relation relates events. Does Davidson address the further question of that in virtue of which a cause is the cause it is?

First, just what is the nature of that question? It is supposed to be metaphysical and not merely epistemological. Recall after making a distinction that dodges the qua problem, then the question to ask is: In virtue of what is a cause a cause? Or, in other words, why does a cause have the causality it has? We may begin by offering one place a cause does not derive its causality from: a cause does not get its causality in virtue of appearing in a causal explanation. It must be a cause already, so to speak, in order for appeal to it in a causal

155
explanation to do any work. As we worked out above, if we have in hand a causal explanation that is good, genuine, adequate or true (that refers to mental properties) then we can know that there is a strict law relating cause and effect under another description. But this is not because of appearing in the causal explanation. The cause is metaphysically a cause already, perhaps in virtue of falling under a law or perhaps in virtue of nothing – just brutally.

Whereas Davidson is widely thought to hold that a cause is the cause it is in virtue of falling under a strict law with its effect, most of the time Davidson’s published words actually remain neutral on the Euthyphronic question of what’s in virtue of what.

There may be a lot of possible answers to what makes a cause a cause. But importantly – and this has not been noticed in the literature – the answer to this question can be separated from the fact that when c causes e we have a law. That is, it could be that a cause is the cause it is, for instance, as a matter of brute fact. And then at the same time it could be that when c causes e there is a law connecting them. How so? Because it could be the case that there is such a connecting law in virtue of the brute causal patterns. This is at least a position in logical space and is consistent with the rest of anomalous monism even though Davidson did not commit himself one way or the other. (I continue to discuss brute causal relations in Chapter Five, next.)

§

4.12 Conclusion

\[^{117}\text{To relate them under the same or explanatory description would be Procrustean (see Section 4.4 above).}\]
I have been at pains to show that the precise way that the critics of anomalous monism formulate their worry about its entailment of type-epiphenomenalism conflates the properties in virtue of which \( c \) is related to \( e \) by the relation of causation with the properties of \( c \) and \( e \) in virtue of which we can achieve epistemic satisfaction and explanatory adequacy in our interest-relative explanations of the primary explanandum (\( e \) under a certain description).

Where the critics go wrong is to think that if mental properties do not feature in the relation that undergirds the metaphysical causal relation between \( c \) and \( e \), then mental properties have no role in causal explanation. That inference relies on an assumption that the properties that satisfy us epistemically are or should be identical to the properties grounding the causal relation, i.e., the properties in virtue of which \( c \) and \( e \) are related by a necessary relation.

In the next chapter we consider a new way to answer the newly disambiguated metaphysical “in virtue of” question. As promised, I momentarily take the nomological view of causation. The laws that back the causes are the laws of nature. And the laws are not mere de facto regularities but are necessary. To achieve the strictness or necessity we need an account of laws. Whence the necessity of generalizations that are actual laws as opposed to the generalizations that are mere regularities?

The overarching structure of Chapter Five is a bit complicated. I introduce a debate among historians as to what Hume really said. I don’t do this in order to assert that philosophers calling themselves Humean do not really understand Hume. No. In the course of presenting the various positions on Hume we discover that there is a position not represented in the laws of nature discussion. I end by articulating and defending that position tentatively and just as a first pass.
CHAPTER FIVE

A CODA: NEW HUME AND THE QUASI-HUMEAN REVERSE VIEW

5.1 The New Hume Debate

5.1.1 Introduction

Against the once widely held positivist interpretation, Hume scholars now agree that Hume did not hold a simple regularity theory of causation devoid of any aspect of necessity. Yet, the contemporary literature on causal laws is split between so-called Humeans and Non-Humeans, where one is a Humean if one denies necessity in causal laws. I do not remark on this to suggest a simple change in nomenclature. Rather, the insights we gain from examining the New Hume Debate will help us in determining relative plausibility in the battle between the two Davids: the Humean David Lewis and the Non-Humean David Armstrong. In the end, Blackburn’s quasi-realist, projectivist Hume offers yet another argument against Non-Humeans about the necessity of causal laws.

5.1.2 The New Hume Debate

If we let “Causation” with a capital “C” and “thick connexion” refer to one event producing another, or necessitating another, but involving something more than mere regular succession, then the old positivist interpretation of Hume holds that there is no such thing as

---

118 Non-Humeans about the laws of nature argue that non-nomic facts, e.g., causal regularities, depend or supervene on nomic facts, e.g., laws. While Humeans argue that the laws depend on the de facto regularities.
Causation or thick connexions in the world. The old positivist interpretation holds that Hume takes causation to be nothing more than constant conjunction or “regularities,” lacking any element of necessity. Hume claims, on this account, that we have no impression of necessity, so any talk of necessary connection is either unintelligible or meaningless. However, this positivist reading has fallen out of favor.

In recent decades there has been a considerable shift in the interpretation of Hume’s philosophical theories about causation. Yet the positivist influence remains, at least in nomenclature. Regularity theories in the contemporary debate about causation have been called Humean. But “this usage just betokens a limited acquaintance with the work of Hume,” Craig writes (2007: 113). While it persists in the nomenclature of the contemporary debate about causation, the positivist interpretation has been thoroughly discredited among historians of philosophy:

Off the agenda now is the idea that he [Hume] taught a strict regularity theory: that there is nothing in reality but regular sequence, and that that is accordingly all that causality amount to, either in our concept of it or in things and events themselves (Craig 2007: 113).

Strawson (1987) charges that the positivist theory of causality attributes to Hume an unhappy inference from epistemology to ontology, from what we happen to know to what exists in reality. The positivist reading, Strawson argues, fails to countenance the legitimate distinction between objects in the world and our ideas about objects.

---

119 The positivist interpretation makes reference to Hume’s “philosophical” definition of “an object followed by another, and where all the objects similar to the first are followed by objects similar to the second,” (E76).

120 In this section, my tone may appear more positive than it actually is towards the New Hume skeptical realist view. I disagree with the New Humeans but I also disagree with the old positivist reading.
Against the positivist reading of Hume, Strawson and others offer what they call a “skeptical realist” reading.\textsuperscript{121} “The skeptical realist interpretations claim that Hume believed we can know that causal powers and objects exist in the world, although we may not be able to know any more about them than that they exist,” (Richman 2007: 1). So the New Hume is a realist about the ontological question but a skeptic with respect to the epistemological question. Hume “takes it for granted… that Causation does exist in reality, although we are entirely ignorant of its ultimate nature,” (Strawson 1989: 219).\textsuperscript{122} So the realism comes from taking it that Hume holds that “causal power is some feature or property of the world, one which does not in any way owe its existence to human existence,” (Craig 2007: 113). The skepticism, then, comes from their claim that Hume consistently maintained the point of view that the real causal powers and forces in nature “are not directly accessible to our senses,” (Wright 1983: 129). As Strawson summarizes, “In the end Hume’s regularity theory of causation is only a theory about causation so far as we can know about it in the objects, not about causation as it is in the objects,” (Strawson 2007: 33). So there exist thick connexions or genuine Causation out in the world among objects but it is beyond our ken.

A good way to get clear on the differences among the possible interpretative positions is to chart where each falls on a classic interpretative problem for Hume studies. It is a problem for any interpreter of Hume to reconcile these three apparent commitments of his:

(a) Endorses beliefs in objects and causes;
(b) Holds that we should not endorse beliefs that do not have appropriate grounding in our impressions (as described in his theory of ideas);

\textsuperscript{121} I am eliding the differences between Wright, Strawson, Popkin and others who’ve written on skeptical realism.

\textsuperscript{122} It is because he is “following Newton, [that] he [Hume] repeatedly insists on the epistemological claim that we know nothing of the ultimate nature of Causation,” (Strawson 2007: 35).
(c) Holds that the beliefs in objects and causes do not have appropriate grounding in our impressions.

The positivist reading achieves consistency by denying or qualifying (a). On the other hand, Strawson accepts (a), and qualifies (b) and (c). That is, Strawson expands what is allowed to be an appropriate grounding by highlighting a supposed distinction of Hume's between what we can “conceive” and what we can “suppose.”

The above formulation is due to Richman (3); Blackburn offers his own inconsistent Humean triad:

1) We have an idea of a thick necessary connexion between distinct events;
2) We have no ideas except those that are preceded by suitably related impressions;
3) There are no impressions that are suitably related to the idea of a thick necessary connexion between distinct events.

The positivist interpretation “takes Hume to be claiming that when we talk of causation we only mean something that strips out the thick element of necessity, and substitutes regular contiguous succession,” (Blackburn 2007: 101). So positivists deny (1). From this perspective, Strawson can be taken to be trying to distinguish between the terminologies of (1) and (2) in order to resolve the inconsistency. Briefly, he says that while we may have no impression of Causation or thick connexions and thus cannot conceive of them, we can nevertheless “suppose” them to be the case and thereby have a “relative idea” of them.

5.1.3 Blackburn’s Hume

Blackburn (1990/2007) offers an interpretation of Hume that is distinct from both the positivist and skeptical realist views. There is a further distinction, Blackburn says, that both the positivist and skeptical realist interpretations fail to countenance. “When we think of a causally connected pair of events, such as the impact of the first billiard ball causing the
motion of the second, we want there to be a further fact than (mere) regular succession,” (103). There is supposed to be more than regular succession. There is supposed to be dependency. We want a fact making it such that when the first happens the second must happen. Blackburn calls this the desire for a “causal nexus.” When Blackburn says we “want” or that “there is supposed to be” a further fact, I take him to be saying that our idea of causation includes the component of thick connexion.

“But now suppose we shift our gaze to the whole ongoing course of nature,” (103) by which Blackburn means for us to imagine not any old pair of causally related events but lawful regularities or patterns of events of said type. We want there to be something in virtue of which these patterns continue. But according to Hume (E37, T90-1),

There is no contradiction in supposing that the powers and forces with which events are endowed at one time cease at another, nor in supposing that any secret nature of bodies upon which those powers and forces depend itself changes, bringing their change in its wake,” (Blackburn 2007: 103).

To imagine that the natural order might at any moment fall apart is to suffer “inductive vertigo,” as Blackburn calls it. We desire a further fact beyond the patterns or regularities in order to avoid this inductive vertigo. What we want,

...is whatever it is that ensures the continuation of the natural order, that dispels the inductive vertigo that arises when we think how natural it might be, how probable even, that the constrained and delicate pattern of events might fall apart. Call the desire for this further fact the desire for a straightjacket on the possible course of nature: something whose existence at one time guarantees constancies at any later time,” (103).

“Nexuses by themselves do not provide a straightjacket,” Blackburn says. A fact alleviating inductive vertigo would be “a very peculiar fact” because: “It has to be something whose own continued efficacy through time is subject to no possibility of change or chance of
failure,” (103). For otherwise, “the fact that it keeps on as it does would itself be a case of coincidence or fluke, another contingency crying out for explanation and engendering inductive vertigo,” (103). Where are we to find such a straitjacketing thing? “Some philosophers think they can point us toward a fact with this potency…Armstrong believes that a kind of necessary, timeless gridlock of universals will do” (103).123

Strawson cites “fundamental forces” as essentially constitutive of “the nature of matter,” and invokes these, as Blackburn writes, “to soothe away inductive vertigo.” However, Strawson is failing to see Hume’s point, Blackburn says. “Even if forces are taken ‘to latch on to real, mind independent, observable-regularity-transcendent facts about reality’ (Strawson 1987: 91) they need something further in order to serve as a straightjacket. They need necessary immunity to change; they need to be things for which the inductive vertigo does not arise,” (Blackburn: 105). One may object that Armstrong and Strawson each individually have offered a straightjacket. But Hume’s point is that at every moment one is vulnerable to inductive vertigo. It is simply part of the nature of human understanding that nothing can be straightjacketed. When one is offered, then it too will need a straightjacket and we get an infinite regress.

“Hume’s main interest in causation is to destroy the idea that we could have such [ultimate] knowledge, and hence ever apprehend a straightjacketing fact: we have no conception of it, nor any conception of how we might approach such a conception,” (105).

“In particular, we must not think of the advance of science as targeted on finding such a thing. The lesson drawn from Newton is that just as Principia gives us the operation of

123 To repeat it here, Armstrong’s attempt at fighting off inductive vertigo is as follow’s. He argues that there is a relation of universal necessity between types or universals. “Suppose it to be a law that Fs are Gs. F-ness and G-ness are taken to be universals. A certain relation, a relation of non-logical or contingent necessitation, holds between F-ness and G-ness. This state of affairs may be symbolized as ‘N(F,G)” (Armstrong 1983: 85).
gravitational force, but does not ‘tell us what it is’, so any conceivable advance in science can only do more of the same,” (105). It can put events in wider and more interesting and exception free patterns, and that is all.” (105).

§

5.2 The Reverse View

5.2.1 Introduction

Someone wise once said that all you have to do to write a philosophy paper is find a claim a writer says is so obviously true that it needs no argument and, then, deny that claim. In arguing that anomalous monism entails epiphenomenalism about mental properties, McLaughlin (1993) makes one such “obviously true” claim.

McLaughlin says it is typically the case that “when a particular bears an extensional relation to another particular, the particulars are so related in virtue of something about each,” (32, my italics). The extensional relation he is talking about, we know, is causation. The hard question he puts to Davidson is: “What are those properties of an event, c, and an event, e, in virtue of which they are related by the causal relation?”

One possible response – not the one he ultimately imputes to Davidson – is considered but is immediately dismissed by McLaughlin. It is this: The two events, c and e, might just be causally related as a matter of brute fact. They might be what they are without being what they are in virtue of something else being the way it is. However, if causal relations were brute, McLaughlin says, “there would be nothing in virtue of which c and e are causally related, no properties and no law in virtue of which c is a cause of e,” (32). McLaughlin says this is “implausible,” (32) and dismisses it out of hand as perhaps absurd.
This supposedly “implausible” or perhaps “absurd” claim is the claim, suitably understood, that I will endorse.

In the following sections, I will outline a view that:

• makes saying “causal relations are brute” no longer engender the worry that McLaughlin is implying when he says that were causal relations brute “there would be nothing in virtue of which c and e are causally related, no properties and no law in virtue of which c is a cause of e;”
• makes the mental causation problem as usually formulated impossible to come up, a non-starter;
• is broadly-speaking empiricist.

The “Reverse View” will reverse the traditional order of priority between causes and laws.

Here’s the structure of my arguments. I pose a Euthyphro question. It works as an argument against the absurdity (inconceivability) of the Reverse View. I then argue that the Reverse View does not have the problems that McLaughlin thought made a view like it “implausible.” The Reverse View also provides a response for the non-reductive physicalist to the “qua problem” of mental causation (a.k.a., type-epiphenomenalism) that McLaughlin and many others leveled at anomalous monism. Finally, I make two objections to the main competitor to the Reverse View, Armstrongism.

5.2.2 The Second “In Virtue of” Question and Laws

In the previous chapters, we arrived at the question of what makes a cause be the cause it is. What we called the second or metaphysical “in virtue of” question asks: When an event, c, is a cause of event, e, in virtue of what are they causally connected?

As we’ve discussed, a standard answer begins with a discussion of the properties in virtue of which a cause and its effect are related. Davidson writes:

“If a caused b, then some descriptions of a and b instantiate a strict causal law,” (p. 243).
Here is Robb’s reconstruction of Kim’s interpretation of the PNCC:

“The principle of the Nomological Character of Causation appears to require that, when one event causes another, it does so in virtue solely of its physical properties,” (Robb 2003, my italics).

Here is another example from Davidson:

“…when events are related as cause and effect, then there exists a closed and deterministic system of laws into which these events, when appropriately described, fit.”

And here is another example of the imputation of a direction of what’s in virtue of what.

“Davidson’s principle of the nomological character of causality implies that… events can cause other events only in virtue of falling under physical types cited in strict laws…” (McLaughlin 1993, italics added).

Davidson’s language in regard to the PNCC is neutral as to what is in virtue of what, laws or causes. No one has noticed that the “in virtue of” locution is slipped in by his critics.

Davidson’s original expression was in fact neutral on the direction or order of metaphysical dependency. Davidson’s PNCC says that when there is causation then you will find that

1) c and e are related by the causal relation and
2) properties of c and e instantiate a strict law.

Let’s let CR(c,e) be read as “c and e are related by the causal relation” and let N(H,J) be read as “properties of c and e, H and J, instantiate a strict law.” The PNCC says CR(c,e) and N(H,J). However, the question remains, is CR(c,e) in virtue that N(H,J) or is N(H,J) in virtue that CR(c,e)? This a question of Euthyphronic form. Are c and e related by the causal relation
in virtue of H and J instantiating a strict law, or do H and J instantiate a strict law in virtue of c’s being causally related to e?

In virtue of its silence on the matter, Davidson’s PNCC gives rise to the question. Nomic subsumption is a way of answering it such that \( CR(c,e) \) in virtue of \( N(H,J) \); but, I suggest, a view is conceivable that answers it such that \( N(H,J) \) in virtue of \( CR(c,e) \). That view would be the Reverse View.

5.2.3 No Catastrophe

As above, the Reverse View is not simply absurd. We can conceive of such a position by imagining a direction of what’s in virtue of what going the other way than usually attributed to Davidson’s PNCC. We argued in the previous chapter that the usual reading of the PNCC did not, as commonly assumed, entail Catastrophe. Now I wish to remark that the PNCC read in the other direction does not entail Catastrophe either. What prevented the entailment before also prevents it here. And that was namely the distinction between epistemology and metaphysics. When we ask “in virtue of” questions we need to be careful to make explicit whether we are asking an epistemological or metaphysical question.

If the causal relation between \( c \) and \( e \) does not stand in virtue of any of their properties instantiating a causal law, then it does not make sense to ask what in virtue of which said causal relation stands. If it is primitive then it stands in virtue of nothing. If it’s primitive other things stand as they do in virtue of it.

5.2.4 Wider Ramifications

The view that I am denying here is central and often tacit in the mental causation debate. The directionality of the PNCC is decidedly from laws down to causal relations. However, I think
it is plausible to disagree with this central idea, which Bennett calls “accepted” and Kim says “makes sense.”

Kim (1993): All that is necessary to see the problem for Davidson is the recognition that it makes sense to ask questions of the form “What is it about events $c$ and $e$ that makes it the case that $c$ is a cause of $e$?” and be able to answer them, intelligibility and informatively, by saying something like “Because $c$ is an event of kind $F$ and $e$ is one of a kind $G$ (and, you may add if you favor a nomic conception of causality, there is a law of an appropriate form connecting $F$-events with $G$-events)”. This is only to acknowledge that the causal relation obtains between a pair of events because they are events of a certain kinds, or have certain properties (p. 22).

Bennett (2006): “[It is accepted that] it makes good sense to talk about the causal efficacy of properties, to say that $c$ (your thirst, say) caused $e$ (your movement towards the kitchen) in virtue of or qua some but not other of its $c$’s properties. This idea – and the worry that establishing the causal efficacy of particular (token) mental events… does not guarantee the efficacy of event-types or properties – attracted attention in the wake of Davidson’s defense of anomalous monism,” (p. 34).

In McLaughlin, Kim and Bennett, and indeed more widely, it is assumed often without argument that properties (often nomic properties) feature in the appropriate answer to the metaphysical question of what it is in virtue of which a cause is the cause that it is. When $B$ stands in virtue of $A$, then $B$ depends on $A$. The direction of this dependence inherent in the nomic subsumption theory of causation can be reversed with good results for our theory of mental causation. No Catastrophe. Indeed The Reverse View suggests itself as an alternative to the Armstrong-backed nomic subsumption view. See Chapter Two for a discussion of nomic subsumption.

5.2.5 The Reverse View Versus Armstrongism

Nomic subsumption funded by Armstrongism about nomological necessity may be a plausible view. I am not disputing that. I am here only arguing that there are good reasons that motivate looking for an alternative.
Here I scope my criticism down to two main arguments.

(1) Something brute in Armstrongism is just as *prima facie* implausible as brute causal relations in the Reverse View.
(2) There are problems with the Armstrongian account of “necessity,” – worries one would expect a Humean to point out.

With respect to (1), I expose the deep and strange commitments involved with the realism of Armstrong. Ancillary commitments of Armstrongism are just as counterintuitive, at least at first glance, as a view that says causal relations are brute.

With respect to (2), I have already asked for what “straightjackets” (in Blackburn’s phrasing of his take on Hume’s “skepticism” about necessity, see Section 5.2.2) the necessary causal nexus that guarantees it will necessarily be a necessary causal relation at a time in the future, t. Armstrong can assert that it is just brutally necessary, but as above in Section 5.2.2, even its brute necessity needs straitjacketing.

I believe these two reasons motivate at least looking for an alternative to nomic subsumption, such as the Reverse View. Let’s now look at (1).

### 5.2.6 Armstrong’s Counterintuitive Realism

We have seen McLaughlin argue that a view asserting that causal relations be brute would be an implausible view. However, the Armstrongian view must itself appeal to something in its account being brute, namely that such and such universals are necessarily related. The fault here lies not with Armstrong himself, who recognizes that he himself must assert something as brute. The fault is with McLaughlin, as discussed above, for thinking assertion of bruteness is immediately damning.

Is there any hope of demonstrating the necessity of the ultimate connections? I do not believe there is (Armstrong 1985: 159).
Bruteness is not immediately damning. What matters is where you locate the bruteness. Armstrong and McLaughlin seem to agree in their preferred location of what is brute; they prefer it to the location of bruteness I advocate. But the views are on level playing field with respect to any liability in assertion of bruteness insofar as each theory asserts it somewhere.

I would argue that a view that held that causal relations were brute was no more *prima facie* implausible than a view that said such and such universals are necessarily related in virtue of nothing at all, i.e., their necessary relation is brute. It is not a terribly strong argument to say that one’s own position is no more *prima facie* implausible than one’s opponent’s position. However, it is not nothing. A view that asserts brute causal relations should not be dismissed out of hand at first glance as we observed McLaughlin (1993) does. It may eventually be found to be “unchoice-worthy” in terms of theory choice, but how choiceworthy it is requires looking into. If it accounts for the relevant phenomena to be explained, brute causal relations may turn out to be even more plausible than the position of the Armstrongian-backed nomic subsumption view which holds that causal relations depend on laws, an important aspect of which turns out to be brute. I leave that for future work. However, I can imagine virtues of the Reverse View, for example, where it locates the bruteness might be preferable to where Armstrong, for example, locates it.

The Reverse View would be more inline with an “empiricist” point of view on philosophy on this score. An empiricist point of view would, for example, find a conventionalist view on what makes laws different than accidents more appealing than a realist view. After all, the empiricist would say, what we have experience of are only the regularities observable in the past to the present. Those regularities or causal relations we
have observed should be deemed “more real” in an empiricist theory, rather than supposing that what’s “more real” is something arrived at in fact by speculation.

Armstrong (1985) himself says:

But if explanation has to stop short of the Absolute, then we have to accept brute fact, that is, contingency at some point. At what point should we do this? That is a question of the utmost delicacy for every philosophy. In my judgement, the Regularity theory of law gives up much too soon. Instead, I have argued, regularities among particular states of affairs can be explained by connections between universals. … Can these connections in turn be explained? The system of connections may be simplified, and brought under higher-order laws. But when all this has been done, is there any hope of demonstrating the necessity of the ultimate connections? I do not believe there is. Necessity can be asserted, but it cannot be demonstrated or even made plausible (159).

I would argue that it’s not a matter of “giving up too soon” versus sticking it out to the end. It’s a matter of where you take the beginning and ends of things to be. We have a choice, so speak, to go in for brute (and in some sense “more real” – for being explanatory and being that on which other things depend) causal relations or going in for brute necessary connections between universals. Well, if “necessity” is something that cannot be proven and causal relations (which are regularities lacking necessity) are right here before us, then why not think that where “the being of being” begins, and where explanation ends, is at the stuff closer to us, more immediately available? Why give ultimate status to the fictions we create in order to sustain our mind-dependent epistemologies, such as necessary connections among universals?
5.2.7 Conclusion

In this chapter, I briefly surveyed the New Hume debate in order to get the concept of “straitjacketing” on the table. I then introduced my novel observation that the PNCC is ambiguous as to the Euthyphronic order of explanation between causes and laws. The traditional view is to say that causes depend on laws. I articulated the Reverse View which says that it goes the other way around – laws depend on causes, metaphysically speaking. This does not entail Catastrophe. I then compared what’s taken as brute in the traditional answer to what would be taken as brute in the Reverse view, with some implications favorable, I think, for the Reverse view.
CONCLUSION

By way of conclusion allow me to return to two notions I’ve put forward and see if they are mutually consistent. I do not know if they are. One of my aims in this dissertation was to present and defend their plausibility. The further question of their consistency is work for the future. However, I may take a stab at it now.

I have asked two Euthyphro questions: one between reasons and causes and the other between causes and laws. I said that a reason is a cause in virtue of being a reason and not the other way around. (That is, it is not the case that a cause is the reason in virtue of being a cause). With respect to causes and laws, the traditional view – manifest in the nomic subsumption view of causation – has it that causal relations are what they are in virtue of the laws being the way they are. I, however, offer a Reverse View, which takes the Euthyphronic direction to run the other way: the laws are what they are in virtue of causal relations being what they are. According to the nomic subsumption view, when backed by Armstrongism, the laws – second order universal relations between universals – are themselves brute. According to the Reverse View, causal relations are brute and the laws derivative of, or supervenient on, the causal relations. How is it possible that causes are brute, as per the Reverse View at the same time as causes, according to my first Euthyphro question, are not the reason for action in virtue of being causal? How is it that causes are brute with respect to laws but between reasons and causes a cause is what it is in virtue of being a reason and not the other way around? Do my answers to these two Euthyphro appear mutually inconsistent? Let me try to explain how I see it such that they are not inconsistent.
First of all, the Euthyphro question between causes and laws is a question about causes metaphysically-speaking and laws metaphysically-speaking. The Euthyphro question between reasons and causes, on the other hand, is a question about reasons epistemically-speaking and causes metaphysically-speaking. That is an important difference, more about which momentarily.

Secondly, I said “posit” four times in this dissertation. It is time to bring them together. My answer to the Euthyphro question between reasons and causes was explained in the context of my Priority of the Phenomenon Thesis and the Strict Qualification Thesis. According to the latter, we are dealing only with true, genuine, successful reason explanations. According to the former, we move roughly speaking from epistemology to metaphysics. So, according to these two theses when we explain action by reference to reasons we attribute a reason to an agent according to what we called “the epistemology of the reason,” which is broadly speaking hermeneutic. Whatever deliverance the epistemology of the reason issues in is the reason for which the agent acted. And, according to the multiple reasons argument, if a reason is the reason then that reason is the cause of the action.

The fact that we move from epistemic-reason to metaphysical-cause solves the identification problem. The identification problem asked which mental event token was going to be token identical to which physical event token. The Priority of the Phenomenon Thesis, the Strict Qualification Thesis, the multiple reasons argument and my answer to the Euthyphro question between reasons and causes jointly have it that the reason is token identical to the cause and for principled reasons. The cause that the reason is identical to is “programmed for” in Jackson and Pettit’s language or “posited” as I say it. The successful epistemology of the reason imputes to us the further metaphysical commitment that the reason is the cause.
In Section 4.11 on Davidson’s distinction between “cause” and “causally explain,” I asked what in virtue of which c is a cause of e, metaphysically-speaking. I said it would not do to suggest that a cause is the cause that it is in virtue of featuring in a causal explanation. The cause must already be a cause, so to speak, in order for reference to it under some description to be explanatory with respect to the primary explanandum. Maybe we should say that a causal explanation “posits” the cause it references. In a true, genuine and successful reason explanation the posit will in fact be the cause. This is what we said with respect to Bontly’s critique of Yablo’s proportionality constraint. There we noted that causal explanation – explanations that refer to causes under a description – could go deeper and deeper or into ever different contexts forever. Each causal explanation will posit the cause it refers to as the cause and posit that every subsequent true explanation will describe the very same event, even if in radically different ways.

So in some sense the causality of the reason is achieved transcendentally. That the reason be the cause is a necessary precondition of a true, genuine and successful reason explanation according to the multiple reasons argument. We’ve said we have a true, genuine and successful reason explanation, so that means we can know a priori that the reason is causal. What remains to ask is this: Are causal relations which are brute with respect to their derivative, supervenient causal laws incompatible with the causality of reasons being arrived at transcendentally? Can the cause that is the reason be brute? I think we may offer a cautious “Yes.” It is transcendentally posited as brute. This is so because a cause does not become the cause it is in virtue of being referenced in a true, genuine and successful explanation. The reason is already the cause before we refer to it under a description in our true, genuine and successful reason explanation. The causal relation between the reason and the given action can be brute with respect to the law supervenient on it; and this is not inconsistent with the
reason’s causality being arrived at transcendentally. To be arrived at transcendentally just means to be arrived at in the move from successful epistemology to what further is imputed to us metaphysically-speaking.
EPILOGUE

In the Prologue to this dissertation, I said that I would lay out the minimal metaphysics that even a Rortyian pragmatist is or must be committed to. I am speaking only of the Rortyian pragmatist who takes it that adopting Davidson’s philosophy of mind, namely anomalous monism, “allows us to treat poetry and physics even-handedly.” This Davidsonian Rortyian pragmatist is sanguine in his anti-metaphysical stance, such that he does not believe the special sciences need any metaphysical undergirding. My argument has been that in the act of subscribing to the truth, genuineness, and success of true, genuine, successful reason explanations certain commitments, call them metaphysical, are imputed to us, those who subscribe to anomalous monism. The commitments are these: the mental event that is the reason is a physical event and is the cause of the action; additionally, the causal reason is covered by some strict-law, which we need not know. It’s just that we can know that there is such a law even if we don’t know what the law is. So the minimal metaphysics is what other commitment are imputed to us besides what the causal explanation commits us to directly. These metaphysics include the causality of reasons and the lawfulness of causal relations.
BIBLIOGRAPHY


