To my mother and in memory of my father
ACKNOWLEDGMENTS

Sometimes a dissertation is a manifestation of academic flourishing. People may contribute to such flourishing in different ways. They may directly affect the dissertation’s content, as my teachers, students, and peers have done; or they may be the backdrop against which the project has been possible in the first place, by fostering intellectual curiosity and by providing emotional support, as some family members and friends have done. While it is impossible to mention every person who has affected this dissertation in some way, I wish to thank at least those to whom I am most indebted.

In the fall semester of 1999, I took my first Philosophy of Mind course with Matthias Vogel at the Goethe University Frankfurt, Germany. It is hard to overestimate the impact that this course had on me. Since we carefully worked our way through Jaegwon Kim’s introduction *Philosophy of Mind*, I was exposed to most of the arguments, positions, and problems that I am still thinking about, including the exclusion problem. Furthermore, Matthias’ enthusiasm for the subject and his striving of clarity infected me and thus set the course of my further academic development. I wish to thank him for that. It was also through him that I got invited to Wolfgang Detel’s colloquium. Populated by inspiring philosophers who possessed much greater knowledge and skill than I did at the time, it forced me to adapt intellectually, much to my own benefit. Though I am indebted to all the participants of the colloquium, I owe thanks particularly to Wolfgang Detel and Alexander Becker. Later, in 2006, they would constitute the committee for my Magister thesis in which I developed a criticism of Stephen Yablo’s proposed solution to the exclusion problem. In the meantime, I became a teaching and research assistant for Wolfgang. Having him as a teacher as well as a philosophical role model provided me with a solid foundation upon which I am still building. Without a doubt, I am most deeply indebted to him.
among all the excellent teachers I had in Frankfurt. I wish to thank him for the many things I owe him for.

There is not much of a point to building a foundation unless one plans to build something on it. Since I was interested in an area and in a kind of philosophy that was and, I believe, still is best done in the United States, it seemed like a good idea to apply to graduate school there. I got an invitation from the philosophy department of the University of Florida and readily accepted. I wish to thank the department as a whole for giving me the chance to study with philosophers who grew up in the analytic tradition and who have been an integral part of it. Particularly the courses I took with John Biro, David Copp, Michael Jubien, Chuang Liu, Kirk Ludwig, and Gene Witmer broadened my philosophical horizon and closed a good number of gaps. More importantly, they deepened my understanding of the arguments, positions, and problems and allowed me to see connections between concepts that I previously had kept separate in my mind. It is largely due to what I have learned from them that this dissertation focuses not on a problem within a particular area of philosophy but on a problem located at the intersection of the philosophy of mind, metaphysics, and the philosophy of science. Since I think this makes the dissertation more interesting and more valuable, I wish to thank them. Chuang, Gene, and John were also members of my dissertation committee. For that, and for feedback along the way, I owe them as well as the external committee member, Jonathan Edelmann, thanks. Without a doubt, I am most deeply indebted to Gene for being the superb advisor that he was and consequently for being the person I interacted most with philosophically. Numerous thoughts and insights on the following pages are the result of the many fruitful discussions I had with him. He made me see that many of the arguments and positions I deemed clear during my Frankfurt years really are not clear after all. Thus, he forced me to clarify where clarification was much
needed, thereby drastically improving the precision of my thinking and the clarity of my writing. I wish to thank him for all of this and more.

Regarding the backdrop, Andy Lücking has been most influential for my intellectual development. Had it not been for our nightlong discussions of Nietzsche and for our innocent marveling at Frege and Wittgenstein as teenagers, I likely would not have become an academic, let alone a philosopher. It is telling that the owner of our local bookstore still remembers us as the only two kids in town that ordered “weird books.” I consider myself fortunate that a few years later I found another friend, Markus Rohling, with whom I could continue that tradition of excessively long conversations. I hope he knows how much I miss those tranquil evenings at the observatory. Without Andy and Markus, I would not be who I am today. They deserve my gratitude more than anyone else. In Frankfurt, I became friends with Hana Gründler, Oliver Schütze, and Jan-Erik Strasser. Fiercely smart and highly analytical, they have pushed my intellectual boundaries ever since. Letting our minds freely roam conceptual pastures has been among the most pleasurable and deeply satisfying activities in my life. I cannot imagine a life without it, thanks to them. I am confident that Hana, too, considers the loss of our beloved “kitchen of absurdity” to be a true tragedy for the realm of ideas. In Gainesville, Ron Claypool and Aubrey Spivey have provided emotional support and social stability beyond what they may be aware of. I will miss our philosophizing and cracking jokes during lunch breaks, and I will miss our long bike rides on the Hawthorne Trail. I hope they know how grateful I am for those times together.

I also wish to thank Ray D. Fennell for having made me more aware of the beauty of simplicity; Michael Cullinan for not losing patience when I kept nagging him with questions about English grammar and punctuation; Larry McKill for offering some culture when I was
starving for it; Roger Meyer for listening when I needed to blow off steam; Allison H. Roulston for going beyond words with me; Charles Stanley—a.k.a. Carlo Pittore—for showing me how to live a life of complete dedication; and William Neville for his encouragement and support when it was needed, for sharing his academic insights as a man of letters who has “been there, done that,” and for loving Jordi Savall as much as I do. These friends all have been a seemingly inexhaustible source of intellectual stimulation. That two of them are not able to witness the completion of my dissertation saddens me beyond words.

Finally, thanks are due to my brother, Christoph Falke, and my parents, Herbert and Rita Falke, for supporting me in numerous ways ever since I decided to study philosophy. I am sorry that my father, too, is among those who are not able to see the result of their support.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACKNOWLEDGMENTS</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>CHAPTER</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 INTRODUCTION</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>1.1 Motivating the Project</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>1.2 Structuring the Project</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td>2 THE EXCLUSION PROBLEM</td>
<td>23</td>
<td></td>
</tr>
<tr>
<td>2.1 Two Challenges</td>
<td>23</td>
<td></td>
</tr>
<tr>
<td>2.1.1 The Lack of a Terminological Convention</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>2.1.2 The Lack of a Canonical Version</td>
<td>32</td>
<td></td>
</tr>
<tr>
<td>2.2 Formulating the Exclusion Problem</td>
<td>37</td>
<td></td>
</tr>
<tr>
<td>2.3 Generalizing the Exclusion Problem</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>2.4 Chapter Conclusion</td>
<td>55</td>
<td></td>
</tr>
<tr>
<td>3 CAUSATION AND THE EXCLUSION PROBLEM</td>
<td>57</td>
<td></td>
</tr>
<tr>
<td>3.1 A Qualification</td>
<td>57</td>
<td></td>
</tr>
<tr>
<td>3.2 Causal Explanations</td>
<td>58</td>
<td></td>
</tr>
<tr>
<td>3.3 Theories of Causation</td>
<td>62</td>
<td></td>
</tr>
<tr>
<td>3.3.1 Regularity Theories</td>
<td>64</td>
<td></td>
</tr>
<tr>
<td>3.3.2 Counterfactual Theories</td>
<td>72</td>
<td></td>
</tr>
<tr>
<td>3.3.3 INUS Conditions</td>
<td>79</td>
<td></td>
</tr>
<tr>
<td>3.3.4 Probabilistic Theories</td>
<td>83</td>
<td></td>
</tr>
<tr>
<td>3.3.5 Transference Theories</td>
<td>87</td>
<td></td>
</tr>
<tr>
<td>3.3.6 Single-Difference Theories</td>
<td>90</td>
<td></td>
</tr>
<tr>
<td>3.3.7 Causal Pluralism</td>
<td>93</td>
<td></td>
</tr>
<tr>
<td>3.3.8 Causal Antirealism</td>
<td>99</td>
<td></td>
</tr>
<tr>
<td>3.4 Chapter Conclusion</td>
<td>102</td>
<td></td>
</tr>
<tr>
<td>4 STANDARD RESPONSES TO THE EXCLUSION PROBLEM</td>
<td>106</td>
<td></td>
</tr>
<tr>
<td>4.1 Organizing the Responses</td>
<td>106</td>
<td></td>
</tr>
<tr>
<td>4.2 Distinctness</td>
<td>109</td>
<td></td>
</tr>
<tr>
<td>4.3 Causal Impact</td>
<td>116</td>
<td></td>
</tr>
<tr>
<td>4.4 Typicality</td>
<td>124</td>
<td></td>
</tr>
<tr>
<td>4.5 Completeness</td>
<td>126</td>
<td></td>
</tr>
<tr>
<td>4.6 Overdetermination</td>
<td>128</td>
<td></td>
</tr>
<tr>
<td>4.7 Exclusion</td>
<td>130</td>
<td></td>
</tr>
</tbody>
</table>
4.8 Chapter Conclusion ................................................................................................................134

5 REASSESSING THE EXCLUSION PROBLEM .....................................................................139

5.1 Striking a Balance ............................................................................................................139
  5.1.1 Strong Exclusion .........................................................................................................140
  5.1.2 Weak Exclusion ..........................................................................................................143
5.2 Conclusion .......................................................................................................................145

APPENDIX: KANT EXCERPT .................................................................................................149

LIST OF REFERENCES .............................................................................................................150

BIOGRAPHICAL SKETCH ........................................................................................................158
EXCLUDING THOROUGHLY: THEORIES OF CAUSATION AND THEIR IMPACT ON
THE EXCLUSION PROBLEM

By

Andreas Falke

December 2017

Chair: Gene Witmer
Major: Philosophy

The exclusion problem casts doubt on the idea that mental events can have a causal impact on other events in virtue of their mental properties. While many responses to the problem focus on the relationship between mental and physical properties, some analyze other concepts involved, such as the completeness of the physical and overdetermination. However, surprisingly little effort has been made to investigate the exact role causation plays in the problem. An underlying premise of the debate thus seems to be that the problem remains the same regardless of what theory of causation one presupposes. To show that this premise is false, I offer a version of the problem that is more carefully developed than the often sketchy versions discussed in the literature. I then analyze how the problem’s dynamic and its nature change depending on what theory of causation one presupposes. The results of this analysis can help to explain why and how some of the standard responses are not convincing, at least as responses to my version of the exclusion problem. In some respects, that version is stronger, and, in other respects, it is weaker than versions more commonly discussed. As a result, standard responses may not convince as responses to my version but succeed as responses to other versions. Because of that, I review some of these alternative versions with respect to their virtues and vices and conclude the dissertation by defending my version as the one most suitable for future debate and research.
CHAPTER 1
INTRODUCTION

1.1 Motivating the Project

Metaphysics investigates what kinds of things there are in the world, what their nature is, and how they are related to one another. Philosophy of mind is commonly, though not necessarily, viewed as a branch of metaphysics. When viewed as such, it is that branch which tries to understand the nature of the mind as well as the mind’s relation to the body in particular and to the world in general. Since beliefs regarding the nature of the mind affect our theorizing about its relation to the world, and vice versa, careful inquiries in the philosophy of mind require keeping an eye on the broader metaphysical context.

To some degree, such context is involved whenever the mental is contrasted with the physical. If, for example, mental events such as feeling pain are essentially private, then they seem to have a feature that physical events lack. In combination with Leibniz’s law of the indiscernibility of identicals, according to which two things are identical only if they have all their properties in common, it seems to follow that at least some mental events cannot be identical with physical events. If, on the other hand, we endorse some kind of identity theory, we either have to deny that mental events are private or we have to explain how privacy can be accounted for in physical terms.

But from time to time we need to broaden the context even further exactly because the entanglement of the philosophy of mind and metaphysics goes deeper. For instance, when philosophers analyze the relationship between mental and physical properties, they often have recourse to other elements of their ontology such as events and laws of nature. Donald Davidson’s anomalous monism may serve as a good example. According to this theory, it is not possible to reduce mental properties to physical properties exactly because there cannot be any
strict laws that would allow us to predict or to explain mental events, despite the fact that every mental event token is identical with a physical event token.\textsuperscript{1} But while philosophers of mind reflect on how metaphysical assumptions about events, properties, and laws of nature may affect their theories of mind, they remain remarkably silent about the concept of causation.

This is remarkable for two reasons. First, in formulating and discussing theories of mind many philosophers rely heavily on causal talk. If this talk remains unexplained, the resulting theories and discussions may remain as unclear as the causal talk they rely on.\textsuperscript{2} Second, in contemporary metaphysics it is taken for granted that theorizing about the nature of events and properties potentially has an impact on theories of causation, and vice versa.\textsuperscript{3} If the philosophy of mind in fact is a branch of metaphysics, the silent assumption that theories of causation do not affect theories of the mind then becomes rather puzzling unless some explicit justification for this assumption is provided.

The relative silence regarding causation may be explained historically. With the exception of substance dualism and arguably idealism, all of the important theories of the mind—property dualism, nonreductive and reductive physicalism, functionalism, type and token identity theories, etc.—are theories about how mental properties relate to physical properties.\textsuperscript{4}

\textsuperscript{1} See Davidson (1970/1980, 208) or Davidson (1993, 3). As Evnine (1991, 51) points out, Davidson later seems to admit the possible existence of psychophysical laws in the context of generalizations needed to spell out causal dispositions; see also Davidson (1987, 45). It is an interesting question whether, or to what extent, this affects or even undermines his anomalous monism.

\textsuperscript{2} Stathis Psillos shares my worry. He argues that theories that rely heavily on the concept of causation, such as causal theories of knowledge, perception, or identity through time, appeal to little more than “shaky prephilosophical intuitions,” unless we “do some serious groundwork to clarify what exactly we refer to when we speak of causation and explanation”; Psillos (2002, 2–3).

\textsuperscript{3} Shoemaker (1980/1997), for instance, uses the notion of causation to define properties. According to him, properties are clusters of causal powers.

\textsuperscript{4} The fact that most, if not all, introductions to the philosophy of mind focus on this relation is indicative of this. In this respect, Kim (2006) is quite typical.
Most of the philosophical energy of the twentieth-century philosophy of mind went into developing sophisticated versions of such theories. Some of them have obvious implications for entities such as events. If, for example, the identity of an event is determined by which properties it essentially bears, a change in the nature of some of these properties might have implications for the identity conditions of events that bear them. But causation seems different. If, for example, causation is a relation between entities that bear properties, a change in the nature of the causal relation may be less likely to affect the identity of the relata. And this, in turn, may explain why philosophers did not primarily worry about causation. It seems natural to reevaluate old problems like the problem of causal interaction and to approach newer problems like the exclusion problem through the eyes of these theories of the mind first rather than throwing causation into the mix right away, thereby potentially muddying the waters.

The problem of causal interaction arises if we endorse the substance dualist’s claim that mental substances do not have any properties in common with physical substances. In particular, it seems mysterious how something immaterial can causally interact with something material. But even if we grant that mental substances have some physical properties, a mystery remains. For now we may be able to explain how mental substances can causally interact with physical substances, but it seems dubious that they do so in virtue of their mental properties. Stripping this problem of its substance ontology, we almost automatically arrive at the problem

---

5 One exception is their both being temporal, i.e., (most of) their properties are instantiated at particular times.

6 Sometimes, the problem is not couched in terms of immaterial versus material but in terms of abstract versus concrete. Descartes famously thought that mental substances are unextended and nonspatial whereas physical substances are spatially extended; see Descartes (1641/1996, II. 8). We might wonder, of course, how plausible this conception of the mind is. If and when I am having a thought while I am in my office on the UF campus, I might not be able to point to the exact location of my thought; but surely the thought takes place here rather than in India or on Mars or nowhere at all.
of mental causation as it is discussed nowadays: how can mental events have a causal impact on physical events in virtue of their mental properties?

The exclusion problem highlights how severe the problem of mental causation is by placing the assumption that some mental events have a causal impact on physical events in virtue of their mental properties into the context of further metaphysical assumptions that all seem quite plausible: mental properties are distinct from physical properties, the physical realm is causally closed, events are not systematically overdetermined, and no effect has more than one sufficient cause unless it is overdetermined. The problem is that, prima facie, not all of the assumptions can be true at the same time, so at least one of them must be false. Though there is disagreement about which assumption, if any, needs to be given up, most responses to the problem argue from the viewpoint of a particular theory of mind. Partly because of that, a good portion of the discussion focuses on the relation between mental and physical properties while it either ignores or marginalizes the importance of the concept of causation.

This ignorance or marginalization is not necessarily problematic. It might only be reasonable to utilize theories of causation if we have some evidence that in fact they matter. In the absence of such evidence, the most reasonable thing to do may be to proceed as if it were true that problems like the problem of causal interaction and the exclusion problem remain the same no matter what theory of causation we presuppose. The fact that, so far, this assumption has not been supported or justified in any systematic manner then is not objectionable. After all, the burden of proof rests on the person who wants to reject the assumption because rejecting the

---

7 In the literature, the labels ‘the problem of mental causation’ and ‘the exclusion problem’ are not used consistently. Some philosophers use ‘the problem of mental causation’ to refer to the exclusion problem; see, for instance, Harbecke (2008, chapter 1). To avoid confusion, we will discuss these philosophers as if they adopted the distinction between the problem of mental causation and the exclusion problem.
assumption implies positing a connection between the concept of causation and the above problems. As far as the burden of proof is concerned, this is correct. And as long as the historically dominating approach of spelling out the relation between mental and physical properties remains fruitful and provides new suggestions for how to answer the question how, if at all, mental events can cause physical events in virtue of their mental properties, there is no immediate need to change the focus of the debate.

However, the historically dominating approach seems to have reached an impasse. Despite an ever-growing body of literature on the exclusion problem, no satisfactory solution seems to have been offered so far. Among the recent publications, the most promising and productive ones seem to be those that shift their attention from utilizing particular theories of the mind to analyzing the concepts involved in the exclusion problem, such as completeness and overdetermination. While neither these alternative approaches nor the traditional, partly property-based responses to the exclusion problem have yielded convincing solutions so far, the recent debate has had some positive effects. One such effect is the realization that the exclusion problem has previously been presented and discussed in too sketchy a manner for it to be clear what the exact inferences are that are required for the problem to arise. More carefully developed versions of the exclusion problem are now available. Through such fleshed-out versions, the mechanics of the problem can be better understood, formerly overlooked defects can be

---

8 We will offer some support for this claim in Chapter 4. Of course, we will not be able to discuss all proposed solutions. Instead, we will look at some representative strategies and illustrate why they are not fully satisfactory.

9 Haug (2009), for example, distinguishes two kinds of completeness and argues that taking these into account allows nonreductive physicalists to solve the exclusion problem. Funkhouser (2002), to give another example, distinguishes three kinds of overdetermination and argues that not all of them are problematic. So, the severity of the exclusion problem depends on which kind of overdetermination we presuppose.

10 A good example for a more carefully developed version, especially with respect to the distinctness claim, can be found in Stoljar (2008).
addressed, and new responses emerge. In other words, the felt need for alternative approaches has brought some fresh air into the debate.

It is in the context of a search for alternative approaches that the lack of justification for the assumption that the exclusion problem remains the same no matter what theory of causation we presuppose becomes particularly puzzling. Just like completeness and overdetermination, causation is a concept that plays a role in several of the assumptions. But unlike those other concepts, causation is, in addition, one of the key concepts in philosophy of mind in general. Since the entire discipline is saturated with causal talk, we should expect causation to be among the first targets of the shift, especially given that rather sophisticated metaphysical theories of causation are readily available to be utilized. But, surprisingly, relatively few philosophers entertain the possibility that theories of causation might contribute to solving the exclusion problem.

Fortunately, we have overdramatized the situation to some extent. While indeed most philosophers endorse, implicitly or explicitly, the assumption that theories of causation will not help solving the exclusion problem, few philosophers claim that they have no impact whatsoever. Indeed, there are counterexamples to the latter claim. Peter Menzies, for instance, utilizes the concept of contrastive causation to undermine the standard version of the exclusion assumption that no effect has more than one sufficient cause unless it is overdetermined. By showing how contrastive causation implies a modified, more plausible version of the exclusion assumption, Menzies illustrates how thinking about causation can indeed have an impact on the exclusion problem.\(^\text{11}\) Of course, such an impact does not by itself establish any previously unnoticed wiggle room for solving the problem. But even if we can show that Menzies’

\[^{11}\text{See Menzies (2008).}\]
particular modification does not offer such wiggle room, not all possible ways in which thinking about causation might affect the exclusion problem have been explored yet. Hence, it seems unreasonable to rule out the possibility that theories of causation might help solving the exclusion problem. And yet, some philosophers endorse such skepticism.

Some such skeptics argue in too sketchy a manner to justify their confidence. Michael Esfeld, for example, devotes only fourteen pages to showing that the exclusion problem is independent of theories of causation. In a book of over four hundred pages, Jens Harbecke spends barely seven pages on arguing for this independence claim. Most of the space on those pages is used to explain theories of causation. It does not take a professional philosopher to suspect that, given the number of theories that need to be covered, the resulting justification is by no means systematic or satisfactory even if the justification provides interesting grounds for further debate and even if the conclusion drawn happens to be true.

But there are laudable exceptions among the skeptics. Instead of arguing against the possibility that theories of causation in general might help solving the exclusion problem, they have developed criticisms against a particular theory of causation, usually the one that is most commonly utilized as a means to tackle the problem: the counterfactual theory of causation. According to this theory, causal claims of the form “Event $a$ caused event $b$” are best analyzed as to mean that, if event $a$ had not occurred, event $b$ would not have occurred. Arguably, $b$ does not counterfactually depend on the occurrence of $a$ simpliciter but in virtue of $a$’s instantiating one or several properties. The general strategy then is to show that there are cases in which the

---

12 See Esfeld (2007).
14 The version proposed by Lewis (1973a) will be discussed in section 3.3.2.
better candidate for the property that \( b \)’s occurrence counterfactually depends on is a mental property rather than a particular physical property. Terence Horgan and Lynne Rudder Baker are among the advocates of this strategy.\(^{15}\) Some critics of this strategy, like Jaegwon Kim, remain skeptical of the counterfactual theory of causation in general and point out that the mere fact of a counterfactual dependence does not establish a causal relevance of the mental property in question. In the absence of an additional argument that establishes such relevance, the overall strategy to utilize the counterfactual theory of causation for solving the exclusion problem is not convincing.\(^{16}\) But even if the critics win in this case, it remains an open question whether other theories of causation might or might not do the trick.\(^{17}\)

Where do these considerations leave us? There is evidence that some theories of causation affect the exclusion problem. At the very least, they help to clarify some of the assumptions as well as some of the inferences required for the problem to arise. Though it has not yet been convincingly shown that theories of causation can be used to solve the problem, it seems reasonable to explore this possibility by systematically analyzing the impact that different theories of causation might have on it. If the analysis reveals that in fact theories of causation cannot, despite their impact, contribute to a solution, an important premise underlying the debate has been justified. If the analysis reveals that theories of causation can establish new ways to tackle the problem, an important contribution to the debate has been made. A classic win-win situation, or so it seems.


\(^{17}\) Lei Zhong, another critic of the counterfactual strategy, acknowledges this point; see Zhong (2011, 131).
1.2 Structuring the Project

The main goal of this dissertation is to demonstrate how theories of causation may affect the exclusion problem. But to get the project off the ground, we will face two initial challenges in Chapter 2. First, several assumptions of the exclusion problem include causal claims. However, there is no consensus on the nature of causal relata. Are they aspects, events, facts, features, situations, states of affairs, or tropes? Depending on how we answer that question, some of the assumptions would have to be phrased differently. To ensure coherence, we will treat events as the appropriate causal relata throughout the dissertation. Second, many versions of the exclusion problem are rather sketchy and differ in the number of assumptions, their strength, and their specificity. Since there is no canonical version, we will develop our own—one that avoids some of the shortcomings of more common versions. Attention to detail will reveal the need for an assumption that is largely ignored in the literature, namely the typicality assumption, which will become relevant in later chapters. Once all assumptions are established, we will generalize the exclusion problem. Traditionally, it is understood as a threat to the causal impact of mental events in particular. However, the problem may arise for other kinds of nonphysical events as well, namely, for all nonphysical events that are claimed to have a causal impact on other events in virtue of their nonphysical properties. Many special sciences assert such causal impact. The exclusion problem can thus be used to question the veracity of such claims, thereby expanding the scope of the problem beyond the philosophy of mind into the philosophy of science.

In Chapter 3, we will pursue the main goal of the dissertation specified in Chapter 1 and Chapter 2 and build the foundation for assessing standard responses to the problem in the following chapter. Since too many theories of causation are available, we will focus on a representative set. Often, these theories have been developed in such detail that an exhaustive discussion even of only a representative set of theories would be unrealistic. Because of that, we
will present the main idea of each theory at a level of detail that suffices to achieve our goal, which is to show that and how they affect the exclusion problem or our responses to it, even if it means that our discussion of each theory will remain sketchy. We will neither discuss the nuances nor argue for or against the plausibility of each theory, unless doing so highlights something important about the exclusion problem. The resulting analysis reveals that theories of causation sometimes affect the strength of particular assumptions of the problem, and other times they affect the likelihood of rejecting particular assumptions. For example, regularity theories imply a particularly strong interpretation of the typicality assumption; counterfactual theories may help to undermine the nonoverdetermination assumption and, by extension, the threat of exclusion for nonphysical events and properties; and causal pluralism reveals that the exclusion problem might rest on the fallacy of equivocation. By showing how the dynamic and the nature of the problem change depending on what theory of causation we presuppose, the results of the analysis will help to explain why and how some of the standard responses to the problem are not satisfying. Chapter 3 thus sets the stage for Chapter 4.

Even though there are over one thousand articles on the exclusion problem, responses are sometimes divided into only three groups: autonomy solutions, inheritance solutions, and identity solutions. However, there are also responses that cut across this taxonomy as well as theories that could be, but have not been yet, applied to the problem. To ensure that we cover a good amount of territory in spite of being forced to be highly selective, we will discuss at least one response from the three types of standard solutions, one that cuts across the taxonomy, and one that has not been discussed in the context of the exclusion problem yet. Our discussion will yield two results. First, the version of the problem that we have developed in Chapter 2 makes it indeed more difficult for some of the responses to succeed. Often, our version changes which of
the assumptions are likely to be given up in the light of a particular response or view. As an example, one traditional response is to distinguish between causal relevance and causal efficacy and argue that the exclusion problem may succeed in excluding mental events from being causally efficacious but not in excluding them from being causally relevant. Since our version specifies the nature of the causal impact of such events in a way that many other versions do not, that response amounts to little more than rejecting the very assumption that it was meant to support in a less fleshed-out version. Unfortunately, the plausibility of such rejection will depend on how well the difference between physical and nonphysical events lines up with the distinction between relevance and efficacy, and we can show that it does not line up well. As a result, the response may fail. Second, some of the responses work very well in combination with particular theories of causation but not at all in combination with others. For example, one response claims that what motivates the completeness assumption also motivates a further claim that is related to completeness but usually not discussed in the context of the exclusion problem, namely, the claim that physics provides a complete inventory of the fundamental kinds and categories that are sufficient for a realization basis for all of the events that have a causal impact. We can show that this assumption in combination with a transference theory of causation may imply that the impact assumption is false while it does not threaten impact if we presuppose a regularity theory of causation. The conclusion of Chapter 4 will thus be that theories of causation have an impact not only on the exclusion problem itself but also on the plausibility or the impact of at least some of the responses to it.

While some of the standard responses may not be convincing in response to our version of the problem, they may be promising as responses to other versions. For our discussion in Chapter 4 to have weight, we need to establish that our version is indeed a good one to endorse.
We will thus spend Chapter 5 on explaining why our version is to be preferred over some stronger and weaker versions sometimes discussed in the literature. This discussion will also serve as a justification for why our version is particularly well suited as a basis for future research. It strikes a balance between specificity and strength on the one hand and the avoidance of unnecessary ontological commitments on the other. Other versions rarely achieve such balance. It makes it easier for philosophers who differ in their views on causation, completeness, events, overdetermination, and the mind to assess the problem and to respond to each other. In the final section, we will summarize the lessons learned and highlight some open ends that need to be addressed in future research.
CHAPTER 2
THE EXCLUSION PROBLEM

2.1 Two Challenges

In a paper published in 1968, Norman Malcolm asks whether it is conceivable that neurophysiological states and processes are sufficient causes of human behavior. His answer is negative for two reasons. If such states and processes were causally sufficient, the truth of such a view could not be asserted consistently because the act of asserting requires mental states or processes to play a causal role in the production of the assertion. And if the assertion could not be produced by mental states or processes, it could not be made on rational grounds; for that requires causally responding to reasons.¹ For these arguments to work, however, he needs to further assume that there being sufficient neurophysiological causes precludes other factors from playing a causal role in the production of the assertion. Malcolm thus arrives at an early formulation of the exclusion problem.² In spite of occasional responses from prominent philosophers such as Daniel C. Dennett, Alvin Goldman, and G. H. von Wright in the 70s, it took almost 20 years for the problem to gain real momentum.³ Partly due to the work of Jaegwon


² While Malcolm’s paper offered an early formulation, as Robb and Heil (2013, section 6.2) correctly point out, it was by no means the earliest. MacIntyre (1957) had developed similar arguments. And even a passage in Immanuel Kant’s Critique of Pure Reason could be interpreted as fusing questions concerning the compatibility of free will and determinism with problems regarding causal exclusion; see Kant (1781/1997, A536/B564), cited in the appendix. If we interpret talk about freedom as talk about mental events that have certain features that are causally relevant, that passage could be viewed as a much earlier predecessor of the exclusion problem. Against this background, it is not surprising that Davidson’s anomalous monism, which can be understood as a response to the threat of the causal exclusion of the mental, is sometimes seen as the continuation of Kant’s attempt to reconcile free agency and physical determinism; see, for instance, Kim (2003, 129). Lately, such Kantian concerns have resurfaced in the form of exclusion arguments against free will. See, for example, List and Menzies (2017) for a criticism of such arguments.

Kim, which helped structure the somewhat scattered responses up to that point, the exclusion problem has become one of the central problems in the philosophy of mind.

The newly gained popularity also yielded an abundance of versions of the exclusion problem. These versions differ in the number of assumptions, their strength, and their level of specificity. This diversity has persisted. As of now, no canonical version of the exclusion problem has been developed. Because of that, our first order of business should be to look at some of these versions, identify their similarities and differences as well as their strengths and weaknesses, and develop a version of the problem that seems suitable for our purposes. Our version needs to keep the assumptions informative enough for the problem to arise; but it also needs to keep them general enough for the traditional responses to the problem to still make sense, although a fleshed-out version might well illustrate how and why these responses are not satisfactory. Finally, it needs to keep the assumptions flexible enough for them to be adjustable to different theories of causation. But for all that to be possible, we need to overcome a second obstacle first: it is unclear what the appropriate relata of causal relations are. Among the suggested candidates are aspects, events, facts, features, situations, states of affairs, and tropes.\(^4\) In order to develop a suitable version of the problem, we thus need to decide which causal relata we wish to treat as the appropriate ones and how we want to phrase causal assertions, that is, we need to establish a terminological convention regarding the ontological elements between which we posit causal relations. In the broader context of the dissertation, the purpose of such terminological convention is not to endorse a particular ontology. Rather, it is to allow for a coherent way of talking about the exclusion problem and to analyze and compare standard solutions to the problem more effectively.

\(^4\) Schaffer (2016, section 1) lists the key proponents and publications for each of these suggestions.
2.1.1 The Lack of a Terminological Convention

In ordinary language, we do not worry much about causal relata. When a friend admires a statue in the garden, we might proudly proclaim, “I made it,” thereby positing a causal relation between a person, us, and an object, the statue. We might explain another friend’s being in the hospital by saying, “She fell down the stairs and broke her left ankle,” thereby citing an event, the falling down the stairs, as a cause of either a state, her ankle’s being broken, or an event, the breaking of her ankle. That is, we often cite different kinds of relata as causes and effects and seem quite happy mingling them. Doing so is not necessarily problematic, for it often is possible to translate a statement citing one type of relatum into a statement citing another, seemingly without loss of information. After all, we could easily switch between talk about the state of our friend’s ankle being broken and talk about the event of her breaking the ankle, even though events and states may be different kinds of ontological entity. However, even if and when such translations seem successful, our metaphysical intuitions regarding the resulting assertions might vary. The worry here is reminiscent of an argument made by Tamar Szabó Gendler in the context of personal identity. She suggests that our responses to thought experiments are sometimes based on how the scenarios are described rather than on what they describe. If she is right and if her point generalizes to other types of descriptions, adopting a terminological convention that leads to particular descriptions of the issues involved in the exclusion problem may not be as innocent as it initially seems.

By the same token, some issues that seem particular to some causal relata may turn out to be general. Roderick Chisholm’s theory of agent causation may illustrate the idea, albeit in the

---

context of the problem of free will. According to Chisholm, sometimes agents rather than events or states of affairs cause behavior. Of course, an action such as intentionally lowering the right hand might involve many intermediate causes: there are neural events in the brain, muscle contractions, and so forth. However, when we trace the causal chain back to its origin, assuming that there is one, at some point we will find a primitive causal relation with no intermediate causes, and that causal relation has the agent and a particular brain event as its causal relata. At first, this view may seem puzzling at best and incomprehensible at worst. It is hard to resist the question how exactly an agent causes that brain event. But Chisholm responds that there is nothing to explain. The agent just does it.

According to him, if we are puzzled, we are puzzled about causation in general, not about agent causation in particular. To see what he has in mind, we need to ask ourselves how we would explain how an event that initiates a causal chain causes that first effect in the chain. If there are no intermediate causes and thus no causal mechanisms to refer to, it is hard to see what we could refer to beyond the fact that events of such type typically produce effects of that type. But Chisholm would make the same point about agents. Agents just are the sort of entities that produce free actions. Of course, this is not to suggest that Chisholm’s theory is free of problems. For starters, we might ask what explains that the agent makes the decision to lower the hand now rather than later. To answer the question satisfactorily, we may have to refer to the agent’s having certain desires, intentions, or thoughts now rather than later. But then the agent causal view seems to collapse into a regular event causal view. Still, Chisholm’s theory illustrates how

---

7 Ibid., p. 150.
some of our intuitions may be taken to be specific to certain kinds of causal relata—agents, for instance—when they really are about causation in general. And sometimes it takes unfamiliar ways of framing the issues to make us see how puzzling some of our conventions are even though they may appear very natural and clear to us. The danger then is that we may fail to pick up on some problems lingering in the background of the exclusion problem exactly because we stopped thinking about some types of relata, for they might highlight these problems in ways that our adopted terminological convention does not.⁹

So, adopting a terminological convention is risky for two reasons: our intuitions may be swayed in unintended ways, and the convention may make us overlook issues that might hold the key to solving the exclusion problem. The good news is that every convention carries these risks, so as a starting point we should just opt for the most commonly used relata in the metaphysics of causation, namely events,¹⁰ since this option will make it easiest for us to apply these theories of causation to the exclusion problem in Chapter 3. Two of the most prominent theories of the nature of events are those proposed by Donald Davidson and Jaegwon Kim.

Davidson treats events as datable and locatable, nonrepeatable particulars that are fundamental in the sense that they are not constituted or realized by anything else. As such, they are unanalyzable. Happily, we can refer to them by way of descriptions. As is typical of particulars, their descriptions can vary drastically. Our friend’s falling down the stairs might also be described as her tripping on the third stair and landing, left foot first, on a particularly hard

---

⁹ Partly because of that, this dissertation should be thought of as a first step towards a larger project, namely, a comprehensive primer to the exclusion problem. Such a primer would have to keep track of the possible impact of different terminological conventions as well as of concepts other than causation, such as completeness, overdetermination, etc.

¹⁰ I am not alone in this assessment. Schaffer (2016, paragraph 1), too, claims that the standard view of causal relata is that “they are of the category of event, and that their number is two…”
slate floor. But while the latter description is more fine-grained, it serves the same purpose: to pick out the event that caused the breaking of our friend’s left ankle. Davidson takes this very seriously. Rather than privileging the latter description as the one that cites the features of the event in virtue of which our friend’s ankle broke, he would view the fall’s preceding the breaking, however described, and its being causally related to the breaking as part of the fall’s identity conditions.\textsuperscript{11} He thus distinguishes between the way we describe and pick out events on the one hand and how events are ontologically individuated on the other hand. His theory might not suit our purposes well because the exclusion problem is based on the idea that causal descriptions and explanations involving the mental may sometimes compete or be inconsistent with those involving the physical. But if mental and physical descriptions just are different ways of picking out causally related events, it is unclear how that inconsistency is supposed to arise. So, rather than utilizing a Davidsonian theory of events as the basis of our terminological convention, we should discuss it later, in Chapter 4, as a response to the exclusion problem.

Whereas Davidson’s account seems to be driven mainly by semantic concerns, Kim construes events in a manner that he deems most useful for our thinking about causation and explanation. Rather than being fundamental, events are complex, structured entities; they are exemplifications of a property by a substance at a time. And by ‘substance’ he means any object or any kind of entity that can be the bearer of properties.\textsuperscript{12} Thus construed, it is natural to cite some of the properties constituting a particular event as those in virtue of which it is causally related to another event. Arguably, if our friend screamed loudly while falling, the scream did

\textsuperscript{11} See Davidson (1969/1980, 179), where he states that “events are identical if and only if they have exactly the same causes and effects.” Since Davidson developed his view over years and articulated bits and pieces in various papers, Simon Evnine’s brief reconstruction of Davidson’s view is a useful starting point; see Evnine (1991), especially pp. 26–33. Compare Bennett (1988) for a criticism of Davidson’s view.

\textsuperscript{12} See, for instance, Kim (1976/1993, 34).
not cause her ankle’s breaking, even though it was part of the fall, whereas her landing with her left foot first, perhaps in a certain angle with a certain force, did. While this sounds like a plausible causal explanation, there is a catch. Kim’s view may entail that, strictly speaking, every property exemplification constitutes a distinct event and that every difference in properties implies a difference in identity. So, our friend’s falling down the stairs may not be the same event as her tripping and screaming loudly, even though Kim can construe the latter as being part of the former. For Davidson, these are merely different descriptions of the same cause. Why does this matter? A Kimean account might sometimes force us to single out events and properties as causes in a way that a Davidsonian account would not. In doing so, it also might exclude other contenders; that is, there may be cases where the exclusion problem arises in virtue of the theory of the nature of events endorsed.

Where does the above leave us with respect to our desired terminological convention? On the one hand, Davidson’s theory is attractive because it allows for an intuitive way of talking about events in that it is not overly fine-grained. But there is a serious disadvantage to adhering to Davidson’s theory in general. The properties cited in causal statements mainly serve, on his theory, to pick out the events between which causal relations are posited. It remains unclear, though, what role they play ontologically, if any. Are any of the properties cited in a description those in virtue of which an event is a cause? Given Davidson’s nominalistic inclinations, he might simply reject such in-virtue-of talk altogether, if ontologically understood. Since most

---

15 While this is not the place to discuss all problems with Kim’s account, it is worth pointing out that one criticism alleges that it lacks the resources to distinguish between events and facts. Hendrickson (2006) discusses this criticism and defends a modified Kimean account against it.
discussions of the exclusion problem heavily rely on in-virtue-of talk, it thus seems unwise to adopt too much of Davidson’s theory. Kim’s theory, on the other hand, is attractive because it is compatible with such in-virtue-of talk. But then again, his theory seems overly fine-grained. In addition, it would be, strictly speaking, false to say that an event has a property because the property exemplification is partly what constitutes that event.\textsuperscript{16} But it seems rather natural to speak that way of events and properties.

In light of these advantages and disadvantages, might there be a way of talking about events without committing to a particular theory of them, perhaps even a way that preserves some of the advantages? Suppose we simply asserted that event $a$ has property $P$ and that $a$ causes event $b$ in virtue of property $P$, without specifying whether events are simple or structured or whether they bear properties or are partly constituted by them. Could we make sense of the assertions from the viewpoint of each theory of events? If we adopted a Davidsonian theory, the role of the in-virtue-of clause might be unclear, depending on whether we adopt his nominalist inclinations in addition to his theory of events. But if we rejected his nominalism, we might understand the statement as claiming that $a$ is the cause of $b$ and $P$ is a property of $a$ that figures into a causal law that $a$ and $b$ fall under.\textsuperscript{17} If we adopted a Kimean theory, we might understand the statement as claiming that $a$ is an exemplification of $P$ by an object at a time and it is in virtue of $P$ that it is the cause of $b$. So, while each theory will require some rephrasing of the assertions, we still can make sense of them.

\textsuperscript{16} This may be a simplification. At times, Kim seems to think that his account may be consistent with treating some properties as properties of events, namely, when the properties are not essential, that is constitutive of it. See Kim (1976/1993, 42–49).

\textsuperscript{17} According to Davidson (1970/1980, 207), “events related as cause and effect fall under strict deterministic laws.”
Sometimes, clarity might require specifying the time at which an event has a certain property or at which an event occurs, as in “event a has property P at t,” or “event a at t, causes event b at t.” A non-nominalist version of the Davidsonian theory can easily accommodate that distinction because events can persist independently of the properties they bear. But within a Kimean theory, that distinction may be more difficult to capture, for the time of the property exemplification is the time at which the event takes place. Within a Davidsonian theory, it is also informative to specify the times in the above way, for the identity of the event is determined by its causal relations to other events. They may hold even when an event in question occurs a bit earlier or later. Not so in Kim’s theory. Since the triad of the entity that bears the property, the exemplified property, and the time of the exemplification constitutes the event, to specify the time of an event’s occurrence in the above way is either redundant or poorly phrased.

So, while each theory, if endorsed, would affect the way we phrase causal statements in general as well as the assumptions of the exclusion problem in particular, it does not seem necessary to commit to either one of them. It seems we understand the statements above without settling whether events are simple or structured and without specifying whether events are bearers of properties or partly constituted by them. Positing causal relations between events, considering the properties had by these events, and specifying the properties in virtue of which the causal relations hold seems to suffice as a basic terminological convention. We may consider the statement and the terminology ontologically neutral in the sense that it can be interpreted in a manner consistent with the Davidsonian and Kimean event theories while preserving some of their virtues. Such neutrality seems to serve the main purpose of a terminological convention

18 Such neutrality is also desirable because neither Davidson’s and Kim’s account of events may have withstood the test of time. Other theories have been developed that might be more plausible as an ontological view. But since our goal was merely to adopt a terminological convention, the discussion above will suffice. For a defense of Kim’s
for our project, namely, to increase coherence. Different authors cite different kinds of causal relata, utilize different theories of events, and so forth. A relatively neutral way of formulating the assumptions of the exclusion problem will then allow us to discuss these authors without switching back and forth between different ways of phrasing the assumptions and the arguments proposed.

2.1.2 The Lack of a Canonical Version

Often the exclusion problem is discussed without being carefully developed. We are to understand the pull of the problem without the details being clear. If we attempted to provide clarification, we would be bound to make interpretative decisions eventually yielding claims of varying strength. As a result, many versions of the exclusion problem exist some of which merely differ in their wording without changing the content much while others vary in strength or even in the amount of assumptions included. Consider, for example, how Karen Bennett presents the problem:

**Distinctness:** Mental properties (and perhaps events) are distinct from physical properties (or events).

**Completeness:** Every physical occurrence has a sufficient physical cause.

**Efficacy:** Mental events sometimes cause physical ones, and sometimes do so in virtue of their mental properties.

**Nonoverdetermination:** The effects of mental causes are not systematically overdetermined; they are not on a par with the deaths of firing squad victims.

**Exclusion:** [N]o effect has more than one sufficient cause unless it is overdetermined.\(^{19}\)

---

\(^{19}\) Bennett (2008, 281); bold in the original.
As suggested in Chapter 1, the exclusion problem consists in the prima facie inconsistency of these assumptions. One of them has to be false, given the truth of the others—or so it seems. The big question is, of course, which of these assumptions should be rejected and for what reasons. But these reasons might be more or less compelling depending on how exactly we understand these assumptions and how exactly we phrase them. We might wonder, for instance, whether the completeness claim needs to be as strong as it is in Bennett’s version. What if there are some events that have no cause? Some say quantum physics allows for such events.  

If that is true, perhaps we need only to claim that every physical occurrence that has a cause at all has a sufficient physical cause. This weaker version of the nonoverdetermination assumption seems less likely to be rejected than Bennett’s strong one; yet it seems to be all that is required for the problem. We then might be better off adopting such a weaker version.

Some authors use the alleged truth of some of these assumptions as a reason for rejecting another. Stephen Yablo, for instance, introduces one such exclusion argument, an argument for epiphenomenalism:

If an event $x$ is causally sufficient for an event $y$, then no event $x^*$ distinct from $x$ is causally relevant to $y$ (exclusion).

For every physical event $y$, some physical event $x$ is causally sufficient for $y$ (physical determinism).

For every physical event $x$ and mental event $x^*$, $x$ is distinct from $x^*$ (dualism).

So, for every physical event $y$, no mental event $x^*$ is causally relevant to $y$ (epiphenomenalism).

\[20\] See Healey (2009, 673).

\[21\] See Kim (2005, 16).

\[22\] Yablo (1992a, 426) and Yablo (1992b, 247–248); italics in the original.
If we compare Yablo’s version with Bennett’s, we will notice a number of important differences. Aside from the different labeling and wording of the assumptions and the fact that the rejection of one of these assumptions is turned into a conclusion, the seeming omission of one assumption is particularly striking. While it looks as if Yablo has dropped the nonoverdetermination assumption, he really seems to have merged it with the exclusion assumption. As a result, the exclusion assumption seems stronger than the one suggested by Bennett. But this comes at a price: the exclusion assumption thus formulated may be implausible, for assuming the existence of some overdetermination cases might be perfectly reasonable. What we should be worried about instead is systematic overdetermination, as Bennett correctly observes. Additional complications linger. What exactly does it mean for an event to be causally relevant to another? What if the threat of exclusion stems not from two distinct events but from two distinct properties of the same event? Can we not expect there to be a very similar threat of exclusion if the causation involved is probabilistic, even though in such cases the exclusion and determinism assumption as formulated are false because in such cases the cause is not causally sufficient for the effect?

Of course, Bennett and Yablo are perfectly aware of some of these issues and address them. Bennett, for example, spends significant time on explaining what the notion of overdetermination requires, namely the nonvacuous truth of certain counterfactuals that she specifies, thus illuminating the nonoverdetermination assumption. And Yablo utilizes the rather

---

23 Some authors include more assumptions rather than fewer. Jesper Kallestrup attributes the view to Jaegwon Kim that the causal exclusion problem involves an additional assumption, namely, a supervenience claim; see Kallestrup (2006, 465). It is not clear, however, whether Kim in fact holds this. Kim (2005, 21) distinguishes between the exclusion argument and the problem of mental causation, and it is only the latter, broader problem that involves supervenience. Kim (1998) claims that the exclusion problem arises from the supervenience argument, but the exclusion problem itself, as described by him, does not include the supervenience claim.

24 See Bennett (2008, 288).
vague notion of causal relevance to explain how thinking about the relation between mental and physical events and properties as analogous to the relation between determinables and determinates can avoid the threat of overdetermination and exclusion.25 Others, such as Matthew Haug and Eric Funkhouser, carefully distinguish between different kinds of completeness and overdetermination and investigate the impact each kind would have on the exclusion problem. Haug, for instance, begins his analysis with the following formulation of the completeness assumption:

\textit{Causal sufficiency completeness (CSC):} For every physical event and every time \( t \), if that event has a sufficient cause at \( t \), then it has a physical sufficient cause at \( t \). In other words, every physical event is completely causally determined, insofar as it is causally determined, by physical events.26

As a starting point, this is not bad, for it avoids the problem that Yablo’s strong formulation of the assumption had and it is more concise than Bennett’s version—and this before he even starts distinguishing different kinds of completeness.

The examples above may suffice to illustrate the range of differences regarding how the exclusion problem is presented and discussed in the literature. Unfortunately, there is no canonical formulation of the problem, just many versions of it. Some versions are so sketchy that they merely appeal to our intuitive grasp of the problem. The inferences required for the problem to arise are far from clear. And if turned into an argument against the truth of one of the assumptions, the result is nowhere near a deductive argument. Others spell out the assumptions more, but do so in different ways resulting in claims of varying strengths. This makes it difficult to compare and assess responses to the problem, for they are responses to particular versions of

\footnotesize{25 See Yablo (1992a, 432–434) and Yablo (1992b, 253ff.).}

\footnotesize{26 Haug (2009, 382).}
it. What might seem like a promising approach to one version might not be promising at all regarding another version. More importantly, if we introduce a particular version of the problem, one that is not used and shared by those philosophers whose approach we assess, it might be easy to get caught in a straw man fallacy of the following sort: Person $X$ claims that theory $T$ offers a solution to problem $Ex_1$, but theory $T$ does not work as a solution to problem $Ex_2$. Therefore, $X$ is mistaken.

Of course, this is not what we are going to do. Since, ultimately, we want to analyze the impact that theories of causation might have on the exclusion problem, details will likely matter. We should thus aim for making the assumptions as clear as possible, especially those that feature causal notions. Striving for such clarity by eliminating ambiguities and vagueness sometimes will result in a higher degree of specificity. Sometimes, such specificity is a virtue, but other times it may be counterproductive. We need to keep a balance between specificity and general applicability. Our version should capture cases in which distinct types of events are causally excluded as well as cases in which a distinct feature or property of one and the same event is excluded from being causally relevant. And it should be applicable to probabilistic causation as well as to nonprobabilistic causation.

In addition, we should strive for making the assumptions as strong as possible without undermining their plausibility. For example, Yablo’s version of the exclusion assumption is stronger in the sense that it excludes the possibility of overdetermination altogether whereas Bennett’s version is compatible with nonsystematic overdetermination. Given that examples of alleged overdetermination are routinely discussed in the literature, Yablo’s version seems unnecessarily strong, if not implausible, in light of such examples. In addition, the plausibility of the assumptions will also depend on how strong a version of the exclusion problem they are part
of. An unnecessarily strong assumption embedded in a set of weaker assumptions that entail relatively little and that are more difficult to reject may sway our intuitions regarding which assumption should be rejected. By the same token, the weaker the assumptions, the less force our discussion will have; for it is possible that an approach avoids problems in weak versions while failing to solve problems in strong versions. A view that entails an occasional nonsystematic overdetermination of events might be used to respond to Bennett’s version of the exclusion problem, but it would be useless as a response to Yablo’s version.

If an approach can convince even as a response to the strong version, that is better evidence for its promise and quality than its managing to deal with weak versions. Applying standard responses to our strong version thus is to be considered a test for the quality of the response. It is not to be understood as a representation regarding what the authors were arguing. It is possible, for instance, that Yablo’s utilizing the determinable-determinate distinction to deal with his version of the exclusion problem might be a promising approach while it might be a nonstarter in response to our version of it. In such situations, the conclusion will be that there is a version of the exclusion problem that such standard responses cannot deal with successfully. This is not yet to say that our version of the exclusion problem is the most plausible one available. In the final chapter, we will assess the plausibility of a particularly strong version of the exclusion problem, compare it to weaker versions, and suggest that our version is a promising candidate for a canonical version that might prove most useful for future discussion.

2.2 Formulating the Exclusion Problem

If mental properties were identical with physical properties, there may be no particular problem with mental causation. Mental events could have a causal impact on physical events in virtue of their mental properties simply because mental properties are just physical properties.
Hence, mental causation may be no more mysterious than physical causation. For the problem of mental causation and the exclusion problem to get off the ground, we need to assume that mental properties are not identical with physical properties. As we have seen earlier, this is more or less how the first assumption sometimes is phrased. However, it could be the case that nature is not that homogenous. Some mental properties may be identical with physical properties, others may not be.²⁷ Hence, we should use a weaker version of the distinctness assumption:

1. **Distinctness:** There is an important kind \( k \) of mental property, such that, if a mental property is of that kind \( k \), it is distinct from any physical property.

It is important to appreciate just how strong the last part of (1) is. If we thought of mental properties as complex, highly relational physical properties, for instance, we would not be thinking of mental properties of the kind \( k \). So (1) does not seem compatible with some versions of reductive physicalism, certainly not those that are effectively type identity theories.²⁸ Proponents of such theories will presumably remain unmoved by the exclusion problem to begin with.²⁹

The second assumption of the exclusion problem states that mental causation with respect to bodily behavior exists; i.e., there are mental events that cause physical events in virtue of their mental properties. Most philosophers phrase this in terms of the causal efficacy of mental events. Again, we have to be cautious, for ‘causal efficacy’ is a term of art that acquires its content partly

---

²⁷ Jaegwon Kim, for instance, claims that while most mental properties are fully functionalizable and physically reducible, qualia are not. See Kim (2005, chapter 6).

²⁸ This is not the place to provide a comprehensive overview of the many variants of reductive physicalism. Suffice it to say that some versions are supposed to be reductive without assuming identity. Ernest Nagel, for example, argues that reduction merely requires that there are bridge laws connecting mental predicates \( M \) with neurological, or for our purposes physical predicates \( P \) such that ‘\( x \) is \( M \) iff \( x \) is \( P \)’ is true. That would seem compatible with (1). For a brief overview, see Stoljar (2015, section 7).

²⁹ David Papineau might agree, for he points out that it is the distinctness assumption that seems to get us into trouble in the first place by generating puzzles about how the brain generates mental states and how such states can ever be causally relevant; see Papineau (2002, 2f.).
through the terms it is contrasted with. Stephen Yablo, for instance, distinguishes causally efficacious events from causally relevant events and explains the latter by analogy to the relation between determinables and determinates.\textsuperscript{30} Fred Dretske, to give another example, calls causally efficacious events triggering causes and distinguishes them from structuring causes. While triggering causes explain why an event happens right now, structuring causes explain why a particular event rather than another event causes that event.\textsuperscript{31} Unlike in Yablo’s analysis of causally relevant events, the historical dimension of structuring causes is important. As a result, ‘causally efficacious’ means something slightly different in Yablo’s theory than in Dretske’s. No doubt, the most general formulation of the exclusion problem should be free of such theory dependencies. To highlight this lack of theory dependence, let us introduce another term of art: ‘causal impact.’

To shed some light on what might be meant by ‘causal impact,’ we can characterize it by its functional role for the exclusion problem. In section 2.1.2, we saw that the exclusion problem is sometimes turned into an argument for epiphenomenalism. Roughly put, epiphenomenalism states that mental events play no causal role in the production of effects.\textsuperscript{32} However, this can be interpreted in two different ways. For an event to play no causal role may mean that it is not strictly involved in the production of the effect; i.e., it is not part of the causal chain that leads to the effect. Alternatively, it may mean that it has no explanatory value; i.e., it is not needed for a satisfactory causal explanation of why an event occurred.\textsuperscript{33} Accordingly, a strong version of

\begin{itemize}
  \item \textsuperscript{30} See Yablo (1992b, 249ff.).
  \item \textsuperscript{31} See Dretske (1993, 122).
  \item \textsuperscript{32} See, for example, Robinson (2015).
  \item \textsuperscript{33} Playing no causal role in this latter sense would rule out both causally relevant events à la Yablo and structuring causes à la Dretske.
\end{itemize}
epiphenomenalism would state that mental events are never strictly involved nor do they ever have any explanatory value. A weak version of epiphenomenalism would state that mental events are never strictly involved but that they could have some explanatory value.\textsuperscript{34}

If an event’s having a causal impact merely implies that the event is strictly involved in the production of the effect, the exclusion problem could indeed be turned into an argument for epiphenomenalism—namely, if it concludes that no mental event can be strictly involved in the production of physical or mental events. This, however, would leave open which version of epiphenomenalism is true. If, on the other hand, an event’s having a causal impact implies that the event is either strictly involved or that it at least has some explanatory power, the problem could be turned into a specific argument for strong epiphenomenalism—namely, if it concludes that mental events can neither be strictly involved nor have any explanatory power. While the second interpretation of ‘causal impact’ would result in a much more severe problem, it might be too strong for what the exclusion problem is usually taken to show. Yablo (1992b), for instance, concludes that the exclusion problem succeeds in establishing that mental events cannot be causally efficacious, but it does not establish that mental events are not causally relevant, i.e., they don’t have any explanatory value at all. To avoid the failure of the resulting arguments because they were too ambitious, it then seems reasonable to adopt the first interpretation of ‘causal impact’ at this point. Hence, whenever we talk about an event’s having causal impact from now on, we will talk about an event that is strictly involved in the production of the effect, i.e., an event that is part of a causal chain that leads to the effect. We have to bear in mind,

\textsuperscript{34} It is an interesting question whether having explanatory value in this context amounts to anything ontologically. Disagreements about how to answer this question might be partly responsible for Donald Davidson’s being charged with commitment to epiphenomenalism despite being a monist. For a discussion of this charge, see Davidson (1993) and McLaughlin (1993), for instance.
though, that this disqualifies the exclusion problem from being turned into an argument for strong epiphenomenalism unless further assumptions are added.

Let us now return to the second assumption of the exclusion problem. The claim that mental events have some causal impact on physical events in virtue of their mental properties is clearly too strong because it is too general. It seems quite possible that while some types of mental events do have a causal impact in virtue of their mental properties, others do not. Thus, the claim needs to be weaker:

2. **Causal Impact:** Sometimes, mental events have a causal impact on physical events in virtue of their mental properties.

Because we were so concerned about the generality of (1) and (2), we now run into another problem. The conjunction of (1) and (2) is consistent with two different stories about the causal impact of mental events.

2.1 Some mental events have a causal impact on physical events in virtue of mental properties that are identical with physical properties.

2.2 Some mental events have a causal impact on physical events in virtue of mental properties of the kind $k$ that are distinct from any physical properties.

As briefly mentioned above, the exclusion problem is sometimes used as an argument against nonreductive physicalism and other times as an argument for epiphenomenalism. To be able to utilize the problem in this way, we need to ensure that the mental properties addressed in (2) are at least sometimes distinct from physical properties, i.e., at least sometimes they need to be of the important kind $k$, as posited in (2.2). We should modify (2) accordingly:

2.* **Causal Impact:** Sometimes, mental events have a causal impact on physical events in virtue of their mental properties of the kind $k$.

There remains a problem. Its severity will become fully apparent only after the assumptions regarding completeness, nonoverdetermination, and exclusion have been introduced and
discussed. Nonetheless, this is the best time to address the problem because it will lead to a further assumption that best is seen as a qualifier of \( (2^* ) \).

Many philosophers believe that a view according to which many events are or might be overdetermined is not plausible. This is why assumption \( (5) \) of the exclusion problem is a claim about the nonoverdetermination of physical events. In a strong version, assumption \( (5) \) then would claim that events are never overdetermined.\(^{35}\) Thus interpreted, the claim would line up nicely with \( (1) \) and the yet to be discussed assumptions of the exclusion problem. If it is true that (a) for every physical event there is another physical event that has caused it, that (b) events are never overdetermined, and that (c) no event has more than one sufficient cause unless it is overdetermined, it seems to follow that \( (2^* ) \) is false. However, we may be inclined to think that under certain circumstances particular events might very well be overdetermined. A popular example is the firing squad scenario. We are asked to imagine a firing squad executing a prisoner. Two of the bullets hit the prisoner in the same deadly manner, say, in the head, at the same time. It seems reasonable to consider both shots as causes of the death of the prisoner. Some judge such instances to be genuine cases of overdetermination.\(^{36}\) Since it at least is possible that such cases exist, it seems wise to look for a weaker version of the nonoverdetermination assumption. Rather than not allowing for overdetermination to occur at

---

\(^{35}\) We have seen earlier that Yablo’s exclusion assumption seems to rely on such strong version.

\(^{36}\) Of course, not everyone agrees. The general puzzle is this. A counterfactual analysis requires the following: if shot \( A \) is the cause of the death of the prisoner, then it must be true that if shot \( A \) had not occurred, the death would not have occurred (at the same time in the same fashion). Similarly, if shot \( B \) is the cause of the prisoner’s death, then had shot \( B \) not occurred, the death would not have occurred. But in the firing squad scenario, both counterfactuals are false. Thus, some argue, neither shot \( A \) nor shot \( B \) can be the cause of the prisoner’s death. See Schaffer (2003a) for a defense of both shots being a cause as well as overdeterminers.
all, it seems to be safer to merely claim that there is no systematic overdetermination. And, in fact, this is the version that most philosophers endorse.\textsuperscript{37}

But now (2*) cannot be rejected any longer based on the other assumptions. If (2*) is true, then it merely follows from the other assumptions that any particular mental event that causes a physical event in virtue of a mental property of the kind $k$ must be an overdetermining event. But on its own, this does not imply systematic overdetermination because there might be something atypical about such cases. Sometimes, the instantiation of a mental property of the kind $k$ might have a causal impact on physical events. Other times, it might not have such an impact. In other words, particular mental events that have a causal impact on physical events in virtue of mental properties of the kind $k$ and thereby overdetermine them might not be representative for their mental types. And as long they are not representative, such cases of mental causation might not be particularly bothersome.\textsuperscript{38}

Consider the following analogy: suppose it is true that in a Philosophy of Mind class only students with yellow socks got an A. This is not necessarily surprising, for maybe only one student got an A. But suppose twelve out of twenty students got an A and all of them wore yellow socks and no other students wore yellow socks. Most likely, this would make us wonder exactly because it would indicate some systematic correlation between the color of socks the students wear and the grades they got. Similarly, if every mental event that has a causal impact on physical events in virtue of its mental properties of the kind $k$ is an overdetermining event,

\textsuperscript{37} We have seen that Bennett endorses such a version. See Stoljar (2010, 217) for a way to spell out the exclusion principle in a way that it is sensitive to a weak version of the nonoverdetermination assumption: “For almost all events $e$ and $e^*$, if $e$ directly causes $e^*$, and if there is a property $F$ which is causally efficacious in $e$’s causing $e^*$, then there is no property $G$ such that (a) $G$ is distinct from $F$ and (b) $G$ is causally efficacious in $e$’s directly causing $e^*$.”

\textsuperscript{38} I wish to thank Gene Witmer for pointing this problem out to me.
there would be a systematic correlation between mental causation of a certain kind and
overdetermination. From there, it is not far from a clear violation of the weak
nonoverdetermination assumption.

While the existence of nontypical mental events that have a causal impact on physical
events is conceptually possible, they not only lead to an odd picture of reality but they also
threaten to undermine the explanatory power of causal explanations that involve them. To
illustrate this oddity and threat, consider the following example. Suppose that my thought that it
might rain today indeed had a causal impact on my behavior; it caused me to carry an umbrella.
If the completeness assumption is true, then there was also a physical event—in my brain, for
instance—that caused my behavior. To avoid systematic overdetermination, we now are
committed to saying that my thought that it might rain today is not representative of that kind of
mental event, i.e., that there are many mental events in general and thoughts in particular that do
not have any causal impact in virtue of being a thought or another mental property. But if
thoughts, even of that particular kind, usually do not have such causal impact, then it seems that
for whatever reason the thought had the causal impact it had, it was not because it was that
thought, nor was it because it was a mental event. And if this is true, the reference to my
thought that it might rain today was not an appropriate causal explanation for why I carry an
umbrella today.

Interestingly, this seems to be more of a problem for theories of causation that presuppose some sort of regularity
than for singularity theories of causation. In particular, reductionist theories of causation may not work if mental
events turn out to be too unrepresentative for their kinds. According to Jonathan Schaffer, for example, causation
reduces to history plus laws of nature; see Schaffer (2008, 5). Typically, laws of nature express nonaccidental
regularities or generalizations of some sort. It then seems that if mental events are too atypical, there might not be
any laws they can fall under. But if reductionism is right, that just means that they cannot have a causal impact. We
will get back to this in chapter 4. A second interesting point may be lingering here: if there are events that have a
causal impact but that are not representative of their kinds in the way described above, there may be events that are
strictly involved in the production of an event but that have no explanatory value. If such causes are possible, the
notion of an adequate cause, which we will introduce below, needs rethinking and perhaps modification.
Given the above considerations and problems regarding completeness and typicality, we may be well-advised to add an assumption not usually listed:

3. **Typicality:** Some mental events have a causal impact on physical events in virtue of mental properties of the kind $k$, and this causal impact is not atypical of events that have mental properties of the kind $k$.

In other words, such mental events are representative of events that have mental properties of the kind $k$ in that they tend to be strictly involved in the production of physical effects in virtue of properties of the kind $k$.40

The fourth assumption of the exclusion problem is motivated by our current scientific worldview. It is usually phrased in terms of causal sufficiency: every physical event has a sufficient physical cause, if it is caused at all. Bearing in mind that later we will discuss probabilistic causation and apply it to the exclusion problem, we cannot simply adopt this formulation. Some events may not have a sufficient physical cause but rather one that made their occurrence to some degree likely. Often this is put in terms of probability-raising: if an event is a probabilistic cause of another event, then it contributed to raising the probability of the occurrence of that event.41 To say that an event has a probabilistic causal impact on another event then is to say that (a) the event raised the probability of the occurrence of that type of event and that (b) the event is part of a causal chain that leads to the effect.

We might wonder whether it is necessary to assume that a probabilistic cause of an event raises the probability of that event to occur. Perhaps any impact on that event’s probability

---

40 It is worth noting that while the typicality assumption is usually not discussed in the context of the exclusion problem, it occasionally surfaces in the philosophy of science. For example, Harold Kincaid points out that at least some models in the social sciences rest on the assumption that effects are uniform, i.e., “the influence of a variable cannot vary according to context”; see Kincaid (2009, 738). This could be read as a version of the typicality assumption (3).

41 For a good overview of various probabilistic theories of causation, see Williamson (2009).
qualifies as having a causal impact, even if the cause lowers the probability. Consider, for example, a woman who gets pregnant in spite of having taken contraceptive pills. On the one hand, it seems natural to say that having taken the pills has had a causal impact on the likelihood of the pregnancy to occur. On the other hand, it seems wrong to say that taking the pills partly caused the pregnancy. We might be more inclined to say that such seemingly probability-lowering event really contributed to the prevention of an event, the pregnancy, albeit not sufficiently so; that is, it increased the likelihood of the pregnancy not to occur. Happily, we don’t need to take sides, for we can simply state more neutrally that if an event has a probabilistic causal impact on another event, it affected the probability of the effect to occur.

What is important, however, is that a complete probabilistic explanation does not require any reference to further events or other properties, even though probabilistic causes do not necessitate their effects. Since both kinds of causes, sufficient ones and probability-affecting ones, can share this feature, we may use it to define another term of art: ‘adequate cause.’ An adequate cause is one that, if it were cited in a causal explanation, would be all that one needs to cite as a causal factor operative at that time in order to provide a satisfactory explanation of why another event occurs. And a satisfactory explanation is one that does not require any reference

42 While Suppes (1970) defends the view of causes as probability-raisers, Rosen (1978, 605–606) argues that some causes may lower the probability of effects to occur. Rosen also discusses the pregnancy example in more detail.

43 The term ‘adequate cause’ has been used by other philosophers before, but it is important to notice that our definition may differ subtly from previous uses. For example, D. M. Armstrong uses the term to refer to causal chains that appropriately connect cause and effect. Similar to our notion of causal impact, this guarantees that the cause is strictly involved in the production of the effect. But unlike our notion, an Armstrongian adequate cause may not always yield a satisfactory explanation for why the effect occurs. It only may be one causal chain among many chains involved in the production of the effect; it may merely be a part of the overall cause. If so, it is that chain or that part that produces and explains the complexity of the effect. For example, human behavior often is very complex. An Armstrongian adequate cause needs to mirror this complexity in some way, or else it would not adequately explain the complexity of the behavior. In the case of mental causation, “the ‘adequate’ cause in the series of causes is what we call the mental event” (Armstrong 1968/1993, 119). But that is not necessarily to say that the adequate cause explains all features of the behavior, i.e., why it occurred in the exact way it did.
to further events or properties.\textsuperscript{44} Because the reference to an epistemic context is conditional, we can still consider the resulting concept a metaphysical one and use it to rephrase the fourth assumption in terms of adequate causes.

4. **Completeness:** If a physical event \( p_2 \) is caused at all, it has at least one adequate cause \( p_1 \) that is physical.

In other words, if \( p_2 \) is caused at all, there is some event \( p_1 \) such that \( p_1 \) is a cause of \( p_2 \) and \( p_1 \) is an adequate cause of \( p_2 \) in virtue of its physical property.

A few remarks on the relation between causal impact and adequate causes are in order. Any cause at the very least involves the property or properties in virtue of which the event has causal impact. In some cases, citing these properties and that event might yield a satisfactory causal explanation of why the effect occurred. However, this is not necessarily all that an adequate cause involves. Suppose that a pigeon is trained to peck at red things. When confronted with a scarlet object, the pigeon pecks. Presumably, the wavelength of the light that the object reflects causes the pigeon to peck. But this, in some sense, is not a satisfactory explanation; for it is not because the object is scarlet that the pigeon pecks but because it is red.\textsuperscript{45} While it is the wavelength corresponding to scarlet in virtue of which the event has its causal impact, the fact that that particular wavelength is in a certain range is causally relevant and needed for a

\textsuperscript{44} Interestingly, Malcolm’s early paper on the exclusion problem already discusses adequate and complete causal explanations, albeit not in the context of probabilistic causation. For him, an adequate explanation is one that references causal conditions sufficient for an effect; see Malcolm (1968, 56). Kim, too, thinks of complete explanations as those that specify “a sufficient set of causal conditions for the explanandum; see Kim (1989/1993, 251). We will need to address one problem at the beginning of Chapter 3, though. Which causal explanations provide satisfying answers to our causal questions may depend on the kind of question we asked as well as on the context in which the question is ask; that is, there may be pragmatic factors external to the situation that influence whether or not we consider a causal explanation satisfactory. But, as we will see, this is not an argument against our notion of an adequate cause but rather indicates that it is not properly defined yet. It is worth noting that the causal questions we ask may also specify how far back on the causal chain leading to the effect we are supposed to go when identifying the cause, and that timespan may be part of what determines whether or not a cause is adequate.

\textsuperscript{45} The example is taken from Yablo (1992b, 257).
satisfactory causal explanation. Hence, adequate causes might include properties in addition to those in virtue of which the cause has its impact. We should interpret the completeness assumption as stating that, if there are physical events that require such additional properties, those properties need to be physical as well.\textsuperscript{46}

Let’s recapitulate. So far, the exclusion problem assumes that sometimes mental events have some causal impact on physical events in virtue of their mental properties of the kind $k$. This causal impact is not atypical of events that have mental properties of the kind $k$. But notice that the completeness assumption, be it on its own or in combination with the other assumptions, does not exclude the possibility that there are mental events that have a causal impact on physical events in virtue of their mental properties, even if those mental properties are distinct from physical properties. It merely implies that whenever there is a mental event that has a causal impact on a physical event in virtue of one of its mental properties, there also is a physical event that is an adequate cause on its own. Regardless of whether the mental event is an adequate cause or not, as long as it has a causal impact in virtue of its mental properties, the physical effect is overdetermined.\textsuperscript{47} As mentioned above, single overdetermination may not seem to be as outrageous as systematic overdetermination, and yet that is what the typicality assumption seems to commit us to. If there is a single case of overdetermination involving mental events, then that is a typical case of a certain type of mental events, mental events that have a causal impact on physical events in virtue of mental properties of the kind $k$ typically overdetermine physical

\textsuperscript{46} This makes our version of the completeness assumption significantly stronger than those commonly discussed. We will address the question whether this might be too strong a claim later in Chapter 5.

\textsuperscript{47} This is only prima facie true. In the next chapter, we will encounter one response to the exclusion problem that attempts to undermine this alleged implication.
events.\textsuperscript{48} If we are inclined to think that nature most likely is not organized in such a way, we agree with the fifth assumption of the exclusion problem:

5. **Nonoverdetermination**: Events are not systematically overdetermined.

Earlier we saw that an adequate cause is one that, if it were cited in a causal explanation, would be all that one needed to cite as a causal factor operative at that time in order to provide a satisfactory explanation of why another event occurs. And a satisfactory explanation is one that does not require any reference to further events or properties. This alone does not guarantee that any explanation that involves more than one adequate cause is also a case of overdetermination. The phrase “more than one adequate cause” should not be read as “two or more adequate causes.” Rather, it should be interpreted as referring to at least one adequate cause and some other causal factor. This additional causal factor could be another adequate cause, but it also could be just another event that has some causal impact or another property that has some explanatory value. Such additional causal factors do not necessarily amount to a second adequate cause. But even if they did not amount to such, they would be explanatorily redundant; for if they were not, the physical cause could not be an adequate one. Hence, we need a further assumption that ties the concept of an adequate cause to the concept of overdetermination and eliminates other ways for an adequate cause to coexist with any further causal impact. This is done by the so-called exclusion assumption:

6. **Exclusion**: If an event has an adequate cause at time \( t \), then it does not have another cause at \( t \), unless it is overdetermined.

We now have established all the assumptions of the exclusion problem. As mentioned earlier, the problem is that (1)–(6) cannot all be true at the same time, i.e., at least one of the

\textsuperscript{48} The systematicity of overdetermination would threaten even more severely if the only events that genuinely overdetermine were events that have causal impact in virtue of properties of the kind \( k \).
assumptions has to be given up. Sometimes, the decision about which assumption to reject is based on which theory of mind an author favors. Other times, the decision is the result of weighing the plausibility of each assumption against the others and rejecting the one that seems the least promising or justified one. Depending on which assumption is rejected, the exclusion problem may in some cases turn into an argument for a particular theory of mind.

2.3 Generalizing the Exclusion Problem

In section 2.1.2, we mentioned that the exclusion problem is sometimes turned into an argument for epiphenomenalism, the view that mental events never have a causal impact on physical events in virtue of their mental properties. That is, assumptions (1) and (3) to (6) are used as premises for the conclusion that the causal impact assumption (2) is false. Of course, (2) is crucial not only for the exclusion problem but for all arguments dealing with mental causation. Critics worry, for example, that (2) would violate conservation laws in physics including those that the completeness assumption (4) may rely on. One type of response to these worries is to make them seem less pressing. Even if (2) implied such violation, so would any nonphysical event for which some causal impact is claimed in virtue of its nonphysical properties. Presumably, many causal explanations in the special sciences invoke such impact. In the claim “toxins cause cell membrane defects,” for example, a particular kind of chemical substance is cited as a cause of a particular kind of biological process or state. But while the properties being a toxin and being a cell membrane might relate to physical properties in important ways, they may not be identical with any of them. Given that we cite such nonphysical events and properties

49 Fodor (1981, 114ff.) and Dennett (1991, 35) use this point as a general argument against substance dualism as well as against anything nonphysical that allegedly causally impacts the physical.


51 See, for example, Burge (1993, 102), Baker (1993, 77), and Van Gulick (1992, 325).
in our causal explanations all the time, why then should we worry about mental causation in particular? Of course, this strategy is a double-edged sword, for we might as well just accept that the causal claims in the special sciences are problematic. But that just illustrates how relevant the problem is for the philosophy of science. Prima facie, this seems like a good reason to generalize the exclusion problem by substituting references to mental events and properties with references to nonphysical events and properties.

While the threat of exclusion may apply to some events and properties cited in the special sciences, the stakes are not necessarily the same in this context. If events never have a causal impact in virtue of their nonphysical properties, we may have to change our understanding of some of our epistemic institutions and practices, including the special sciences that invoke these properties in their causal explanations. Some of their explanations would be literally false, but they still might serve as abbreviated versions of more complex physical explanations. But if events never have a causal impact in virtue of mental properties such as having particular beliefs, desires, and intentions, we would have to revise the way we think about ourselves as human beings and, arguably, should strive to change some of our moral and social practices. Neither our traditional concept of agency as behavior caused by intentions nor the view of ourselves as morally responsible agents might survive such drastic causal exclusion of mental events and properties.

Caution is in order, though. While there are good reasons for trying to save some kinds of mental events and properties from exclusion, excluding others might not have any effect at all on

---

52 Kim (1997, 288) makes this very point: “[p]erhaps, the correct moral to be drawn is not that mental causation is not a problem for us, but rather that geological and biological causation, too, is a problem. If anyone were to retort, ‘Who needs an account of the efficacy of geological or physiological properties?’, I don’t believe that is the kind of philosophical attitude we should commend.”

53 Robb and Heil (2013, section 1.1) agree that at least some types of moral responsibility require mental causation.
our understanding of human nature and practices. Jaegwon Kim, for example, believes that qualia, the phenomenal qualities of experience, resist functionalization and face causal exclusion. But this mental residue, as he calls it, does not figure into our most promising physicalist accounts of intentions and other mental phenomena required for human agency. A plausible and useful generalized version of the exclusion problem thus should allow for the possibility that some kinds of nonphysical events and properties face exclusion whereas others do not.

Our version of the problem seems particularly well-suited in that respect, for, unlike Bennett’s and Yablo’s versions, our distinctness assumption (1) is not overly general. It does not claim that all mental events and properties are distinct from physical ones but merely that there is an important kind of mental property that is distinct from any physical property. (2) then assures that it is in virtue of mental properties of such kind that some mental events have a causal impact on physical events. Generalizing the assumptions to nonphysical events and properties would then not yield unduly strong assumptions, for they would merely state that some, but not necessarily all, nonphysical events have a causal impact in virtue of nonphysical properties of an important kind. This may include some of the events and properties cited by the special sciences, but it does not have to. It could be the case that the only nonphysical events that have a causal impact are those that have an impact in virtue of a special kind of mental property. Using our version of the exclusion problem as the basis for the generalization thus has the virtue of potentially applying to the special sciences without committing us to the claim that it does.

Replacing all references to mental events and properties with references to nonphysical events and properties yields the following generalized version of the exclusion problem:

54 Kim (2005), especially pp. 161–174. This is not a new idea. Armstrong (1968/1993, 234) already suggested that perceptions that are decoupled from beliefs may not have the same or even any causal impact on behavior. The beliefs do the causal work, but they can be accounted for in terms of his central-state materialism.
1. **Distinctness:** There is an important kind $k$ of nonphysical property, such that, if a nonphysical property is of that kind $k$, it is distinct from any physical property.\(^{55}\)

2. **Causal Impact:** Sometimes, nonphysical events have a causal impact on physical events in virtue of their nonphysical properties of the kind $k$.

3. **Typicality:** Some nonphysical events have a causal impact on physical events in virtue of nonphysical properties of the kind $k$, and this causal impact is not atypical of events that have nonphysical properties of the kind $k$.

4. **Completeness:** If a physical event $p_2$ is caused at all, it has at least one adequate cause $p_1$ that is physical.

5. **Nonoverdetermination:** Events are not systematically overdetermined.

6. **Exclusion:** If an event has an adequate cause at time $t$, then it does not have another cause at $t$, unless it is overdetermined.\(^{56}\)

Of course, this is not the only way to generalize the problem. In fact, we may worry that contrasting the physical with the nonphysical may not be very informative, for it is not clear what we are referring to by ‘the physical.’ According to Hempel’s Dilemma, we refer either to properties currently cited in physics or to properties cited in a future completed physics.\(^{57}\) On the one hand, a future physics may qualify as being complete exactly because it includes all properties that in fact are physical, assuming that *being physical* is a natural kind property. Since we don’t know which properties such future physics will include, we don’t know what we are referring to when using the term ‘physical.’ Consequently, we do not know which properties are

---

\(^{55}\) The motivation for contrasting nonphysical with physical properties is explained further below. Thus contrasted, it is analytically true that nonphysical properties are not physical. But whether or not a particular property is in fact nonphysical is an empirical question. When discussing particular properties, the nonphysical property in question is one that is treated by one or several of the special sciences as a serious contender for a nonphysical property.

\(^{56}\) From now on, numbered assumptions in the text will refer to these generalized assumptions, unless otherwise noted.

\(^{57}\) For a discussion of Hempel’s Dilemma in the context of the question how to best understand physicalism, see Melnyk (1997, 623) and Gillett and Witmer (2001, 302–303). See also Crane and Mellor (1990) for an argument against the suggested contrast. They argue that there is no good reason for excluding the mental, and perhaps by extension the nonphysical, from the physical realm. But then physicalism may be vacuous if understood as a theory of the mind, and perhaps by extension of the nonphysical in general.
threatened by exclusion. On the other hand, our current physics is likely subject to change and not complete in the above sense. But if it gets the extension of ‘physical’ wrong at least in some cases, the exclusion problem would threaten the exclusion of some events and properties that do not face such a threat metaphysically.

Alternatively, we might consider a contrast between fundamental and higher-level events and properties.\(^58\) However, the main problem with that approach is that it commits us to the existence of hierarchically stratified levels as well as to the assumption that there is a fundamental level in the first place. It might not be unreasonable to assume that reality is thus organized.\(^59\) It might even be possible to provide criteria for what constitutes a particular level or how to individuate levels. But there are good reasons to remain critical of simply presupposing that a fundamental level exists. For example, Jonathan Schaffer claims that there is no scientific evidence that actually supports this assumption, even though some scientists endorse it as a presupposition or conjecture.\(^60\) Furthermore, rejecting the assumption might allow for “greater explanatory scope” in that the workings of every single entity is explicable in terms of the workings of its parts, and infinite division yields a more elegant hypothesis in that the pattern of division embraces the whole structure of nature.”\(^61\) Put more simply: not endorsing the assumption that there is a fundamental level might yield better theories. What seems most important in the context of the exclusion problem, however, is that the notion of a fundamental

\(^{58}\) We could also contrast physical events with events that supervene on them. Humphreys (2009, 642), for example, suggests that the exclusion argument, and by extension the exclusion problem, can be applied to any kind of event that supervenes upon physical events.

\(^{59}\) John Tienson, for example, claims nonreductive physicalists must endorse such a level ontology. However, he argues that, while there is a genuine problem of mental causation in such ontology, that problem is not “an instance of a general exclusion problem”; Tienson (2002, 100).

\(^{60}\) See Schaffer (2003b, 502ff.).

\(^{61}\) Ibid., p. 500f.
level might carry the connotation of ontological or perhaps even causal priority. Or, put more colloquially, the fundamental level is where the real causal action takes place. Such suggestive priority might further bias us in unwanted ways regarding the possibility of a causal impact of nonphysical events, especially in the context of an already very strong completeness assumption (4).

Thus, it seems best to avoid using the contrast between fundamental and higher-level events and properties and opt for contrasting physical and nonphysical events and properties instead. While that carries the risk of it remaining somewhat unclear which particular events and properties are threatened by exclusion metaphysically, it may suffice to treat those events as serious contenders that, according to the special sciences, are said to have a causal impact in virtue of their nonphysical properties. Whether or not these properties are in fact nonphysical remains to be seen, but currently the special sciences offer the best contenders for such properties.

2.4 Chapter Conclusion

So far, we have accomplished three things. First, in order to develop a suitable version of the exclusion problem, we needed to adopt a convention regarding the ontological entities between which we posit relations. We have opted for treating events as nonrepeatable particulars that have properties at a time. This should be understood as a way of talking about events, properties, and causation rather than as endorsing a particular theory of events. What matters is that we can specify the properties in virtue of which events have causal impact, not whether events in fact are the bearers of properties or partly constituted by them. Even though our convention is relatively neutral with respect to particular theories of events such as Davidson’s and Kim’s, it does treat events as the appropriate causal relata. We are thus bound to ignore alternative views according to which the relata are aspects, facts, features, situations, states of
affairs, or tropes. Whether we can modify our framework to incorporate such views goes beyond the scope of this dissertation, as does the question whether these views yield responses to the exclusion problem that are distinct from those discussed here. Answering that question could be part of a larger project, namely, a comprehensive primer on the exclusion problem.

Second, we have developed a version of the exclusion problem that is more precise than most versions discussed in the literature. References to rather vague concepts such as being causally relevant have been replaced by references to more clearly defined concepts such as having a causal impact and being an adequate cause. While some versions of the problem boldly assume that mental properties are generally distinct from physical properties, our version is sensitive to and compatible with the possibility that some mental properties are identical with physical properties whereas others are not. And, as a result of trying to be more precise, the need for an additional assumption became apparent, namely the typicality assumption, which helps illuminating some of the inferences needed for the problem to get off the ground.

Third, we have generalized the exclusion problem. If the problem indeed extends to some nonphysical, nonmental events and properties, it threatens the status of some explanations in the special sciences. Not only does this highlight the gravity of the problem, it also allows us to contribute to arguments made in metaphysics and the philosophy of science regarding the proper scope of the problem. In return, we can utilize some of those arguments for bringing our concepts of causation and causal explanation into sharper focus. Generalizing the problem thus naturally sets the stage for merging debates in metaphysics, the philosophy of mind, and the philosophy of science.
The goal of this chapter is twofold. First, we want to test whether the exclusion problem is affected by particular theories of causation. Perhaps we are more inclined to endorse or reject one or several of the assumptions when we presuppose a particular theory of causation; or perhaps the problem appears to be more or less severe depending on which theory of causation we presuppose. Second, we want to set the stage for assessing some of the standard responses to the problem in Chapter 4. If theories of causation indeed affect the dynamic of the exclusion problem, they may also affect how successful some of the responses to the problem are, both in virtue of affecting the exclusion problem itself and in virtue of changing the meaning of claims involved in these responses.

However, covering all theories of causation would be too ambitious a project. Over the past two and a half millennia, many different theories of causation, and even more modifications and versions of these theories, have been developed. By the same token, in section 1.1, we complained that the assumption that the exclusion problem remains the same no matter what theory of causation we presuppose has not been justified in any systematic manner. We granted, though, that a potential impact of a few particular theories such as the counterfactual theory of causation has been explored. To decide whether or not the above assumption can be justified, we thus need to look at more than just two or three theories, or else we may accidentally miss the theories that are most promising regarding a potential impact. So, while we need to cover enough theories for our inquiry to be systematic, we also need to exclude some theories and versions for the project to be feasible. An easy way to achieve this is to use standard anthologies such as Beebee, Hitchcock, and Menzies (2009) and Sosa and Tooley (1993b) to determine which
theories of causation are still relevant in contemporary metaphysics and which may be worth applying to the exclusion problem. But rather than just using the most prominent ones, we may also include a few lesser-known theories because they might be interestingly different. The resulting set of theories may be considered representative enough for our project to be systematic, but it will not be comprehensive.

While many theories of causation started out relatively simple, most of them have been developed into fairly elaborate and complex theories. Naturally, we will not be able to do justice to that complexity, and, in a way, we do not need to. Since our goal is to show how such theories may affect the exclusion problem and the responses to it, we need to explain them enough for that impact to surface, but not beyond. It may well be that a closer look would reveal further trouble for the problem or the responses; but the goal of this dissertation is more limited in scope, namely, to put pressure on the silent assumption that theories of causation do not matter for the exclusion problem. If we can establish that they do matter without further elaboration, we should discuss more theories of causation and assess more responses. But before we do so, it will be useful to briefly look at the relation between causation and causal explanations.

3.2 Causal Explanations

Except for the distinctness assumption (1), all assumptions of the exclusion problem include causal assertions. But the assertions involve different types of causal explanation. The nonoverdetermination assumption (5) and the exclusion assumption (6) reference an adequate cause that would satisfactorily explain the occurrence of the effect.\footnote{At first glance, it may not seem that the nonoverdetermination assumption (5) is doing this. But when we established the assumption, we explained what it means for events not to be systematically overdetermined, namely, that they do not typically have more than one adequate cause, be it another adequate cause or just an additional causal factor. Thus understood, (5) uses the notion of an adequate cause.} Whenever we explain the occurrence of an event by referring to a cause of it, let us call such explanations ‘basic causal
explanations.’ But the causal impact assumption (2), the typicality assumption (3), and the completeness assumption (4) refer to properties that explain why the cause-effect relation holds in first place, namely, in virtue of properties had by the cause. Let us call explanations of this type ‘causal grounding explanations.’ In the context of the exclusion problem as a whole, both types of explanations may also be understood contrastively. For example, we may take (2) to emphasize that sometimes nonphysical events have a causal impact on physical events in virtue of their nonphysical properties of the kind $k$ rather than in virtue of some properties not of the kind $k$—whether they are themselves physical or nonphysical. Let us call explanations that specify alternative or contending causes or properties ‘contrastive causal explanations.’ It is important to distinguish these types of explanations because they have different explananda and can be used to answer different kinds of causal question.\(^2\)

In everyday life, we tend not to clearly articulate and carefully distinguish between different kinds of causal question. Often, it is the epistemic interests or the conversational context that determine what we are really asking and what type of causal explanation we expect and would accept as satisfactory. If we wish to give an honest answer to a causal question or an explanation of a causal problem, we need to be sensitive to that context. Interestingly, Paul Grice’s conversational maxims seem to apply and may serve to illustrate the point. For example, the maxim of relation prescribes that our assertions be relevant to the purpose of the conversation.\(^3\) In causal contexts, that may limit how far back in the causal chain leading to the effect in question a cited cause can be. If we were asked why the friend mentioned in section

\(^2\) Lipton (2009, 624) points out that not all causes are explanatory with respect to all explananda. For example, our love of Broadway shows may causally explain why we went out to see a show, but it would not explain why we went to see *Jumpers* rather than *Candide*. Merely citing a cause in response to a contrastive question often does not yield a satisfactory explanation.

\(^3\) See Grice (1989, 27).
2.1.1 is in the hospital, responding with a reference to the Big Bang would violate the maxim of relation, even if the Big Bang indeed were an adequate physical cause of our friend’s being in the hospital. But in an ordinary conversation, we would expect the cause cited to be either a disease or an accident, such as her falling down the stairs and breaking her ankle. The maxims of quantity, to give another example, ask us to make our contributions as informative as is required but not more informative than is required by the context. If we responded to the question why our friend is in the hospital by giving a microphysical description of the event of her falling down the stairs, this description may again cite an adequate cause, at least if the completeness assumption (4) is true; but even though we avoid the mistake of going too far back in the causal chain, we now violate the second maxim of quantity as well as the maxim of relation by providing too much detail. Why is this important?

First, it sometimes may seem as if the causal explanandum—say a particular physical event—in different domains of discourse is the same. But attention to the conversational context might reveal that to be illusory. If nonphysical events sometimes have a causal impact on physical events in virtue of their nonphysical properties of the kind \(k\) rather than in virtue of some properties not of the kind \(k\)—physical or nonphysical—, we need to tread carefully whenever the explanandum seems the same because properties of the kind \(k\) may be cited in explanations of different special sciences than properties not of the kind \(k\). That is, they may be cited in different domains of discourse. This may be particularly important when thinking about the nonoverdetermination assumption (5) and the exclusion assumption (6), for both presuppose a common explanandum, the physical event \(p_2\).

---

Second, if in fact there is reason for thinking that the explanandum is never or rarely ever exactly the same when we cite physical causes and events that have a causal impact in virtue of their nonphysical properties, we may wonder whether physical causes really are adequate, as posited by the completeness assumption (4), at least for those aspects of the physical event that the nonphysical cause is meant to explain. Alternatively, we may wonder whether the nature of the causal relation involved in each explanation is the same because we may take the lack of matching or overlapping causal explananda as a reason to endorse causal pluralism.

Finally, if the conversational context is as important as suggested, we may wonder that whether or not a causal explanation is satisfactory depends too much on our linguistic and, by extension, normative practices. More specifically, we may worry that this might turn the concept of an adequate cause, which is used by several assumptions, into something epistemic. As a reminder, an adequate cause is an event that, if it were cited in a causal explanation, would be all that one needs to cite as a causal factor operative at the time in order to provide a satisfactory explanation of why another event occurs. And a satisfactory explanation is one that does not require any reference to further events or properties. But if the conversational context rather than the causal situation determines whether or not references to further events or properties are required, how then can an adequate cause be something ontologically robust, i.e., be a relatum in a mind-independent causal relation?

---

5 Different theories of events may require us to phrase the issue here differently. As we have suggested in section 2.1.2, a fine-grained Kimean account may construe every aspect of a physical event as a distinct event, for it construes events as exemplifications of a property by an object at a time. The question then is whether adequate physical causes would satisfactorily explain the occurrence of those events that nonphysical events are meant to explain.

6 Menzies (2009, 342) suggests that this is a serious worry, albeit not specifically in the context of satisfactory causal explanations and not for the reasons given here. He argues that “the concept of causation has essential elements that are contrastive, normative, and context-sensitive in character, and that these elements do not seat easily with the conception of causation as a natural relation.”
The worry may be alleviated in the following way. The reason why we recommended paying attention to the context is that communication is fragile. Grice’s maxims are meant to increase the likelihood of a successful communication. Communication can fail, for example, when too many assumptions are not explicated or when assertions, arguments, and questions are not articulated clearly or phrased properly. But arguably, physics and the special sciences attempt to articulate their background assumptions as explicitly as they can, certainly much more so than we ordinarily do. They also utilize definitions, formalisms, mathematical models, and such to maximize the clarity of their assertions. To the extent that they do this, they fix what exact causal question we are asking or should be interested in and what types of causal explanations would count as satisfactory. But once the different types of explanations and questions are properly distinguished and recognized, there are objective facts in the world that determine what an adequate cause of a given event is.\footnote{Maslen, Horgan, and Daly (2009) may deny this. They argue that, within a counterfactual analysis, it always depends on the explanatory context which possible worlds will qualify as relevant and which do not. Section 3.3.2 explains what is meant by this.} Hence, while we should be cautious regarding the types of causal explanation given and the types of causal question asked, we should not worry too much about the concept of an adequate cause and explanations that reference it.

### 3.3 Theories of Causation

Theories of causation are theories about the nature of causal relata and relations.\footnote{The following sections will discuss various accounts of what the conditions are for a causal relation to hold between events. It is not clear in what sense exactly the mere listing of such conditions amounts to a theory of causation. The notion of a theory is thus not a very rich one. It may not mean anything more than there being an account or view of something. Yet, it is customary in the literature to talk of theories of causation. We will simply adopt this custom without discussing what constitutes a theory.} Since we have settled for events as the appropriate causal relata as a matter of terminological convention in Chapter 2, we will focus on the nature of causal relations in this chapter. To
motivate the order in which we will discuss the theories of causation, it will be useful to briefly
distinguish between causal realism and antirealism, causal reductionism and antireductionism, as
well as causal monism and pluralism.\(^9\)

Causal realism assumes that there are genuine, mind-independent causal relations in
nature. Such causal relations would hold between events even if no beings with a mind were
around referencing those relations in their explanations.\(^10\) Causal antirealism in its boldest form,
eliminativism, denies that causation is a genuine, mind-independent relation. Eliminativists
typically point out that our best epistemic practices and institutions, such as physics and the
special sciences, do not rely on or need not make use of the concept of causation and take that as
one reason to eliminate it from our ontology altogether.

Causal reductionism views causation as a complex phenomenon that can be analyzed in
terms of its constituent parts. While reductionism does not entail antirealism, some reductionists
may be content to eliminate the term ‘causation’ from our explanations, for it can be replaced by
whatever they think the concept and phenomenon of causation reduces to and is identical with.
But to explain causal relations by reference to other entities or relations in our ontology is not the
same as denying the reality of causal relations. Thus, causal reductionists are not eliminativists.
Most reductionist acknowledge the central role that causation plays in many of our explanations
and thus would be reluctant to eliminate the term from our explanations even if they consider it
successfully reduced. Some types of nonreductionism assert that the concept of causation is
simple in the sense that it cannot be broken down into further concepts. Nonreductive realists

---

\(^9\) The distinctions introduced in this paragraph are loosely based on Humphrey (2009, 632–336).

\(^10\) See Psillos (2009, 133f.).
may see this as one reason to view causation as essential to our explanatory practices and epistemic interests.¹¹

Causal monism asserts that there is only one type of causal relation, but causal pluralism denies this. Unlike regular causal realists, pluralists assert that there is more than one type of causal relation. Often, these types are domain specific. For example, mental causation may be one type of causation, and physical causation may be another. In theory, causal pluralists can mix and mingle all previous stances. Some type of causal relation may be reductive whereas others are not. Some type of causal relation may be real whereas others may just have an epistemic or pragmatic value. Other types of causal relation may not hold up to scrutiny and should be abandoned altogether. The pluralist stance will partly depend on the number of causal relations posited and the function assigned to each of them.

We will first discuss theories of causation that are usually considered instances of reductive realism. These are regularity theories, counterfactual analyses, INUS conditions, probabilistic theories, transference theories, and single-difference theories. We will then look at causal pluralism, which, depending on how it is fleshed out, could be reductive; but it does not have to be. Finally, we will have a brief look at eliminativism as a kind of antirealism.

3.3.1 Regularity Theories

According to Stathis Psillos, regularity theories are reductive in that they try to explain causation in terms of spatiotemporal contiguity, succession, and regularity:

\[ c \text{ causes } e \text{ if and only if } \]
\[ \text{i. } c \text{ is spatiotemporally contiguous to } e; \]
\[ \text{ii. } e \text{ succeeds } c \text{ in time; and} \]

¹¹ There are many types of reduction, so this paragraph simplifies things considerably. It still might suffice to hint at some key differences between various theories of causation. Humphrey (2009, 632–336) briefly walks through some of the standard views of reduction in the context of causation. The reader may recruit his text for further details.
iii. all events of type C, i.e., events that are like c in some relevant respect, are regularly followed by, or are constantly conjoined with, events of type E, i.e., events that are like e in some relevant respect.\textsuperscript{12}

A naïve regularity theory\textsuperscript{13} would treat (i), (ii), and (iii) as individually necessary and jointly sufficient conditions for causation. If they are necessary, notice that (i) may put pressure on the causal impact assumption (2), at least for some nonphysical properties. It is not clear, for example, whether mental events such as having beliefs and desires are spatially extended and locatable and in what sense they can be spatiotemporally, rather than merely temporally, contiguous with other events.\textsuperscript{14} In fact, some nonmental events may face similar problems. For example, consider the closing down of a business. On the one hand, it is locatable in the sense that it is registered in a particular city, state, or country. One the other hand, some businesses might not be tied to a particular building, for the event mostly has to do with revoking ownership. We might say that the closing down takes place wherever an official processes the legal documents; or we might say that this just means that, while the event has a location, its boundaries are fuzzy. But either response seems somewhat arbitrary.\textsuperscript{15}

\textsuperscript{12} See Psillos (2009, 131). While Psillos phrases (i) and (ii) as listed, (iii) is slightly modified, so this is not a direct quote. On a side note, discussions of regularity theories usually start with Hume. Often, such theories are called ‘Humean.’ However, it is utterly unclear what exactly Hume’s metaphysical view on causation is. Psillos’ preliminary definition is reductionist as it is. Sklar (2009) reads Hume specifically as an eliminativist. And according to Garrett (2009, 73–74), projectivists hold that, while a felt necessity is merely a feeling of the determination of the mind that then gets projected into the world, the necessity is nonetheless an essential part of the concept of causation, which charges the concept of causation epistemically. But it is not clear whether Hume endorsed any of the above views. From the fact that we merely observe regularities but generate an impression of necessity in our mind, it does not follow that causal relations in nature do not possess features similar to necessity or sufficiency; see Garrett (2009, 89). To avoid a possibly misleading misnomer, we will avoid calling regularity theories ‘Humean.’

\textsuperscript{13} Armstrong (1983, 11) calls it that.

\textsuperscript{14} Of course, this echoes Descartes’ view that mental substances are spatially unextended, as mentioned in footnote 6 in section 1.1. However, the issue may arise even if one does not endorse a Cartesian substance dualism.

\textsuperscript{15} I thank Gene Witmer for this point and example.
Furthermore, a naïve regularity theory would deny that the regularity expressed in (iii) needs further explanation, for example in the form of laws of nature that might explain why the regularities hold in the first place.\textsuperscript{16} Of course, this denial does not suggest that there are no laws of nature. But such laws would merely express the regularities rather than ground and explain them.\textsuperscript{17} Typically, the regularity referenced in (iii) is thought to be one without exception. If there is one event of type $C$ that is not followed by an event of type $E$, then $c$ is not the cause of $e$. This has interesting consequences for the typicality assumption (3) and the nonoverdetermination assumption (5).

Recall that Stephen Yablo’s version of the exclusion argument in section 2.1.2 offered a very strong exclusion assumption, one that merges exclusion with nonoverdetermination. His version does not allow for even a single case of overdetermination.\textsuperscript{18} Many philosophers, including Karen Bennett, think that that is too strong; isolated instances of overdetermination may be conceptually and perhaps metaphysically possible. But in the context of the exclusion problem, only systematically overdetermined effects should concern us, which is what we tried to capture in (5). In part, this was what motivated the introduction of the typicality assumption (3). Nonphysical events that have a causal impact on physical events in virtue of the nonphysical properties would be threatened by exclusion qua (5) only if they are representative of their kind $k$, i.e., only if they are typically strictly involved in the production of effects in virtue of properties of the kind $k$; otherwise, their causal impact would not imply systematic


\textsuperscript{17} See Armstrong (1983, 11).

\textsuperscript{18} He is not alone. Bunz (1979, 146) denies that some of the alleged standard cases are genuine overdetermination cases and doubts whether there could be some. See Kim (1989/1993, 250–254) for the claim that they are improbable. See Schaffer (2003a) for a criticism of Bunzl and Kim.
overdetermination. We thus restricted the scope of the exclusion problem to nonphysical events that have such representative causal impact. This restriction, however, allows for there being atypical nonphysical events that only sometimes overdetermine effects in virtue of their nonphysical properties. After all, the distinctness assumption (1) merely states that, if a nonphysical property is of the kind $k$, then it is distinct from any physical property. Since this is merely an implication and not a biconditional, (1) allows for nonphysical properties that are not of the kind $k$ and yet distinct from physical properties. Some events may have a causal impact in virtue of such nonphysical properties of the non-$k$ kind, say of the kind $l$, without that impact being typical. But if the naïve regularity theory is correct, such atypical causes could not exist. The assertion that $c$ causes $e$ in virtue of its nonphysical property of the kind $l$ in cases where typically events of type $C_l$ are not followed by events of type $E$ would imply that (iii) is false and, consequently, that $c$ is not the cause of $e$. So, while (3) on its own is compatible with there being atypical nonphysical causes that overdetermine events that have adequate physical causes, (3) in combination with the naïve regularity theory captures all events that may have a causal impact in virtue of their nonphysical properties. In fact, we could just omit (3) for the naïve regularity theory would guarantee typicality. By the same token, the falsity of any theory of causation that assumes or implies exceptionless regularities threatens the typicality assumption (3) of the exclusion problem.

If we accept the naïve regularity theory, the exclusion problem thus seems particularly severe, for all nonphysical events would be threatened by causal exclusion. The problem could be turned into a deductive argument, for instance by using the causal impact assumption (2) to set up a reductio ad absurdum that results in the negation of (2). We may try to cast doubt on the argument’s validity by driving a wedge between the typicality of the causal impact and the
systematicity of the overdetermination. Perhaps instances of nonphysical events that have a causal impact are typical but rare enough not to amount to systematic overdetermination. But that seems confused. Consider a standard overdetermination case such as two assassins that attempt to kill the same person. They pull the trigger of their respective rifles, both bullets hit the victim in a deadly manner, and as a result the person dies. However, if the bullets hit the victim at different times, this may not qualify as genuine overdetermination, for the earlier bullet may explain why the victim died an instant earlier. Similarly, if the bullets hit the victim at different places of his or her body, each shot may contribute to the victim’s death in different ways, which may not amount to overdetermination either.\textsuperscript{19} For the assassination to be a genuine case of overdetermination, both bullets need to hit the victim at a time in such a way that their effects contribute to the death of the victim in the exact same manner. Such overdetermination cases may be very rare, but that is not what makes them lack systematicity. Rather, it is the fact that it is a mere coincidence that they happen to have the exact same effect,\textsuperscript{20} for it is in virtue of the circumstances rather than in virtue of the properties of the causes that they overdetermine the death of the victim. Increasing the prevalence of overdetermination by frequently recreating similar circumstances would not lead to systematic overdetermination. If, by contrast, we found that whenever two assassins try to shoot a person the bullets hit the victim at the same time in the same way, i.e., if there is little to no variability in the effects in spite of varying circumstances, that would be surprising exactly because it suggests that there is something about the causes that

\textsuperscript{19} We are thus describing what Paul (2009, 178) calls “fine-grained overdetermination,” “where a difference in properties amounts to a difference in events.” We thus need to describe the assassination case in such a way that the effect occurs in precisely the same way regardless of which bullet hits the victim. Requiring such fine-grained overdetermination is much more natural in a Kimean account of events than in a Davidsonian one; see section 2.1.1 for details.

\textsuperscript{20} This point is inspired by Sider (2003, 722), even though he discusses coincidence and systematic overdetermination in a different context.
explains the overdetermination. That is what makes the overdetermination systematic, and that is what nonphysical causation would look like according to the typicality assumption (3) and the completeness assumption (4). Every effect that has a nonphysical cause also has an adequate physical cause. Consequently, all nonphysical causes that have a causal impact in virtue of their nonphysical properties are threatened by exclusion by way of the nonoverdetermination assumption (5).

However, it seems difficult to support (5) if the naïve regularity is true. So perhaps we should not question the argument’s validity but deny its soundness instead, for instance by rejecting (5). Armstrong and others have argued that the theory is bound to treat many regularities as laws of nature and as causal relations that intuitively are not. For instance, there seems to be a correlation between per capita cheese consumption and the number of people who die by becoming tangled in their bedsheets. But it would be odd if there were an actual law these types of events fell under. Most likely, the correlation is spurious and purely coincidental. If Armstrong is right and the naïve regularity theory is ontologically profligate, it is difficult to see what reasons we could have for endorsing the nonoverdetermination assumption (5).

Systematic overdetermination by way of nonphysical events may yield an odd picture of reality, but it is not odder than an overabundance of laws of nature and causal relations. In fact, if nothing grounds regularities and if laws of nature supervene on those regularities, it is rather unclear what is odd about systematic overdetermination to begin with.

Under the naïve regularity theory, an event e’s being overdetermined implies that there are at least two events, c and d, that are spatiotemporally contiguous with e, that precede e, and

---

21 The example merely serves to illustrate the problem, for the correlation only covers 94.71 percent of the data entries. To satisfy (iii), we would need a perfect correlation. The example is drawn from the amusing collection of spurious correlations offered by Vigen (2015).
for which there are laws they, together with \(e\), fall under. There “is a law \([L_1]\) such that events of the type \(C\) (where \(c\) is one of them) are followed by events of type \(E\) (where \(e\) is one of them)”\(^{22}\) and there is a law \(L_2\) such that events of the type \(D\) (where \(d\) is one of them) are followed by events of type \(E\) (where \(e\) is one of them). But what exactly is odd about there always being two causal laws \(L_1\) and \(L_2\)? One answer may be this: especially if one of these causes, \(c\), and by extension one of these causal laws, \(L_1\), is ontologically privileged by way of the completeness assumption (4), \(d\) and \(L_2\) seem epistemically redundant,\(^{23}\) for they don’t explain anything that is not already explained by \(c\) and \(L_1\). We may respond by applying Ockham’s Razor and deny that \(L_2\) is a genuine causal law. Or we may accept that it is a law that expresses said regularity but deny it is also a causal law. The regularity expressed by \(L_2\) might supervene on the one expressed by \(L_1\). If causes are events that bring about or produce effects, as we have suggested when we defined ‘causal impact,’ this solution may seem appealing at first. But the naïve regularity theory lacks the resources to draw the needed distinction in this case. While it can distinguish between laws that are causal and those that are not, it can do so only when the events in question are not spatiotemporally contiguous or when \(e\) does not succeed \(d\). Since the relation between \(d\) and \(e\) satisfies (i), (ii), and (iii), \(L_2\) qualifies as a causal law. We are left with overdetermination, or so it seems. However, the regularities that \(L_1\) and \(L_2\) express are not grounded in anything else. In some sense, citing these laws does not explain the occurrence of \(e\), at least not in a way that causal production or necessitation would. But then it remains unclear why overdetermination is problematic, even when it is systematic. All it means is that two distinct events of certain types regularly precede effects of a certain type.

---

\(^{22}\) Psillos (2009, 146).

\(^{23}\) Sider (2003) discusses this epistemic argument against overdetermination in addition to a metaphysical objection and a coincidence objection.
To summarize: applying the naïve regularity theory to the exclusion problem has two interesting results. First, all nonphysical events that allegedly have a causal impact in virtue of their nonphysical properties are threatened by exclusion. If overdetermination is a problem for nonphysical events, it always is. Second, what exactly is problematic about overdetermination remains a mystery. When assessing which assumption of the exclusion problem should be rejected, assuming that (1) to (6) are in fact inconsistent, the nonoverdetermination assumption (5) may thus seem like the best candidate.

To avoid an overly profligate naïve regularity theory, we can either ground (iii) in something else or we can add further necessary conditions to (i) to (iii). Pursuing the first strategy, we may offer a thick notion of the laws of nature that might allow us to distinguish between accidental and nonaccidental regularities. The Ramsey-Lewis view on laws of nature may provide such a thick notion. According to this view, regularities can be laws only if properly embedded as axioms in a deductive system, i.e., a particular kind of scientific theory, that balances simplicity and strength best. The best candidates for nonaccidental regularities, then, are those that can be subsumed under a law that improves the quality of the deductive system. Within such systems, we may also have the resources to distinguish between laws that are causal and those that are not, to privilege some laws over others, and to thus support the nonoverdetermination assumption (5). Still, it seems possible that regularities that lead to systematic overdetermination may be among those that improve the quality of the deductive

---

24 Strawson (1987, 266) agrees that regularities need to be grounded in something, or else “the regular succession [would be] completely astonishing [...]: all those totally unprincipled physical objects on the loose out there, interacting in seemingly perfectly law-governed ways—and for no reason at all.” Psillos (2009, 134) points out that whatever grounds regularities would itself be an “unexplained explainer” and thus not solve the problem.

25 See Psillos (2009, 142) for a brief sketch and Ramsey (1928/1978) and Lewis (1973b) for the fully developed view. Psillos also addresses the point that the Ramsey-Lewis view charges the notion of laws of nature epistemically.
system. Since regularities and laws are grounded by the occurrence of events rather than vice versa, it seems rather unclear what principled reason we could have to reject that possibility. To the extent that we do not have such reason, we should not be particularly invested in the nonoverdetermination assumption (5).

Alternatively, we may declare laws of nature to be fundamental, not further analyzable items in our ontology that would allow us to distinguish between accidental and nonaccidental or causal and noncausal types of regularity. While this may seem rather ad hoc, for we could have postulated two distinct types of regularity instead, it would at least lend more plausibility to the completeness assumption (4) if the only laws that do not supervene on other laws happen to be physical laws. Pursuing the second strategy, we may add further necessary criteria to the causal biconditional by fusing the regularity theory with other theories of causation, such as the counterfactual analysis of causation.

3.3.2 Counterfactual Theories

The best-known counterfactual theory of causation is the one developed by David Lewis. It shares two important features with the regularity theory discussed in section 3.3.1. First, it is reductive because it explains causation in terms of another relation, in this case counterfactual dependence rather than regularity. Second, it is a realist theory, for the counterfactual dependence is based on facts about and relations between possible worlds, which Lewis construes as entities that are as real as our actual world. Lewis himself presents his

---

26 See, for example, Maudlin (2007, 17–18).

27 Not all counterfactual theories are reductive, according to Paul (2009, 164).

28 See Lewis (1986) for a defense of this modal realism. However, the counterfactual analysis of causation does not rely on this type of realism. Plantinga (1976), for example, construes possible worlds as abstract states of affairs that exist but do not obtain.
counterfactual analysis as an alternative to regularity theories of causation, even to those that are more advanced than the naïve regularity theory we discussed in the previous section.²⁹ So he is not interested in grounding (iii) or in adding further necessary conditions to (i), (ii), and (iii) of the regularity theory. However, Helen Beebee has convincingly argued that his counterfactual analysis ultimately rests on an analysis of counterfactuals and laws of nature which in turn depend on regularities. It thus may be best understood as a sophisticated version of the regularity theory.³⁰ Let us now look at the analysis itself.

Recall our friend from sections 2.1.1 and 3.2 who went to the hospital because she broke her ankle when she fell down the stairs. The first thing to notice is that, for the breaking of her ankle to be a cause of her visiting the hospital, there needs to be a causal chain connecting the two events. If our friend broke her ankle but treated it herself and went to the hospital to visit her brother, the causal explanation would be false. If and when there is such a causal chain connecting the two events, it seems natural to say that her visiting the hospital causally depends on her falling down the stairs. After all, had she not fallen down the stairs and broken her ankle, she would not have visited the hospital, unless some other accident had happened instead. Lewis thinks that we should take this ordinary way of talking about causal dependence very seriously and offers his counterfactual analysis as a way of capturing it.³¹ To say that event c caused event e is best understood as to mean that, if event c had not occurred, event e would not have occurred. But this leaves it open whether causation explains the counterfactual dependence or vice versa. Lewis argues for the latter. What makes such counterfactuals true are not causal

²⁹ After sketching a few allegedly unsurmountable challenges, Lewis (1973a, 557) concludes that “it is time to give up [on trying to save regularity theory] and try something else.”

³⁰ See Beebee (2006, 511f.) for a defense of this claim.

³¹ See Lewis (1973a, 557f.).
relations; rather, it is the degree of similarity between relevant possible worlds, where the degree of similarity is taken to be a primitive relation.\textsuperscript{32}

Possible worlds are ways our world might have been.\textsuperscript{33} To say that something was possible, even though it did not actually happen, is to say that there is at least one possible world in which it did happen. To say that something happened necessarily is to say that it happened in all possible worlds. One problem is that few things, if any, are metaphysically necessary in the strong sense that they hold in all possible worlds. Even the laws of nature are typically thought to be contingent, i.e., they could have been different. But that is just to say that there are possible worlds in which those laws do not hold. It then may seem difficult to analyze causation in terms of counterfactual dependence. For example, there may be worlds in which broken ankles heal instantly. But in such worlds our friend probably would not have gone to the hospital after falling down the stairs. Similarly, there may be worlds in which bones break all the time so that people go to the hospital at least once a day. In such worlds, it would not be true that, had our friend not broken her ankle, she would not have gone to the hospital. Since most counterfactuals will thus turn out to be false unless we specify the relevant worlds that are to be compared to our world, Lewis suggests that “a counterfactual is nonvacuously true iff it takes less of a departure from actuality to make the consequent true along with the antecedent than it does to make the antecedent true without the consequent.”\textsuperscript{34}

\textsuperscript{32} That is, one that is not further analyzable; see Lewis (1973a, 559). Interestingly, Lewis begins his paper with a brief discussion of problems with the regularity theory. But, as just noted, Beebee (2006) claims that Lewis’ counterfactual theory of causation can be understood as a sophisticated version of the regularity theory.

\textsuperscript{33} While Lewis (1973a) presents the basic counterfactual analysis, illuminating additional points can be found scattered throughout his work. Psillos (2002, 92ff.) and Horwich (1987/1993) both highlight some of these points that can be found in Lewis’ work. The discussion here draws partly from their summaries as well as from Lewis’ own writing.

\textsuperscript{34} Lewis (1973a, 560).
The above yields the following analysis of causation:

i.  
   c is a cause of e if and only if
      i.i  e depends causally on c, or\(^{35}\)
      i.ii there is causal chain leading from c to e.\(^{36}\)

ii.  
   e causally depends on c if and only if c and e occurred and e counterfactualy depends on c.

iii. e counterfactually depends on c if and only if
       iii.i if c occurred in the relevant worlds, e occurred as well, and
       iii.ii had c not occurred in the relevant worlds, e would not have occurred.
       iii.iii the relevant worlds in (iii.i) and (iii.ii) are those in which c and e occurred and
            that are closer to the actual world than worlds in which c occurred but not e.\(^{37}\)

To compare possible worlds with respect to their relative closeness to the actual world, we need to weigh the following maxims against each other: (a) avoid widespread violations of laws of nature, (b) maximize the spatiotemporal region with perfectly matching facts, (c) avoid small, localized violations of laws of nature, and (d) it is not very important to secure an approximate similarity of particular facts.\(^{38}\) The earlier maxims are supposed to weigh more than the latter ones. Worlds in which the laws are very different from ours tend to be more remote from ours than worlds in which only some local facts are different. A world in which bones heal instantly would be more remote than a world in which our friend ends up not going to the hospital merely because her car is out of gas, for example. All key components of Lewis’s counterfactual analysis are now on the table. We can start applying it to the exclusion problem.

---

\(^{35}\) We may be tempted to think that causal dependence is necessary for causation, not merely sufficient. But, as Lewis (1973a, 563) points out, causation is usually considered transitive whereas causal dependence is not because counterfactual dependence is not. Thus, causation does not imply causal dependence.

\(^{36}\) Ibid., p. 563.

\(^{37}\) This is a simplification, for Lewis phrases this in terms of propositions. See Lewis (1973a, 560f.) for a more careful way to spell this out.

\(^{38}\) See Lewis (1979, 472). Lewis (1973, 563) explains the notion of a causal chain by way of causal dependence: “Let c, d, e, … be a finite sequence of actual particular events such that d depends causally on c, e on d, and so on throughout. Then this sequence is a causal chain.” However, while such causal chains are transitive, not all causal dependence is.
One standard objection to the counterfactual analysis of causation is that it has trouble handling some overdetermination cases. Consider the assassin case mentioned in section 3.3.1. Each shooter’s pulling the trigger supposedly causes the death of the victim. So, we have two independent causes, \( c \) and \( d \), that satisfy (iii.i) because they each lead to the death of the victim \( e \). If \( e \) counterfactually depends on \( c \) and \( d \), this implies that, had \( c \) not occurred, \( e \) would not have occurred, and that, had \( d \) not occurred, \( e \) would not have occurred. But that seems false. Had \( c \) not occurred, \( d \) still would have caused \( e \); and had \( d \) not occurred, \( c \) would have caused \( e \). Thus, \( e \) counterfactually depends neither on \( c \) nor \( d \). True, there are two causal chains, one that connects \( c \) and \( e \) and one that connects \( d \) and \( e \), such that (i.ii) is satisfied and \( c \) and \( d \) still qualify as causes. However, if we look at the events in each chain that occur right before the effect such that we don’t have any intermediate events that are part of a causal chain, then the analysis might yield the conclusion that neither one of these events is a cause because neither (i.i) nor (i.ii) is satisfied. This may be a sign that the analysis is incomplete at best, but it may ultimately fail.\\footnote{We have introduced only what Paul (2009, 159) calls the “simple account.” Her paper offers a good overview of problems with this simple account as well as of various ways the account can be developed further. See pp. 177–182 for a discussion of how counterfactual theories may deal with overdetermination cases. Much more could be said here, but the point here is merely to explain how the counterfactual theory may affect the exclusion problem. It may thus suffice to merely indicate that the analysis may have problems with some type of overdetermination cases, namely those that don’t involve causal chains with intermediate events.}

While the counterfactual analysis may be able to handle some overdetermination cases and not others, it remains uninformative regarding systematic overdetermination. The nonoverdetermination assumption (5) should not be endorsed or rejected based on it. In that respect, it may differ at least from the naïve regularity theory, unless Beebee is correct when she states that the counterfactual analysis ultimately rests on a regularity theory of sorts, in which case our attitude towards (5) should not differ significantly.
Some have argued that the counterfactual analysis of causation can be used to reject the completeness assumption (4) and, by extension, the exclusion assumption (6). Supposedly, there are cases in which the better candidate for the property that e’s occurrence counterfactually depends on is a nonphysical rather than a physical property.\textsuperscript{40} Suppose that our friend was in severe pain after she fell down the stairs and broke her ankle and we want to explain why she went to the hospital. “Because it hurt terribly” seems like a natural response. Assuming that her pain is realized by the neuronal activity in her brain, we should also consider the neuronal activity as a potential cause, especially with the completeness assumption (4) in mind. If it is true that her being in pain caused her to go to the hospital, then, according to (iii.ii) it is true that, had she not been in pain, she would not have gone to the hospital.\textsuperscript{41} By the same token, if it is true that her neuronal activity caused her to go to the hospital, assuming it is physical in nature to satisfy (4), then it is true that, had she not exhibited that neuronal activity, she would not have gone to the hospital. To avoid overdetermination, we need to decide which of these counterfactuals designates an actual cause of our friend’s going to the hospital. Either the physical cause is adequate, in which case attributing actual causal impact to the pain implies overdetermination, or the physical cause cited in (4) is not adequate, i.e., (4) is false, in which case the pain might have causal impact. Thus, we need to determine which of the counterfactuals describes a world closer to ours.

\textsuperscript{40} Zhong (2011, 130) reads Horgan (1989), Horgan (1997), and Baker (1993) as defending this strategy.

\textsuperscript{41} Strictly speaking, this is too quick. Since e is a cause of e if and only if (i) or (ii), it could be the case that the pain is a cause of her going to the hospital merely by being part of a causal chain that leads to that. We cannot simply assume that her going to the hospital counterfactually depends on the pain. We will get back to this point in a bit.
While our friend’s pain may be realized by neuronal activity in the brain, such pain can be expected to be multiply realizable. The same kind of pain might have been realized by slightly different neuronal activity. Since securing approximate similarity of particular facts is not very important by comparison to more extended spatiotemporally mismatching regions, a world in which our friend does not exhibit the particular neuronal activity and does not go to the hospital seems more remote than a world in which our friend exhibits a slightly different neuronal activity and still goes to the hospital. But a world in which our friend does not feel pain and does not go to the hospital will arguably be closer to one in which she doesn’t feel pain and yet still goes to the hospital. This indicates, first, that our friend’s going to the hospital causally depends on her being in pain and perhaps, second, that her being in pain is a better candidate for being a cause than a particular neuronal activity. While the ordering of possible worlds here involves conjecture and may be affected by our epistemic interests, this may at least indicate that a counterfactual theory of causation might make it easier to deny the completeness assumption (4).

Let us generalize: there may be cases where the better candidate for a cause of a physical event is the nonphysical event that has a causal impact on the event in virtue of its nonphysical property. If that is so, the exclusion threat may be circumvented because the counterfactual’s proximity to our actual world prevents seemingly competing causes from actually being overdetermining, quite simply because the completeness assumption (4) is false. However, this strategy faces two problems. First, it may be question-begging. Recall that we simply assumed

---

42 Putnam (1967) points out that organisms with quite diverse brain structures seem capable of experiencing pain. Furthermore, it is likely that even within one organism there will be slight differences in the brain areas that realize pain. Against that background, it seems reasonable to assume that pain is multiply realizable.

43 At least that is what Zhong (2011, 134) argues.
that our friend’s going the hospital counterfactually depends on her being in pain. But since \( c \) is a cause of \( e \) if and only if (i) \( e \) depends causally on \( c \), or (ii) there is a causal chain leading from \( c \) to \( e \), this is too quick. For if being in pain were a cause merely in virtue of being part of a causal chain leading to the hospital visit, we would not have compared it to the adequate physical causes by way assessing the proximity of possible worlds. Second, the strategy may not appreciate what adequate physical causes entail. Arguably, the part that realizes the pain likely will be only a partial cause that contributes to the effect without being adequate itself. Adequate causes of physical events may often be extremely complex, for they must explain the occurrence of the effect in such a way that no reference to other events and properties is required. But the more complex they are, the more factors there are to keep in mind in figuring out which world is closest. So, unless we have a clearer idea of what being an adequate cause typically entails, we don’t have a clear idea of whether or not nonphysical events and properties are in fact the better candidates for causes.\(^{44} \) Still, we may flag this and keep in mind that a counterfactual analysis may have an impact on how likely it is that we hold onto the completeness assumption (4).

3.3.3 INUS Conditions

According to regularity theories and counterfactual theories, what counts as cause and effect depends on facts that go beyond the relation between these events themselves. Regularity theories construe causation as a relation that depends on how other events of the same type relate. Counterfactual theories capture the modality involved in causation by way of the similarity between and the proximity of possible worlds. By contrast, INUS conditions locate the modality in proximity to the causal relation itself.

\(^{44} \) Maslen, Horgan, and Daly (2009, 544) make an argument similar to Zhong’s. But while we have treated the proximity of the possible worlds to the actual one as a feature of the world, they argue that the ordering of the worlds is always contextual. In some explanatory contexts, it may be in virtue of nonphysical properties that \( c \) causes \( e \), as Zhong claimed, whereas in others it may not be.
The basic idea of INUS conditions is best conveyed by way of an example. Suppose a short-circuit caused a fire in a house. On the one hand, the short-circuit was not necessary for the fire. Had this particular short-circuit not happened, another one might have happened shortly after. On the other hand, the short-circuit on its own was not sufficient either, for without the presence of a complex set of conditions such as inflammable material, oxygen, and so forth, there would not have been a fire. So, the short-circuit on its own was neither necessary nor sufficient for the fire. But together with these other conditions, the short-circuit was sufficient for the fire. While the conditions and the short-circuit were not necessary for the fire because something else might have caused it, the short-circuit was a necessary part of the conditions, for the other conditions on their own would not have produced the fire. Thus, the short-circuit is “an insufficient but necessary part of a condition which is itself unnecessary but sufficient for the”\textsuperscript{45} fire. We can express this idea more schematically as follows:

\[
\text{If } c \text{ is a cause of } e, \text{ then } c \text{ is an INUS condition of } e, \text{ i.e., } c \text{ is an insufficient but necessary part of a condition which is itself unnecessary but exclusively sufficient for } e. \textsuperscript{46}
\]

However, which conditions are relevant depends on what exactly it is that the cause is supposed to explain. Often, the explanatory context of causal statements will illuminate this further. When we ask why the house caught fire, we may be interested in different things. As we have pointed out in section 3.2, such a question may be understood contrastively. Perhaps we know that other houses in the neighborhood have circuitry of the same type installed, and yet they have not caught fire. Or perhaps we know that the circuitry in this particular house was malfunctioning for some time already, so that we wonder why the house caught fire now rather than at an earlier

\textsuperscript{45} Mackie (1965, 245). The short-circuit example is drawn from Mackie as well.

\textsuperscript{46} This is a slightly modified version of a summary of Mackie’s view that Sosa and Tooley (1993a) provide.
time. In the former context, the conditions will include facts about houses with the same
circuitry. In the latter context, the conditions will include historical facts about the condition of
this particular circuitry. Mackie calls these regions in which causes and effects are assessed
‘causal fields’ and includes them in his definition:

If \( c \) is a cause of \( e \) relative to a particular causal field \( F \), then, given the features
that characterize \( F \), \( c \) is an INUS condition of \( e \), i.e., \( c \) is an insufficient but
necessary part of a condition which is itself unnecessary but exclusively sufficient
for \( e \). \(^{47}\)

The first thing to notice here is that this account, like the counterfactual theories of
causation, has trouble handling overdetermination cases. Consider the assassination case again.
In the absence of the other assassin, each assassin’s pulling the trigger, together with the
conditions relevant for the causal field in question, would be sufficient for the victim’s death.
But when both pull the trigger, neither is a necessary part of the condition which is itself
unnecessary but exclusively sufficient for the victim’s death. The analysis then implies that
neither trigger-pulling was a cause of the victim’s death. \(^{48}\) This is, of course, consistent with the
nonoverdetermination assumption (5). However, if overdetermination is impossible in general, it
is not merely systematic overdetermination we have to deny. As a consequence, atypical
nonphysical events of the sort not captured by the causal impact assumption (2) could not have a
causal impact. In that respect, INUS conditions seem to yield results similar to the naïve
regularity theory we discussed in section 3.3.1. However, the naïve regularity theory rejects
atypical causes because they would violate the regularity requirement. INUS conditions reject
them because the concept of atypically overdetermining causes is as incoherent as any

\(^{47}\) See Mackie (1965, 249). As before, this is a slightly modified version to match our previous notation.

\(^{48}\) Only the disjunction of both trigger pulls is necessary
overdetermination. Counterfactual theories, on the other hand, have been able to handle overdetermination cases that involved intermediate events. Thus, they may allow for events that have a causal impact in virtue of their atypical nonphysical properties of the kind $l$ that are not excluded because they don’t imply systematic overdetermination. They face exclusion in cases only where overdetermination does not involve intermediate events.

A second problem may be this. In cases where nonphysical events supposedly have a causal impact in virtue of their nonphysical properties of the kind $k$, they supposedly overdetermine the physical effects that supposedly already have an adequate physical cause. But in genuine overdetermination cases, Mackie’s analysis implies that neither of the alleged causes is in fact a cause. While the nonphysical event may be excluded from having a causal impact in virtue of its nonphysical properties, i.e., while (2) and (3) may turn out to be false, some physical events too would be excluded, namely those that causally explain the same features of the effect that the nonphysical properties would explain. But now this threatens the adequacy of physical causes in (4) and (6) and, by extension, the truth of those assumptions.

However, proponents of an analysis of causation in terms of INUS conditions may avert such grave consequence by pointing out that, in all likelihood, the causal fields involved in explanations citing nonphysical causes and properties will differ from those that cite physical causes and properties. If so, both physical and nonphysical events may qualify as a cause of the same event, but relative to different causal fields. As a result, there is no genuine overdetermination, and the nonphysical cause is not threatened by exclusion. This response is only convincing, though, if the adequacy of a cause must be relative to the causal fields as well. That requires some justification. Still, INUS conditions help to substantiate the worry we voiced
in section 3.2 that our epistemic and conversational context may determine what type of causal explanation we expect or would accept as satisfactory could be substantiate.

3.3.4 Probabilistic Theories

Probabilistic theories of causation treat causes that necessitate their effects as special cases of a more general relation, namely, as events that raise the probability of the effect’s occurrence to 1. In general, such theories merely state that causes make a difference to or influence the probability of their effects. Patrick Suppes expresses that idea in the following way:

An event c is a prima facie cause of an event e if and only if
i. c occurs earlier than e, and
ii. the conditional probability of e occurring when c occurs is greater than the unconditional probability of e occurring.49

As mentioned in section 2.2., we may wonder whether a positive influence as required by (ii) is needed. Perhaps causes sometimes negatively affect the probability of an effect.50 For example, we may assert that taking contraceptive pills lowers the likelihood of a pregnancy. While this sounds like a natural thing to say, we may also worry. After all, if the pregnancy occurs, the contraceptive pills were not among its causes. Rather, we think the pills had a causal impact on the prevention of the pregnancy, albeit not sufficiently so. For the purposes of this project, we don’t need to take a stance. (ii) can easily be rephrased so that it can accommodate both the view that causes typically raise probabilities and the view that they sometimes lower probabilities:

ii.* if an event c is the cause of another event e, the probability of e occurring when c occurs is different from the unconditional probability of e occurring.

This would capture that causes make a difference, regardless of whether they raise or lower the probability of an event to occur.

49 See Suppes (1984, 151). This version slightly modifies Suppes’ own presentation to match the way we refer to particular events.

50 Rosen (1978, 605–606) seems to think this is not unreasonable.
Even though this is barely even a sketch of a probabilistic theory, there are already interesting implications for the exclusion problem, especially with respect to the nonoverdetermination assumption (5). Recall the overdetermination case of two assassins, A and B, who pull the triggers of their respective rifles so that the bullets hit the victim at the same time in exactly the same way. Intuitively, it is correct to say that both assassins killed the victim. But consider how awkward the probabilities must be for this to qualify as a genuine case of overdetermination. Suppose that the probability of the victim’s dying from being hit by one of the bullets is 0.7 whereas the probability of the victim’s dying at the same time without being hit by any of the bullets, say, by a heart attack instead, is 0.00000001. If only assassin A took a shot, A’s pulling the trigger would clearly qualify as a cause of the victim’s death, according to Suppes’ analysis. The trigger-pulling occurred before the victim’s death, and the conditional probability of the victim’s death when the trigger is pulled, 0.7, is different from the unconditional probability of the victim’s death, 0.00000001. (i) and (ii) of the causal biconditional would be satisfied and the trigger-pulling turns out to be a cause according the probabilistic analysis. In section 3.3.1 we pointed out that genuine overdetermination cases require that the effect brought about by each cause is the same in all its features.51 We thus should assume that the probability of the victim’s death when B pulls the trigger in the absence of A would be 0.7 as well. The analysis applies again and B’s trigger-pulling in the absence of A is a cause as well.

51 In Suppes’ fully developed theory, the time at which an event occurs is always specified; see, for example, Suppes (1970, 12). Consequently, it would be impossible to have a genuine case of overdetermination where c and d would cause the event to appear at slightly different times in the absence of the other cause, say, if the second shot killed the victim a split second later. So, the analysis requires temporal identity of the effects of c and d, as would Kim’s account of events.
However, overdetermination cases are supposedly different from regular joint causes. In the case of regular joint causes, even if we assumed that $P(e \mid c) = 0.7$ and $P(e \mid d) = 0.7$, the joint probability of $e$ given $c$ and $d$ would be different than the probability given just $c$ or $d$. For example, we might get $P(e \mid c, d) = 0.9$. Of course, we cannot just add up the probabilities. If the victim is already likely to bleed to death from one shot, the second shot might just lead to an increased blood loss which makes it more likely that the victim dies, but not necessarily drastically more likely. So, what distinguishes a genuine overdetermination case from a regular case of joint causes? Answer: the fact that the effect needs to be exactly the same. It looks like we have to say that the probability of $e$, given $c$ and $d$, must equal the probability of $e$, given either $c$ or $d$. But that is problematic, for now neither (ii) nor (ii*) is satisfied for $c$ and $d$. The conditional probability of $e$ occurring when $c$ occurs is no different from the probability when $c$ does not occur: $P(e \mid c, d) = P(e \mid d)$. Therefore, $c$ is not a cause of $e$. And the conditional probability of $e$ occurring when $d$ occurs is no different from the probability when $d$ does not occur: $P(e \mid c, d) = P(e \mid c)$. Therefore, $d$ is not a cause of $e$. Consequently, the simple probabilistic theory above, at least on its own, has something in common with the simple counterfactual analysis in section 3.3.2: it has difficulties handling overdetermination cases.

While the probabilistic analysis, like the INUS conditions discussed in section 3.3.3, can endorse (5) in that the analysis agrees with the claim that there is no systematic overdetermination, it does so because the alleged causes involved turn out not to be causes at all.

---

52 Of course, this is very sketchy. See Pearl (2000) for one of the most detailed accounts of how to model probabilistic causation and how to calculate the impact of causes especially in Bayesian networks.

53 Paul (2009, 179) seems to think that is the correct assessment: “In the case of indeterministic laws, the question we should ask is why the presence of an additional, overdetermining cause would not change the probabilities of the outcome.”
In that respect, it seems stronger than the counterfactual analysis discussed in section 3.3.2., which could provide a convincing analysis of at least some overdetermination cases and which seemed otherwise neutral with respect to systematic overdetermination.\textsuperscript{54} But like INUS conditions, the probabilistic analysis has the further interesting consequence that it may cause trouble for the completeness assumption (4).

As a reminder, in cases where nonphysical events supposedly have a causal impact in virtue of their nonphysical properties of the kind $k$, they supposedly overdetermine the physical events that supposedly already have an adequate physical cause. But in genuine overdetermination cases, the simple probabilistic analysis concludes that neither of the alleged causes are, in fact, causes. While the nonphysical event may be excluded from having a causal impact in virtue of its nonphysical properties, i.e., while (2) and (3) may turn out to be false, some physical events would also be excluded, namely those that causally explain the same features of the effect that the nonphysical properties would explain. But now this threatens, like INUS conditions, the adequacy of physical causes in (4).

As in the case of the counterfactual analysis of overdetermination by two events with no intermediate events, it seems strange to think that the mere presence of another event that is a cause undermines the status of an event as a cause. It then seems natural to search for features intrinsic to the causal relations that would explain why $c$ and $d$ are causes, after all.\textsuperscript{55} While process and transference theories do not have to be probabilistic, they can be and have been combined with the latter kind of theory and may be able to circumvent this issue, as we will see in the next section. If that were not an option, we might be well advised not to apply the simple

\textsuperscript{54} Unless Beebee is right in thinking that counterfactual theories ultimately rest on regularity theories of some sort.

\textsuperscript{55} We will get back to the point in the context of causal pluralism in section 3.3.7.
probabilistic theory of causation to the exclusion problem. If a theory of causation leads us to reject not only (2) and (3) but also (4), on analytic grounds rather than on empirical ones, the best explanation may be that there is something wrong with the causal analysis as it stands. Given that we merely provided a sketch of the basic idea underlying probabilistic theories, and given that most probabilistic theories are rather sophisticated, this does not seem particularly surprising. Alternatively, we could just accept that in the context of the exclusion problem probabilistic theories imply that there are never nonphysical events that cause and overdetermine physical events, i.e., the threat of exclusion is very serious.

3.3.5 Transference Theories

Like process theories, transference theories are usually reductionist. Some of the former theories explain causation in terms of processes and interactions; some of the latter theories explain it in terms of quantities such as momentum transferred from one object to another.\textsuperscript{56} The basic idea may be construed as follows:

An event $c$ causes an event $e$ if and only if

i. $e$ involves a change in an object,

ii. at the time $e$ occurs, the object the change in which constitutes the cause of $e$ is in contact with the object that undergoes the change which constitutes $e$,

and

iii. prior to the time of the occurrence of $e$, the body that makes contact with the effect object possesses a quantity which is transferred to the effect object when contact is made and manifested in $e$.\textsuperscript{57}

\textsuperscript{56} At this point, the most obvious way to proceed would be to discuss Salmon’s process theory, developed in Salmon (1984), for it connects well to our discussion of probabilistic theories and promises to overcome at least some of the problems addressed. However, it is essential for Salmon’s view to replace events with processes. This would make a discussion of his view and an application to the exclusion problem rather unnecessarily complicated and difficult. We thus opted to discuss Aronson’s transference theory instead, for, although it is less known, it is a reasonably close relative to process theories such as Salmon’s, while it is easier to apply to the exclusion problem.

\textsuperscript{57} This way of presenting the general idea of the transference theory is inspired by Aronson (1971, 421f.) but drastically modified in order to exclude claims that don’t seem relevant for our purposes. For example, Aronson distinguishes between natural and unnatural changes, where natural changes are internal changes in an object which he does not regard as instances of causation. Only external changes qualify.
We can easily make sense of (i) within a Kimean account of events, according to which events are exemplifications of a property by an object at a time.\textsuperscript{58} However, the account itself leaves it open whether the property exemplified needs to be distinct from properties previously exemplified by the object. (i) seems to require such distinctness; at least it seems natural to understand change in that way. (ii) is basically capturing the contiguity requirement we already encountered in the context of the naïve regularity theories in section 3.3.1.\textsuperscript{59} And (iii) specifies that the quantity transfer happens between the objects constituting the events that are cause and effect, at least on a Kimean reading.

Regardless of whether this kind of theory is ultimately defensible or not,\textsuperscript{60} it may appeal because it agrees with a layman’s understanding of physical processes. Paradigm cases of causation such as two billiard balls hitting each other usually involve physical contact. Furthermore, the way we describe such cases suggests that we are often aware that physical quantities matter. For example, to explain why the red five had a certain impact on the black eight when hitting it, we may specify the speed of the red ball, presumably because we are aware that the velocity of an object is related to its kinetic energy which in turn is needed to explain the impact the red ball has. Thus, it should not surprise us that causation requires contact and involves change and the transfer of some quantity. Plausible enough. When applying it to the exclusion problem, however, problems arise quickly.

\textsuperscript{58} See section 2.1.1.

\textsuperscript{59} If causation consists in the transmission and transference of a quantity, contiguity seems particularly important. However, we may think that this is a disadvantage. In fact, Hall (2009, 168ff.) considers it an advantage of counterfactual theories that they can handle situations where there is no process, no transfer of energy, or in general no contiguity.

\textsuperscript{60} For a critical assessment of the Aronson’s original version, see Earman (1976).
The examples of transferred quantities are usually physical: kinetic energy, momentum, velocity, and so forth. If they are, we can expect the physical cause in the completeness assumption (4) to account for all of them. It is hard to see how else it could be adequate. In that sense, the theory might even lend support to (4). But now consider the causal impact assumption (2). Supposedly, the nonphysical event has a causal impact on a physical event in virtue of its nonphysical properties of the kind $k$. Should we assume that the properties that ground the transfer of such physical quantities have to be physical as well? If nonphysical properties can ground the transference of physical quantities, we need to explain how this can be.\footnote{This, of course, is a particular version of the problem of mental causation sketched in section 1.1.} If we could, what would it mean for the adequate cause in (4)? If there is also a nonphysical property that grounds that transference, we would face systematic overdetermination. If there is not, (4) may be false depending on whether or not we can preserve the adequacy of the cause in (4) even in the presence of transfer that is grounded by nonphysical properties only. If only physical properties can ground such transference, then (2) turns out to be false.\footnote{Dennett (1991, 35), for example, takes energy conversation laws to be part of an argument against assumptions such as (2) exactly because no physical energy or mass is associated with minds.}

Alternatively, we may consider the existence of nonphysical quantities. But then it seems the impact a nonphysical event can have on a physical event cannot be of the same sort the adequate physical cause has, for one transfers a nonphysical quantity and the other a physical quantity. Unless nonphysical and physical quantities can convert into each other somehow,\footnote{A passage in Collins (2008, 38) may be read as suggesting something along these lines: “A dualist could argue that, like the gravitational field, the notion of energy simply cannot be defined for the mind, and hence one cannot even apply PEC to the mind/body interaction. The mind, like the gravitational field, could cause a real change in the energy of the brain without PEC applying to the interaction,” where PEC is the principle of energy conversation as used in physics.} this would yield an implausible picture of the world in which nonphysical events can have a causal
impact on physical events, but only by way of transferring nonphysical quantities that in turn cannot affect physical quantities. If, however, nonphysical quantities can convert into physical quantities, nonphysical events might indeed have a causal impact on physical events; but that would violate the completeness assumption (4)\(^\text{64}\) and the nonoverdetermination assumption (5). Considering how obscure such a view would be, it may be more reasonable to expect nonphysical events either not to have a causal impact at all or to have it in virtue of physical properties and qualities they possess. But in that case, the causal impact assumption (2) is false, as we said earlier. That is not a particularly surprising result, of course. Given that the standard examples of transferred qualities are usually physical, we might suspect that proponents of transference theories tend to generally be friendly towards a physicalist world view and skeptical of assumptions such as (2).

### 3.3.6 Single-Difference Theories

As mentioned in section 3.3.4, regularity theories construe cause and effect as a relation that depends on how other events of the same type relate. Single-difference theories, by contrast, deny that causation depends on regularities. Whether two events relate as cause and effect is solely a matter of the nature of that relation and of the events’ immediate environment. In that respect, single-difference theories resemble INUS conditions, except that the latter can construe causal fields as spatiotemporally widely extended. Like some transference theories, single-difference theories may conceive of events as changes in the property of objects. If such an event is sufficient for another event to occur, it is to be considered the cause of that event.\(^\text{65}\) Given such

\(^{64}\) According to Papineau (1996, 4), postulating such nonphysical “force alongside the fundamental physical forces of gravity, the electroweak force, and the strong nuclear force […] might once have made sense, but the cumulative evidence of two centuries of physiological research weighs heavily against it.”

\(^{65}\) See Ducasse (1926/1993, 126)
understanding of events and causes, C. J. Ducasse claims that $c$ was the cause of $e$ if $c$ was the only difference in $e$’s environment before $e$ occurred. Or, put more schematically:

\[
c \text{ is a cause of } e \text{ if }
\begin{align*}
i. & \quad c \text{ occurred during a time and through a space terminating at the instant } I \text{ at the surface } S, \\
ii. & \quad e \text{ occurred during a time and through a space beginning at the instant } I \text{ at the surface } S, \text{ and} \\
iii. & \quad \text{no change other than } c \text{ occurred during the time and through the space of } c, \text{ and no change other than } e \text{ during the time and through the space of } e.\text{ }^{67}
\end{align*}
\]

For example, when our friend’s foot hit the floor, her ankle broke. Assuming that nothing else was happening in the immediate environment, we can infer that her foot’s hitting the floor caused her ankle to break. We can infer this without ever having seen other people breaking their ankles and without knowing anything about physical laws. In many respects, this is one of the most striking features of this kind of analysis. It does not refer to any laws, regularities, or external factors beyond the immediate environment of the events. Because of that, $c$ can be the cause of $e$ even if events of their type never occur again. Of course, this does not rule out regularities. But if there are regularities, they supervene on the causal relations, not vice versa.\text{ }^{68}

This has interesting consequences. Suppose $c$ and $e$ fall under event types $C$ and $E$ respectively. Even if another event of type $C$ appears in a similar environment, there is nothing about the analysis that would require an event of type $E$ to follow. That is, the analysis is perfectly compatible with atypical causes.

\text{---}

\text{66} We relied on Psillos (2009, 146) for this concise formulation. Ducasse himself phrases this general idea of his view rather awkwardly: “…we may say that the cause of a particular change $K$ was such particular change $C$ as alone occurred in the immediate environment of $K$ immediately after”; see Ducasse (1926/1993, 127).

\text{67} See Ducasse (1926/1993, 127). We have adjusted the notion of events to match the one we are using in other sections. It remains somewhat unclear why Ducasse uses a conditional rather than a biconditional. He seems mainly interested in providing sufficient conditions for events to qualify as cause and effect. Still, this allows for there being causal relations due to factors other than those specified in (i) to (iii). Perhaps he merely thinks of the single-difference account as one that may accompany other definitions and theories.

\text{68} See Ducasse (1969, 21)
The theory is riddled with flaws. One, for example, is akin to a problem the naïve regularity theory faces: the analysis seems to lack the resources to distinguish between events that follow each other accidentally and events that are causally related. But let us put such worries aside for now. After all, we are not in the business of assessing the quality of theories of causation but in the business of applying them to the exclusion problem. So, let us do just that.

Whereas the naïve regularity theory implies that if there are nonphysical events that have a causal impact, they always need to be typical, the single-difference theory does not imply that at all. While the theory is compatible with the typicality assumption (3), especially given that (2) restricts the scope to a particular kind of events, the theory might allow for atypical events to be widespread. So, it may seem somewhat arbitrary that we restrict the exclusion problem to those kinds of events and properties. But if widespread atypical causal impact is possible, the completeness assumption (4) seems rather strong. Of course, a proponent of the single-difference theory might still endorse it for reasons unrelated to the theory. But that is quite different from the transference theory which, depending on what kind of quantity it allows for, might actually provide support for (4).

Furthermore, from the viewpoint of the single-difference theory, it is not clear why we should worry about overdetermination, systematic or not. After all, there seems nothing surprising about the fact that the two events \(c\) and \(d\) occur in close proximity right before \(e\) occurs. Accordingly, it should not be surprising that they are both a cause of \(e\), provided that nothing else happened nearby. So, there does not seem to be a threat of inconsistency, as we suspected lingers in the case of counterfactual theories, INUS conditions, and probabilistic theories.

---

69 See Psillos (2002, 68–71) for a brief discussion of problems with the theory.
Finally, it is unclear in what sense causes explain their effects, beyond the fact that nothing else happens in the immediate environment. We may worry what that lack of explanatory force does to our concept of an adequate cause. Suppose a physical and a nonphysical event occur in close proximity right before another physical event occurs. The completeness assumption suggests that the physical event on its own should explain the occurrence of the effect satisfactorily—that is, no reference to other events is necessary. But why should we privilege the physical event in such a way over the nonphysical, given that there are no laws that connect these events and that might have some hierarchical structure?

The single-difference theory might not be satisfying, metaphysically speaking. However, at least it served to emphasize how drastically our responses to particular assumptions of the exclusion problem may change depending on what kind of theory of causation we presuppose, especially in contrast to the previous theories discussed.

### 3.3.7 Causal Pluralism

All of the previous theories of causation face serious challenges—often in the form of counterexamples—that suggest that they are at best incomplete accounts. While not all of these challenges are fatal, each theory faces at least one challenge that seems insurmountable.\(^\text{70}\) If this is true, there are at least four ways to respond: (a) merge two or more of the above reductive theories into one complex theory of causation, (b) give up the assumption that there is merely one type of causation, as pluralists do, (c) give up causal reductionism, as some singularists do, or (d) give up the concept of causation altogether, as eliminativists do. We may expect that merging two or more of the above theories would have an impact on the exclusion problem that developed so far. She agrees that all suffer from fatal counterexamples and takes that to be an argument for causal pluralism. Psillos (2009, 154) also claims that there are counterexamples to all existing theories of causation.

---

\(^{70}\) Cartwright (2004) walks through some of these challenges, at least for the most promising theories of causation developed so far. She agrees that all suffer from fatal counterexamples and takes that to be an argument for causal pluralism. Psillos (2009, 154) also claims that there are counterexamples to all existing theories of causation.
is identical or similar to the combined impact of whichever theories are merged. We thus will not pursue (a) further. Instead, we will address (b) and (c) in this section and (d) in the following two sections.

There are several kinds of causal pluralism. One kind claims that our ordinary notion of causation conflates several distinct concepts. Once we avoid that ambiguity, we can identify a discrete set of criteria for each kind of causation. It then can be shown that allegedly fatal challenges rest on mingling criteria from different sets. Since these criteria can come apart, the challenges lose their force as a response to each kind of causation. A second kind of causal pluralism agrees that there are several kinds of causation but denies that there are discrete sets of criteria that distinguish each kind. Instead, these kinds resemble each other because their criteria overlap. Ned Hall pursues the former, ambiguity, strategy and Elizabeth Anscombe adopts the latter, family-resemblance, strategy.\(^\text{71}\)

Hall distinguishes between dependence and production. To see the difference, consider a forest fire that breaks out in June rather than in May because of heavy rains in April and assume that, had it not rained in April, a fire would have broken out in May instead. The fire in June counterfactually depends on the rain in April. Had the rain not occurred in April, the fire in June would not have occurred. By the same token, it seems odd to claim that the rain produced the fire. The counterfactual analysis fails to capture this difference between dependence and production and yields the claim that the rain in April was a cause of the fire in June.\(^\text{72}\) According to Hall, this failure is one in principle. Counterfactual theories are bound to fail because they

\(^{71}\) Of course, this is not an exhaustive list of the kinds of causal pluralism that have been developed. Skyrms (1984), for example, argues that there is an “amiable jumble” of criteria which we draw from as needed in different explanatory contexts. For a brief discussion of Anscombe, Hall, and Skyrms, see Godfrey-Smith (2009).

\(^{72}\) See Hall (2004, 229–231). See also section 3.3.2 for a basic introduction to the counterfactual analysis.
cannot handle overdetermination and double-prevention cases at the same time, even if combined with further assumptions. To demonstrate this, Hall introduces the following five theses:

Transitivity: If event $c$ is a cause of $d$, and $d$ is a cause of $e$, then $c$ is a cause of $e$.

Locality: Causes are connected to their effects via spatiotemporally continuous sequences of causal intermediates.

Intrinsicness: The causal structure of a process is determined by its intrinsic, noncausal character (together with the laws). […]

Dependence: Counterfactual dependence between wholly distinct events is sufficient for causation. […]

Omissions: Omissions—failures of events to occur—can both cause and be caused.73

The problem is that the counterfactual analysis fails for some types of overdetermination unless it adopts intrinsicness, locality, and transitivity. To illustrate, Hall discusses a late preemption case of the following sort:

Neurons $a$ and $c$ fire simultaneously, so that $e$ fires at the same time as $b$; the inhibitory signal from $e$ therefore prevents $d$ from firing. If $c$ hadn’t fired, $e$ still would have; for in that case $d$ would not have been prevented from firing and so would have stimulated $e$ to fire. Likewise for every event in the causal chain leading from $c$ to $e$: If that event had not occurred, $e$ would nevertheless have fired. So the strategy of finding suitable intermediates breaks down; for it to succeed, $e$ would have to depend on at least one event in the chain leading back to $c$, and it does not.74

But $e$ lacks the right sort of counterfactual dependence only because of an accidental feature of its surrounding, namely, the fact that there are two events, $a$ and $c$, and two causal chains, each of which would lead to $e$ in the absence of the other chain. However, intuitively the relation

73 Ibid., p. 225; italics in the original. On page 339, Hall elaborates on the intrinsicness thesis in a useful way. It “states that any possible structure of events that exists in a world with the same laws, and that has the same intrinsic character as our given structure, also duplicates this aspect of its causal character—that is, each duplicate of one of $e$’s causes is itself a cause of the $e$-duplicate.”

74 Ibid., p. 235. Late preemption is a type of overdetermination, according to Hall.
between $c$ and $e$ seems causal, regardless of the lack of counterfactual dependence in virtue of the presence of another causal chain.\textsuperscript{75} The intrinsicness thesis tries to make sense of this. But if we adopted the intrinsicness thesis to get the right results for late preemption cases such as the above, we would fail in our analysis of double-prevention cases such as following:

Neurons $a$, $b$, and $c$ all fire simultaneously. The firing of $c$ prevents $e$ from firing; if $e$ had fired, it would have caused $f$ to fire, which in turn would have prevented $g$ from firing. Thus, if $c$ had not fired, $g$ would not have. So $c$ is a cause of $g$.\textsuperscript{76}

For $c$ to qualify as a cause of $g$, we need to accept that omissions can be causes. After all, $c$ prevents $e$ and $f$ from occurring. There is no causal chain connecting $c$ and $g$. This violates the intrinsicness thesis because nonoccurring events cannot enter nomological relations. Hall then demonstrates that similar inconsistencies arise if we try pairing the counterfactual dependence with locality and transitivity. The problem can be solved by assuming that we are really dealing with two distinct kinds of causal relation. The dependence relation captures the dependence and omission theses whereas the production relation captures the intrinsicness, locality, and transitivity theses.\textsuperscript{77}

How does this distinction between dependence and production affect the exclusion problem? Recall that assumptions (2) to (6) use the notion of causal impact. In section 2.2, we said that when events have a causal impact, they are strictly involved in the production of the effect, i.e., they are part of a causal chain that leads to the effect. It then looks as if the causal impact relation matches Hall’s production relation because being part of a causal chain implies

\textsuperscript{75} Interestingly, this point echoes one that we made in section 3.3.1. We suggested that overdetermination cases such as the assassin case lack systematicity in virtue of the circumstances rather than in virtue of the properties of the causes that overdetermine the death of the victim.

\textsuperscript{76} Hall (2004, 241f.).

\textsuperscript{77} Of course, this is a very rough sketch of Hall’s argument that leaves much to be clarified. However, here we are merely trying to illustrate why Hall distinguishes between two rather than three or more kinds of causation and to explain why he adopts dependency and production rather than other kinds of relation.
locality and transitivity. It is, however, not clear whether causal explanations invoking nonphysical properties of the kind $k$ always fit the bill. Suppose a colleague is surprised that we are still in the office at nine o’clock in the evening and comments, “I thought you promised your kids to take them out to the movies tonight.” Shocked, we may reply, “Oh, I totally forgot!” Having forgotten a promise made earlier seems like a perfectly natural, arguably causal explanation of why we are still in the office. But if it is, it rests on an omission, which according to Hall is captured by counterfactual dependence rather than by production. Hall’s distinction then may highlight a shortcoming in the way we have defined ‘causal impact.’ But changing that definition to allow for both counterfactual dependence and production is risky business, for it may lead to ambiguous assumptions. To avoid an equivocation, we need to make sure that the kind of causal relation involved in the assumptions (2) to (6) remains the same.

Of course, the easiest way to avoid complications is simply to ignore mere counterfactual dependence and omission cases and stick to production. But this easy opt-out may not always be an option. Different kinds of causal pluralism distinguish between different kinds of causal relation. Some may introduce causal relations that are all compatible with our concept of causal impact, resulting in different kinds of causal impact. Excluding one kind of causal impact may seem ad hoc and unjustified, so we may have to allow for several kinds. To avoid ambiguity, any application of the exclusion problem to specific events and properties would then have to specify what kind of causal impact is used in (2) to (6). However, Elizabeth Anscombe’s kind of causal pluralism may undermine this strategy.78

78 Sometimes, Anscombe’s view is also listed as a kind of singularism. Even though the terms ‘singularism’ and ‘pluralism’ suggest some kind of contrast, they do not refer to mutually exclusive positions. As will become clear shortly, Anscombe’s view is pluralist in that it assumes the existence of many different kinds of causal relation, loosely connected by way of family resemblance. But it also is singularist in spirit in that it is antireductionist. Causal relations could hold even if there were no laws that connected the types of events in question. That is why Moore (2009, 4) views Anscombe as a singularist.
Anscombe’s view suggests that the criteria for different kinds of causation don’t neatly line up in the way that Hall suggests. When we speak of causal relations in ordinary language, we often refer to very specific causal concepts such as “scrape, push, wet, carry, eat, burn, knock over, keep off, squash, make (e.g. noises, paper boats), hurt.”\(^79\) And sometimes we posit specific relations between various kinds of relata, such as “the pistons compress the air in the carburetor chamber, the sun attracts the planets, the loss of skill among long-term unemployed workers discourages firms from opening new jobs …”\(^80\) In English, we can refer to the relata in these relations as ‘cause’ and ‘effect’ and to the relation between them as ‘causation’; but when we do so, we are merely grouping them based on their family resemblance. Doing so requires mastery of the concepts in question. Because of that, such concepts seem semantically prior to the concepts of cause, effect, and causation. However, it is not as if every causal relation shared a single set of criteria that all others share. For any kind of causal relation there may be another kind that shares some of its criteria, but it is not guaranteed that any two relations share any such criteria.\(^81\)

While our notion of causal impact may exclude some such causal relations, it likely will group together several specific causal relations, especially given how different the vocabularies of various special sciences can be. Suppose we wanted to explain how a cell developed the shape it has. In biology, we might state that the cell membrane anchors the cytoskeleton and attaches

---

\(^79\) Anscombe (1971/1993, 93); italics in the original.

\(^80\) Cartwright (2004, 814); italics in the original. Cartwright’s pluralism seems very similar to one developed by Anscombe. Aronson (1971, 417f.) would be skeptical regarding Anscombe’s and Cartwright’s claim. He suggests that sentences with a specific causal verb can be replaced by sentences using ‘caused by,’ usually without significant loss of information. He does acknowledge exceptions, though.

\(^81\) Famously, Wittgenstein (1953, paragraphs 66–67) introduces the concept of family resemblance. Cartwright suggests we are abstracting from more specific content of the description and phenomena they describe. We thus are bound to lose information. For her, that is why these concepts have semantic priority.
“to the extracellular matrix and other cells to hold them together to form tissues.”

Suppose then that anchoring, attaching, and holding together are specific kinds of causal impact that events have in virtue of whichever properties of the cell membrane explain said effects. If we treated these as contenders for nonphysical events and properties, the exclusion problem then would get off the ground only under two conditions. Either the kinds of causal impact in (4), (5), and (6) match the kinds of causal impact in (2) and (3), or we explain how the latter kinds relate to the former kinds. The less they resemble each other, the less obvious it may be in what way such nonphysical events and properties are excluded by an adequate physical cause. We may be able to explain what the exact relation between these different kinds of causal impact is in specific instances, but without investing extra effort we may have no way of telling whether the exclusion problem is valid when we apply it to specific events and kinds of causal relations.

3.3.8 Causal Antirealism

All of the previous theories of causation were realist in spirit; they assumed that there are genuine, mind-independent, causal relations in nature. Most of these theories were also reductive: they attempted to explain the nature of causal relations by reference to other items in our ontology—for example, laws of nature. But even when they were reductive, they did not eliminate the concept of causation altogether. After all, providing an analysis of the concept by breaking it down into its constitutive parts is not necessarily to deny that the analyzed relation is real. Eliminativists such as Bertrand Russell want to do just that, deny its reality, albeit not necessarily for reasons related to reduction.83


83 The best anthology for contemporary views that at least in part rely on Russell is Price and Corry (2007b), although the views included do not always amount to eliminativism. In Price and Corry (2007a), the authors defend what they call ‘causal republicanism’: “the view that although the notion of causation is useful, perhaps indispensable, in our dealings with the world, it is a category provided neither by God nor by physics, but rather constructed by us.” This could be read in two ways. Either they merely reject the mind-independence required for
Russell offers two main arguments for eliminating the concept of causation from our ontology and explanatory practices. First, the concept itself is confused; and second, our best science, physics, does not employ it but relies instead on laws of nature.\textsuperscript{84} To develop the first argument, Russell presupposes a necessitarian theory of causation, one that endorses what he calls the law of causality: Given any event $c$, there is an event $e$ and a time-interval $T$ such that, whenever $c$ occurs, $e$ follows after an interval $T$.\textsuperscript{85} Since $c$ and $e$ are events, they are temporally extended; that is they must endure for some time.\textsuperscript{86} But if they do, we face a problem. For $c$ to be a cause of $e$, $c$ needs to be spatiotemporally contiguous with $e$, as we saw in section 3.3.1. But then only the latest part of $c$ can have a causal impact on $e$, for the earlier parts are not contiguous with $e$.\textsuperscript{87} Since they are not contiguous, there is a time interval between them and the earliest part of $e$. Because things could happen during this interval that prevent the occurrence of $e$, we must ensure that they do not. But then $c$ is not sufficient for $e$. Even worse, varying environments will require varying ways to ensure no interference. But the more of the environment we need to include, the less likely it is that $e$ occurs whenever $c$ occurs.\textsuperscript{88} While Russell seems to think that this ultimately undermines the necessitarian theory of causation, he mainly uses the point about varying environments to establish his second argument. Scientific realism as we have defined it in 3.3 or they propose some kind of epistemic pragmatism but ultimately endorse an ontological eliminativism.

\textsuperscript{84} In fact, Russell (1912–13) offers a third reason related to determinism and teleology, but will ignore it here.

\textsuperscript{85} Ibid., p. 4. We have adjusted the way Russell refers to particular events to match the way we have done throughout this dissertation.

\textsuperscript{86} This is Russell’s view. The terminological convention regarding events that we have established in section 2.1.2 is silent on the issue.

\textsuperscript{87} Ibid., p. 5.

\textsuperscript{88} Ibid., pp. 7f.
explanations rely on laws of nature rather than on the law of causality.\textsuperscript{89} The antecedents in actual nomological relations are likely so complex that the effect never will occur in exactly the same way. Consequently, the regularities suggested by the law of causality, that whenever $e$ occurs $e$ follows, have no equivalent in science.\textsuperscript{90}

Regardless of whether Russell’s view of science is accurate,\textsuperscript{91} let us assume that he is right in thinking that the concept of causation is expendable and should be eliminated from our ontology and explanatory practices. How would this affect the exclusion problem? As it stands, all but the distinctness assumption (1) include causal assertions. If causation is not a real relation, (2) to (5) all turn out to be false and the threat of exclusion is averted. Well, not really. It seems possible to construct a nomological version of the exclusion problem instead. To illustrate what such version might look like, we may replace the causal impact assumption (2) with a claim involving laws of nature:

\textbf{Nomological Subsumption:} Sometimes, nonphysical events are subsumed under a law in virtue of their nonphysical properties of the kind $k$, and the law connects these nonphysical events and physical events.

Assuming that laws are exceptionless, we might omit the typicality assumption (3) and rephrase the completeness assumption (4) as follows:

---

\textsuperscript{89} Not all eliminativists replace causation with laws of nature. Niels Bohr, for example, tried to replace causation specifically with complementarity on the atomic domain; see Stöltzner (2009, 112).

\textsuperscript{90} See Russell (1912–13, 9). Later, on page 12, he also points out that the laws in physics are probabilistic and thus don’t square well with the law of causality; but given that promising probabilistic theories of causation have been developed, as explained in section 3.3.4, this point seems rather moot nowadays.

\textsuperscript{91} See Lange (2009, 249ff.) for a critical discussion. Schaffer (2016, section 2.1) summarizes the main worry as follows: “The main objection to eliminativism is that causation is too central to eliminate. Causation, according to various contemporary philosophers, is required for the analysis of metaphysical concepts such as persistence, scientific concepts such as explanation and disposition, epistemic concepts such as perception and warrant, ethical concepts such as action and responsibility, mental concepts such as functional role and conceptual content, and linguistic concepts such as reference. Elimination is not just unjustified; it would be catastrophic.”
Completeness: If a physical event \( p_i \) is subsumed under any law at all, there is at least one physical law it is subsumed under; i.e., there is at least one law that connects \( p_2 \) and another physical event \( p_1 \).

Likely, we will need to specify the nomological subsumption assumption further. The problem is that there are many laws that supervene on physical laws without competing with them in any way. But without some kind of competition, we would not get the threat of overdetermination and exclusion. One solution might be to assume that the laws referenced in the nomological subsumption assumption do not themselves supervene on physical laws. But this assumption would require a proper defense. So, it may be trickier to rephrase the nonoverdetermination assumption and the exclusion assumption in terms of laws.

Still, doing so seems possible in principle. If it is, causal antirealism, especially in the form of eliminativism, does not have much of an advantage when dealing with the exclusion problem, at least not more of an advantage than those reductive theories of causation that rely on laws of nature. And, as far as the theory’s impact on the problem is concerned, it will probably depend on what kind of relation causation is replaced with. If laws of nature take its place, the impact may be identical to the one that the naïve version of the regularity theory or any theory of causation that grounds causal relations in the laws of nature has. If something like complementarity takes its place, as suggested by Niels Bohr, the impact may be different.

3.4 Chapter Conclusion

After restricting the scope of our inquiry to a representative rather than a comprehensive set of theories of causation, we cautioned against mingling domains of discourse too easily. Doing so may obscure what the causal explananda in a given explanatory context are. We then explained key features of various theories of causation and illustrated how they may affect the exclusion problem. Let’s review the main lessons learned.
The exclusion problem allowed for, but did not apply to, events that have atypical causal impact on physical events in virtue of their properties of the kind $l$ rather than $k$. This partly motivated the introduction of the typicality assumption to allow for nonsystematic cases of overdetermination. However, if we presuppose a naïve regularity theory, such atypical causes are not possible. That makes the way we phrased the causal impact assumption (2) unnecessarily complicated and renders the typicality assumption (3) superfluous. While the theory is compatible with the nonoverdetermination assumption (5), it is not clear why we would be particularly invested in it if we accepted the naïve regularity theory. If we were forced to give up one assumption to avoid exclusion, (5) might be a particularly promising candidate.

For some kinds of overdetermination, the basic counterfactual theory suggests that the allegedly overdetermining events are not causes at all, for the effect neither depends on them counterfactually nor is connected to them by way of a causal chain. Perhaps that provides particularly strong reasons to hold onto (5). Unlike the naïve regularity theory, the basic counterfactual theory may provide reasons for rejecting the completeness assumption (4) as well. There may be cases where the better candidate for a cause of a physical event is the nonphysical event that has a causal impact in virtue of its nonphysical property of the kind $k$.

While the basic counterfactual theory sees trouble only for specific kinds of overdetermination cases, at first glance INUS conditions must consider all genuine cases as impossible. The implications would be much more severe than in the case of the naïve regularity theory or the basic counterfactual theory. Not only would we have to reject the causal impact assumption (2) and the typicality assumption (3) because all alleged cases of nonphysical causal impact would lead to overdetermination, even the adequacy of physical causes in (4) would be threatened by overdetermination. Proponents of INUS conditions thus would have even stronger
reasons to hold onto (5) when deliberating which assumption to give up. To avoid these consequences, they could postulate that the causal fields involved in explanations citing nonphysical causes and properties are generally different from those citing physical causes and properties. If they are, there is the danger of an equivocation lingering, unless we specify the causal fields involved in the causal claims made by each assumption.

This last option is not one available to the basic probabilistic theory. It denies that the alleged causes in overdetermination cases are in fact causes and sees that as a threat to the adequacy of physical causes in (4) and (6). As the previous theories, it can either be particularly invested in holding onto (5), or it can merge with other theories of causation such as process theories which may provide the resources needed to deliver a more convincing analysis of overdetermination cases.

The basic transference theory seems to affect the exclusion problem in a manner that is quite different from the previous theories. If such a theory understands causation to be the transfer of physical quantities in particular, it can provide support for the completeness assumption (4) and will not give it up lightly. Furthermore, the causal impact assumption (2) and the typicality assumption (3) may turn out to be false, for these require the causal impact of the event in virtue of its nonphysical properties. If such properties can ground the transfer of physical quantities, this consequence may be avoidable, though.

Unlike the naïve regularity theory, the single-difference theory is compatible with there being many events that have an atypical causal impact. While it can endorse (2) and (3), it is unclear whether it would be particularly invested in the typicality of the causal impact of nonphysical events. In addition, if widespread atypical causal impact is possible, we might not be
particularly invested in the completeness assumption (4) when we are asked to give up one or several of the assumptions to avoid exclusion.

Depending on what kind of causal pluralism we endorse, if at least two of the causal relations distinguished from each other are compatible with our notion of causal impact, we may run the risk an equivocation again unless we qualify which causal relation is involved in each assumption. Hall’s distinction between dependence and production did not affect our problem because the notions of causal impact and adequate cause line up nicely with his notion of production. But if a view such as Anscombe’s is right and we have an abundance of different causal relations loosely related by way of family resemblance, it may be rather difficult to get the exclusion problem off the ground.

Finally, causal eliminativism does not seem to undermine the exclusion problem because it seems possible to construct a nomological rather than a causal exclusion problem. How such a nomological version may be affected by different theories of laws of nature may be a topic for another project.
CHAPTER 4
STANDARD RESPONSES TO THE EXCLUSION PROBLEM

4.1 Organizing the Responses

We started our project wondering why many philosophers seem to assume, implicitly or explicitly, that differences among theories of causation do not change the nature or the dynamic of the exclusion problem. Prima facie, it seems reasonable to expect that they at least might have an impact on how we respond to particular assumptions of the problem. If the problem can be resolved only by rejecting one of the assumptions, it would matter if different theories incline us to endorse or reject different assumptions. However, Chapter 3 has shown that, while theories of causation indeed affect the degree to which we might be invested in particular assumptions, they affect the exclusion problem far beyond that.

Some theories allow for, or even suggest, the existence of events that have an atypical impact, but the exclusion problem as we have developed it does not capture such events. From the viewpoint of such theories, the exclusion problem seems incomplete. It might succeed in excluding some nonphysical events and properties; but depending on how widespread we think atypical causes are, this exclusion might not be much of a threat. Other theories of causation imply that such atypical causes cannot exist. From the viewpoint of those theories, the threat of exclusion is much more severe. The typicality assumption (3) also would be redundant and could be omitted.

Some theories would also allow for or imply the existence of different kinds of causal relation. But our version of the exclusion problem seems to utilize only one, based on the notion of causal impact. From the viewpoint of such theories, the problem then might either not be particularly serious or involve an equivocation. It would not be particularly serious if it only applied to cases where nonphysical events enter the same kind of causal relation that physical
events do. But if there are other kinds of causal relation, we might not be particularly invested in the idea that nonphysical events enter the same kind of causal relation to begin with, for they could be said to be causes in another sense. The exclusion problem would involve an equivocation if there were different kinds of causal relation that might qualify as different kinds of causal impact. If we distinguished between these kinds, completeness or overdetermination worries, and by extension the threat of exclusion, might dissolve.

We have thus accomplished what we set out to do, namely, to show that theories of causation indeed have an impact on the exclusion problem. We still might wonder, though, whether this impact in turn affects arguments made in the literature regarding the exclusion problem. This is the task of Chapter 4. There are two parts to this task. First, in many respects, our version of the exclusion problem is more carefully developed than many versions discussed in the literature. If used as an argument for exclusion, where the assumptions (1) and (4) to (6) serve as premises to reject (2) and perhaps (3), our version seems valid whereas most other versions are not. Consequently, we need to test whether some of the responses to the other versions work as responses to our version as well. Second, given the potential impact of theories of causation on the exclusion problem, we should investigate whether they also affect the plausibility of the responses to the problem.

In section 3.1, we established that we could look only at a representative set of theories of causation, for there are too many theories and modifications of theories out there to discuss them all. This is even more of an issue for responses to the exclusion problem. Philpapers.org lists over one thousand publications that include the term ‘exclusion problem’ in their abstract. Consequently, we will need to be very selective. To avoid picking responses at random, we can use taxonomies offered in the literature to make sure we cover at least a reasonable amount of
territory. For example, David Robb and John Heil suggest that many responses fall in one of three categories: identity solutions, autonomy solutions, and inheritance solutions.\(^1\) Put briefly, identity solutions reject the distinctness assumption (2) to prevent the exclusion problem to get off the ground. Autonomy solutions reject the exclusion assumption (6) by arguing that causal explanations involving the physical and the nonphysical tend to be domain-specific and independent of each other in such a way that physical and nonphysical causes do not compete. Inheritance solutions avoid such competition by pointing out that physical and nonphysical properties are intimately related in such a way that nonphysical properties inherit the causal impact, or rather their causal powers, from physical properties.

We will make sure to include at least one alleged solution from each of these categories. However, some responses do not fit into this seemingly neat taxonomy, either because they do not fall into any of these categories or because they cut across them. If some of the alleged solutions respond to particular assumptions of the problem, it may be interesting to see whether such alternative responses explore options that the standard responses do not explore. Finally, some philosophers hold views that could be used as a response to the exclusion problem but that have not yet been applied to the problem. Applying such views may offer insights that have not been addressed in the literature yet.

Given the variety of responses to be covered, the easiest way to structure this chapter is by running through the assumptions of the exclusion problem one by one and discussing particular responses whenever they have something interesting to say about an assumption. Sometimes they may provide reasons for rejecting one; other times they may provide reasons to hold onto one; and yet other times they may just introduce important distinctions that help us to

\(^1\) Robb and Heil (2013, section 6).
get clearer on what the claims are and how the concepts involved relate to other important concepts.

4.2 Distinctness

In some ways, the easiest response to the exclusion problem is to reject that there are nonphysical properties of the kind \( k \) that are distinct from any physical properties. Eliminative materialism claims that standard examples of nonphysical properties such as beliefs, desires, and intentions do not exist. If they do not exist, then there are no events to exclude. Some kinds of realization physicalism claim that seemingly nonphysical properties are in fact identical with their physical realizers. If they are identical, there are no events to exclude either, for any case of alleged nonphysical causal impact really is a case of physical impact.

Eliminative materialists typically focus on mental events and properties in particular. Paul Churchland, for example, argues that we ordinarily explain human behavior by reference to beliefs, desires, intentions, perceptions, and so forth. Such explanations presuppose regularities among mental events and properties cited in the explanations and the behavior to be explained.\(^2\) Sometimes such common-sense laws are used to explain behavior, other times they are used to predict behavior, and yet other times they are modified when they do not work. Given that they are used in this way, they should be considered to be part of a genuine theory, usually called ‘folk psychology.’\(^3\) Unfortunately, this theory is radically incomplete and likely to be wrong. It is incomplete because it offers no explanation of a wide range of important phenomena. For example, the theory is unable to explain the function of sleep or the nature, dynamics, and

---

\(^2\) If the explanatory force of folk psychology indeed rests on regularities, the falsity of any theory of causation that assumes or implies exceptionless regularities would threaten folk psychology. We have made a similar point before, in section 3.3.1, with respect to the typicality assumption (3).

\(^3\) See Churchland (1981, 68f.).
origins of mental illness. In spite of this radical incompleteness, folk psychology has been more or less stagnant for several thousands of years. We are not much better than the ancient Greeks were at explaining human behavior when we exclusively rely on folk psychology. It is likely to be wrong because we have an admittedly incomplete but increasingly powerful set of theories—neuroscience, evolutionary biology, organic chemistry, physics—that connect reasonably well with each other and that provide increasingly nuanced explanations of human behavior in particular and the world in general. But folk psychology does not integrate into this larger corpus of theories and knowledge. The best explanation for this failure of folk psychology may be that the kinds of events and properties it posits do not actually exist, especially if our best theories do not cite them. Therefore, we should eliminate them. If we can eliminate them, we should reject the distinctness assumption (1) of the exclusion problem. As a result, the problem does not even get off the ground, as we stated in section 2.2 when establishing (1).

However, to the extent that eliminative materialism is a defensible philosophical position, it may serve as a response to versions of the exclusion problem that threaten to exclude mental events and properties in particular. But our version is generalized. It threatens to causally exclude all events that allegedly have a causal impact in virtue of their nonphysical properties of the kind \(k\). These properties include those cited by the special sciences that Churchland claims are part of

\[\text{\[4\] Ibid., p. 73.}\]
\[\text{\[5\] Ibid., p. 74.}\]
\[\text{\[6\] Ibid., p. 75.}\]
\[\text{\[7\] Note that this move with respect to the mental events and properties is similar to the one that causal eliminativists make regarding the concept of causation (compare section 3.3.9). Note also that the vocabulary of folk psychology may not be sophisticated and nuanced enough to square well with some of the theories of causation we have discussed. That may be an additional weakness. For example, Salmon’s principle of mark transmission requires the that processes remain uniform in the absence of interactions with other processes. But most of the causal explanations employed by folk psychology are much too crude to include such level of detail.}\]
“the greatest theoretical synthesis in the history of the human race.”

Unlike folk psychology, these special sciences have not been stagnant—on the contrary. Consequently, Churchland could not use his main argument against the existence of mental events and properties as a reason to reject the distinctness assumption (1). Doing so might undermine his reason for rejecting folk psychology in the first place.

There are two obvious ways for Churchland to respond. He could consider the special sciences to be useful heuristics that cite properties that ultimately can be eliminated as well, once physics will have reached a sufficient degree of complexity. Or he could argue that there is an important difference between folk psychology and the special sciences that warrants treating them differently. One such difference we hinted at earlier, namely, that the special sciences connect well with physics whereas folk psychology does not. At least for some properties in the special sciences bridge laws that connect them with physical properties and laws have been established. The existence of such bridge laws is often thought to be an important condition for a successful reduction of one theory to another. While reduction may suggest identity, it does not require elimination. And while at this point in history we have not yet succeeded at reducing all special sciences to physics, we may have good reasons to expect a future science to exhibit such reductive unity.

However, proponents of scientific pluralism reject this picture of science progressing towards a unified theory. John Dupré, for example, claims that the reduction of one theory $T'$ to

---

8 Churchland (1981, 75).

9 See Horgan (1978) for a brief explanation what a reduction by way of bridge laws may look like. However, Horgan suggests that successful reduction does not necessarily commit us to property identity. The classical model of intertheoretical reduction by way of bridge laws is spelled out in Nagel (1961), especially chapter 11.

10 The unity of science view is popular, but it can be understood in various ways. See Cat (2013) for a general overview. See sections 2, 3.1, and 4.1 for a discussion of reductive unity in particular.
another theory $T$ relies on the identification of structurally complex objects cited by $T'$ to structures of simpler objects cited by $T$. For this mapping to work, we need to assume that the objects have essential properties that are both necessary and sufficient for something to belong to a kind that subsumes the objects in question. Within each theory, these essential properties are the ones that fall under laws specific to the theory that then explain phenomena in the respective domain. Dupré argues that many, if not all, special sciences do not assume the existence of such domain-specific essential properties. To explain ecological systems, for example, ecology combines “elements from at least three levels—multicellular organisms, single cells, and molecules (as nutrients in the environment)—but the understanding of such systems typically involves such factors as climate and geology, factors that would have to be assigned to parallel branches.” 11 Such mixing and mingling of objects and properties from different levels of organization to explain events amounts to categorizing and dividing up the world in radically different ways, so much so that it undermines any promise of mapping the properties and laws in the way required for reduction. While different disciplines may categorize and divide up the world in different ways, the causal and nomological relations they posit between events involving different kinds of object are perfectly real. But since they do not map onto those posited by other disciplines, there is no threat of exclusion even if (1) to (3) are true. 12

While scientific pluralism might allow for the completeness assumption (4) and the overdetermination assumption (5) to be true within physics without undermining the causal

---

11 Dupré (1993, 103).

12 Of course, this is a very sketchy summary of a position that is based on the analysis of various special sciences. Specifically, our summary does not explain how we should understand the claim that the world has a pluralistic structure. We also may be sympathetic to a pluralism in the sciences without assuming some kind of ontological pluralism, however that is to be understood. Horst (2007) offers such an epistemic kind of pluralism, which he dubs “cognitive pluralism.” For a more comprehensive interpretation and criticism of Dupré’s view, see Witmer (2003).
impact of nonphysical events in other disciplines, it is not clear why we should think that the kinds of events, objects, and properties which figure into the explanations of various disciplines have to be so different that they never map onto each other. Why think, for example, that biologists and ecologist cannot both explain features of the very same organism? If such is consistent with his scientific pluralism, Dupré has two options. He may admit that this can happen. If these sciences attempt to explain not just the same events and objects but the same properties of these events and objects as well, (4) and (5) might be false. Or he may admit that it can happen but deny that these sciences explain the same properties of the events and objects in question. The latter strategy resembles the dual-explanandum strategy pursued by Terry Horgan, which we will get back to in section 4.6.\(^\text{13}\) The general idea is that causation is a four-place relation which always includes the specific properties of the effect that are necessitated by a cause’s having particular properties. Consequently, different properties of an effect might sometimes be explained by different causal chains. Dupré could then add that these different causal chains might figure into the explanations of different sciences, which is why they can all include true explanations without competing with each other. If that is a possibility, scientific pluralism may indeed be compatible with (4) and (5).\(^\text{14}\)

Regardless of whether we sympathize more with Churchland’s eliminativism or with Dupré’s pluralism, it is important to recognize that these views relate differently to the theories of causation discussed in Chapter 3. Whereas Churchland’s view seems compatible with all the theories, Dupré’s view is incompatible with a good number of them. Since various sciences assume probabilistic laws, both the naïve regularity theory and INUS conditions seem like an

---

\(^\text{13}\) Horgan (1989, 50).

\(^\text{14}\) Dupré himself merely states that his view is not committed to completeness; see Dupré (1993, 216).
unlikely match. And the transference theory might be incompatible depending on what kind of quantities it assumes there are. Dupré himself endorses probabilistic causation as a promising candidate in the context of scientific pluralism.\textsuperscript{15} However, that might not work the way he envisions. Recall that we argued in section 3.3.4 that at least basic probabilistic theories have good reasons to endorse the nonoverdetermination assumption (5). But if scientific pluralism implies widespread overdetermination, assuming that various sciences at least sometimes attempt to explain the same events and properties, probabilistic theories would be at odds with such pluralism. As pointed out at the end of section 3.3.4, we could combine it with other kinds of theory such as process theories, in which case Dupré might be able to offer a consistent view. Still, in the context of the exclusion problem, the incompatibility with several theories of causation might make it less attractive than Churchland’s view, for it will only be a convincing response if it is plausible to begin with and if the theory of causation it relies on is likely to be true. On the other hand, if we deem eliminativism too high a price to pay, we may consider type-identity theories.

Unlike eliminativist materialism, realization physicalism does not deny that there are mental properties. It is just that every “property instance is either an instance of a physical property or a physically realized instance of some functional property.”\textsuperscript{16} Functional properties are specified by way of the causal role they play in a given physical system. For example, to be in pain is to be in a state that is typically produced by some kind of injury, that produces the

\textsuperscript{15} See Dupré (1993, chapter 9).

\textsuperscript{16} Melnyk (2003, 26).
desire to be out of that state, that often leads to wincing, and so forth.\textsuperscript{17} Usually nonphysical properties such as biological, chemical, or economical ones are considered to be functionalizable as well. But the way we generalized the exclusion problem in section 2.3 may yield an awkward-sounding analysis because we opted to contrast physical with nonphysical properties. Applying the functional analysis to the distinctness assumption (1), we then would have to say that instances of nonphysical properties are really just physically realized instances of some functional property. That makes it sound as if we first treat it as an analytic truth that nonphysical properties are distinct from physical properties just to then assert that they are not distinct at all, in spite of the distinctness being analytic. That makes little sense. Recall that we specified in section 2.3 that by “nonphysical events and properties” we meant those events and properties treated by the special sciences as serious contenders for being nonphysical events and properties.\textsuperscript{18} Whether they are indeed nonphysical remains an empirical question. Against that background, the analysis would simply state that allegedly nonphysical properties really just are physical properties or physically realized functional properties.

How strong a response to the exclusion problem this so-called identity solution\textsuperscript{19} is ultimately depends on how defensible realization physicalism is. That discussion would lead us very far astray. Since the exclusion problem as well as the entire project of this dissertation is premised on the assumption that we can make sense of the distinctness claim and should take it

\textsuperscript{17} The example is taken from Levin (2013, section 1). By the way, it would be perfectly fine to refer to other mental states such as beliefs or desires in such functional analyses, for they in turn are physically realized instances of functional properties.

\textsuperscript{18} Melnyk (2003, 32) calls such properties “special-scientific or honorary-scientific types.” According to Melnyk, physicalism does not need to deny the existence of such types, but it does require to treat them as identical with physical types.

\textsuperscript{19} Robb and Heil (2013, section 6) consider identity solutions as one of three standard responses to the exclusion problem, the other two being autonomy solutions and inheritance solutions, to which we will return to in later sections.
seriously, it seems most useful to look at standard responses to the other assumptions of the problem.

4.3 Causal Impact

In section 2.1.2, we contrasted Karen Bennett’s version of the exclusion problem with Stephen Yablo’s exclusion argument to illustrate just how different many of the versions discussed in the literature are. Since there is no canonical version of the problem, we developed our own. But apart from dropping a few hints here and there, we have not yet discussed Yablo’s own response to the exclusion problem. To do so, let us remind ourselves of how exactly he sets up the problem as an argument for epiphenomenalism:

If an event \( x \) is causally sufficient for an event \( y \), then no event \( x^* \) distinct from \( x \) is causally relevant to \( y \) (exclusion).

For every physical event \( y \), some physical event \( x \) is causally sufficient for \( y \) (physical determinism).

For every physical event \( x \) and mental event \( x^* \), \( x \) is distinct from \( x^* \) (dualism).

So, for every physical event \( y \), no mental event \( x^* \) is causally relevant to \( y \) (epiphenomenalism).\(^{20}\)

Yablo’s exclusion premise seems to be a blend of our nonoverdetermination assumption (5) and the exclusion assumption (6), except that he uses the notions of causal sufficiency and causal relevance to state it. He then argues that this exclusion premise is false. From the fact that there is a sufficient cause of an event, it does not follow that no other event or property is causally relevant to it. The mistake, according to Yablo, is to think that physical and mental causes compete. If the first place of a causal relation is already occupied by an event, in this case by a physical event, then no other event can also occupy that place. This view in turn relies on the assumption that the competitors for the place are independent of each other in important

\(^{20}\) Yablo (1992a, 426) and Yablo (1992b, 247–248); italics in the original.
ways. We may be swayed by the dualism premise, for it asserts that the events in question are distinct. Alas, distinctness does not imply independence.

Consider the properties being scarlet and being red. The former is a determinate of the latter, and the latter is a determinable of the former. Being thus related implies that, if anything has the property being scarlet, it necessarily has the property being red as well, whereas the reverse does not hold. Yablo then suggests that there are cases where an event is causally sufficient for another event in virtue of its determinate, while it is also true that, had it not been the determinate of said determinable, it would not have been the property in virtue of which the causal relation holds. Consider a pigeon that has been trained to peck at all and only red shapes. Since scarlet is a shade of red, the pigeon also pecks at scarlet shapes. When it does, the pigeon’s seeing scarlet shapes is causally sufficient for its pecking, and yet it seems the redness of the shapes is what the pigeon is responding to and what causally explains its pecking in light of its past training. But then it seems the redness was causally relevant after all. Yablo then suggests that mental and physical properties may stand in a relation to each other that is similar, though not necessarily identical, to the determinable-determinate relation. Consequently, mental properties, which are like determinables of physical properties, may remain causally relevant even if physical properties are the ones in virtue of which physical events are causally sufficient for other events. Epiphenomenalism thus does not follow from his premises of exclusion, determinism, and dualism.

Let us now see how Yablo’s strategy fares as a response to our version of the exclusion problem. In some respects, adequate causes play a role in our version that is similar to the one

21 Yablo (1992a, 429).

22 Ibid., p. 432.
that sufficient causes play in Yablo’s argument. To be an adequate cause is to be an event such that, if it were cited in a causal explanation, no reference to further events or properties would be needed for the resulting explanation of the event to be satisfactory. We may be tempted to conclude that if an adequate physical cause occupies the first place of a causal relation, it excludes all other events from having a causal impact as well. But this works only if we presuppose particular types of theories of causation. If a nonphysical competitor is thought to have a causal impact in virtue of its nonphysical properties of the kind \( k \), then the typicality of such impact as specified in (3) would indeed imply systematic overdetermination and consequently the event would face exclusion. But remember that in section 2.2—and then later in section 3.3.1 on regularity theories of causation and in section 3.3.6 on single-difference theories—we pointed out that there may be nonphysical causes that have a causal impact in virtue of an atypical kind \( l \). Such atypical causes would not be threatened by exclusion because they don’t necessarily lead to systematic overdetermination. So, from the fact that there is an adequate physical cause, it does not follow that no other events can have a causal impact. It merely follows that nonphysical events of the kind referred to in the causal impact assumption (2) are threatened by exclusion. In that sense, the threat is weaker in our version than in Yablo’s, unless we adopted a type of theory of causation that would not allow for an atypical causal impact of some nonphysical events. The naïve regularity theory seems to be such a theory whereas single-difference theories may not be.

But while causal sufficiency might resemble causal impact to some degree, there is no counterpart to causal relevance in our version of the exclusion problem. Assumption (2) does not merely state that nonphysical events are causally relevant; it claims that they sometimes have a causal impact on physical events. They may not ever be adequate causes like physical events.
sometimes are, but physical and nonphysical events both can have a causal impact; that is, they can both be relata in the same kind of causal relation. Whereas Yablo’s mental events and properties are threatened not even to be causally relevant, which is a more severe threat, nonphysical events and properties in our version are merely threatened not to have a causal impact, i.e., to not be part of a causal chain that leads to the effect. This may be compatible with their being causally relevant in some way, but such relevance must be of a sort that does not undermine the adequacy of the physical causes referred to in the completeness assumption (4).

In this context, it may be important how exactly the notion of ‘causal relevance’ is to be understood and what determinables and, by analogy, mental and some nonphysical properties are supposed to explain. Determinable properties are relevant in that they specify the degree to which the effect lacks sensitivity to the highly specific physical properties of the cause.23 But what does that mean? Yablo himself is not very clear in that respect, partly because he never defines or characterizes ‘causal relevance.’ At times, Yablo seems to affirm that determinate properties explain the causal sufficiency of the cause, albeit in too much detail. That sounds like a causal grounding claim: an event is causally sufficient for another event in virtue of its determinate properties. But how then are we to understand the claim that determinable properties explain as well why an effect occurred, specifically that they are relevant for the effect? We could take Yablo to advocate a dual-explanandum strategy of the sort we will discuss in more detail in section 4.6. While determinate properties explain the determinate properties of the effect, determinable properties explain determinable properties of the effect. But if he indeed adopts a dual-explanandum strategy, he could easily do without his determinable-determinate distinction, as we will see in section 4.6. In fact, even the distinction between causal relevance

---

23 Yablo (1992b, 278).
and causal sufficiency may not be needed. Be that as it may, the dual explanandum strategy may lead to a rejection of the completeness assumption (4) because there are properties of the effect that are not explained by the physical properties of the cause. If causes are adequate only if they explain all the properties of an effect,\(^{24}\) then (4) would turn out to be false.

Alternatively, we might take Yablo to advocate two distinct kinds of grounding relations. Though he himself does not state it that way, this view may be consistent with most, if not all, of his claims. While determinate physical properties ground causal relations, i.e., they are the ones in virtue of which a cause event is causally sufficient for an effect event, determinable nonphysical properties are causally relevant because they ground that grounding relation. It is in virtue of the properties’ being determinates of determinables that the cause event is causally sufficient in virtue of its determinate property. But if determinables play this role in causal explanations, our version of the exclusion problem is guilty of an equivocation because it fails to distinguish between two types of in-virtue-of relation, one that grounds causal relations and one that grounds that grounding relation.\(^{25}\) For our version of the exclusion problem to get off the ground, it needs to interpret the in-virtue-of relation in the assumptions (2) to (6) in the same way.

\(^{24}\) This is a simplification. As explained earlier, an adequate cause is one that, if it were cited in a causal explanation, would be all that we need to cite as a causal factor operative at that time in order to provide a satisfactory explanation for why another event occurs. Such satisfactory explanation would have to locate the cause in time; that is, it needs to identify the event in the causal chain leading to the event the citing of which would yield a satisfactory explanation. Our friend’s tripping on the stairs rather than the big bang qualifies, for example. But our friend’s being in the hospital may involve all kinds of properties we don’t causally explain by way of citing the tripping. Our friend in the hospital may be rather exhausted, but that may be the result of a party the night before and have nothing to do with the tripping. An adequate cause thus only needs to explain those properties of an effect that makes it an effect of that cause but not necessarily all properties.

\(^{25}\) While the charge is reminiscent of the one addressed in section 3.3.7 on causal pluralism, it locates the equivocation elsewhere. Curiously, Robb and Heil (2013, section 6.4) list Yablo as a proponent of an inheritance solution. According to such a solution, some nonphysical properties inherit their causal powers from the property they are intimately related to. The determinable-determinate relation may sound like a good candidate for such intimate relation, but if the suggested analysis is accurate, determinables do not inherit causal powers from determinate properties, for their role in causal explanations is very different. Never mind that Yablo’s view does not seem to presuppose that the causal power theory is correct, at least not as it was discussed in section 3.3.6.
way. Because of that, Yablo would have to reject the causal impact assumption (2) as well as the typicality assumption (3).

However, whether or not we should reject (2) and (3) depends on how plausible it is to treat mental properties, and by extension at least some nonphysical properties, as if they were related to physical properties as determinables are related to determinates. There are at least two problems with this. First, a slight modification of the pigeon example demonstrates that the causal relevance of nonphysical properties does not depend on their being like determinables of physical properties. Second, the description of the pigeon example includes a historical explanation which plays no role in Yablo’s use of the determinable-determinate distinction.

Suppose the color red corresponds to a wavelength range of 620 to 750 nanometers. Now imagine that we trained the pigeon not to peck at all red shapes but only at shapes in two color ranges, for example those that correspond to 620 to 660 and 700 to 750 nanometers. Scarlet falls within the lower of these two ranges. If our pigeon now pecks at a scarlet shape, it would be false to say that it did so because scarlet is a shade of red. Consequently, the determinable being red was not causally relevant. Either we posit that there are determinables that correspond to whatever wavelength range we pick or sometimes determinables are not causally relevant in the way Yablo describes.26 The first option seems to suggest that the pigeon was trained to peck at two distinct determinables rather than at one. One feature of determinables is that they are less specific, i.e., more general, than their determinates.27 It is hard to see how they could then include gaps of the sort needed in our example. But this is a risky response, for we could train the pigeon to peck at shapes of specific colors scattered all over the spectrum so that there is no

26 Falke (2006, 41ff.) makes this point.
27 See Wilson (2017, section 2.1).
determinable that corresponds to this scattered range and that could be causally relevant.

However, the causal explanation of the pigeon’s behavior would remain the same, namely, that
the pigeon pecks at shapes of these particular colors because we trained it to. It is the training,
not the determinable-determinate relation, that grounds the fact that the pigeon pecks at a scarlet
shape. It thus may be good to offer a causal analysis that captures this historical dimension of the
original causal explanation of why the pigeon pecks at scarlet shapes.

In section 2.2, we pointed out that Fred Dretske’s concept of a structuring cause has such
a historical dimension. For Dretske, there is a clear difference between causally explaining why
the pigeon pecks at a scarlet shape at a particular time and why what it did was pecking at a
scarlet shape rather than doing something else.28 Triggering causes, like events that have a causal
impact, explain the former; structuring causes explain the latter. To clarify the distinction,
consider the following example:

A terrorist plants a bomb in the general’s car. The bomb sits there for days until the
genral gets in the car and turns the key to start the engine. The bomb is detonated
(triggered by turning the key in the ignition) and the general is killed. Who killed
him? The terrorist, of course. How? By planting a bomb in his car. Although the
general’s own action (turning on the engine) was the triggering cause, the terrorist’s
action is the structuring cause, and it will be his (the terrorist’s) action, something
he did a week ago, that will certainly be singled out, in both legal and moral
inquiries, as the cause of the explosion that resulted in death.29

While structuring causes provide the conditions for triggering causes to do their work,
they are not sufficient for the effect; they depend on later events such as turning on the engine.
Since they depend on later events to conspire in the right sort of way for the effect to occur, they
don’t involve regularities in the way triggering causes do.30 Furthermore, structuring causes can

30 Ibid., p. 123. There is an interesting parallel to Mackie’s INUS conditions here, which we discussed in section
3.3.3. Recall that one of the reasons for establishing INUS conditions was that the cause was itself not sufficient for
persist through time in ways triggering causes do not. Each peck at a scarlet shape is explained by a different triggering cause, the perception of each individual color and shape. But the structuring cause explains all of these instances, which the training of the pigeon does as well. It then looks like Dretske’s distinction is more useful than Yablo’s. The consequences, however, seem to be the same. If the causal impact assumption (2) and the typicality assumption (3) cite structuring causes rather than triggering causes, our version of the exclusion argument involves an equivocation. Structuring causes explain why certain physical events cause their physical effects in virtue of their physical properties. Since they are relata in a different kind of in-virtue-of relation than the one that figures into assumptions (4) to (6), we should reject (2) and (3) as they stand.

But this strategy seems promising only under two conditions. First, the way nonphysical events and properties figure into special-science explanations need to match the way structuring causes support triggering causes in Dretske’s view. It is not clear whether that is so. Consider our friend with the broken ankle again. She may have a background belief that whenever she experiences serious pain she should go to the hospital. Such belief may serve as a structuring cause, for it would explain every instance of her going to the hospital when she is in severe pain. But at least in some cases the pain itself seems to be a triggering cause, even though there are

---

Dretske’s distinction resembles Lombard’s distinction between enablers and triggers. Both structuring causes and enabling conditions depend on triggering causes to set off the causal chain leading to the effect; both have a historical dimension that Yablo’s determinables lack. But Lombard’s enablers are closer to what we ordinarily would call background conditions; see Lombard (1990, 203ff.). However, Sartorio (2009, 582f.) points out that one could view enablers as causes. If that is right, the difference between structuring causes and enabling conditions might blur further.
adequate physical triggering causes, according to the completeness assumption (4). Of course, this is not yet a reason to reject Dretske’s strategy. But we would need to see whether the nonphysical events cited by the special sciences are usually treated as structuring causes or not. Second, even though structuring causes figure into a different in-virtue-of relation, this may undermine the degree to which physical causes are adequate. If a satisfactory explanation of a physical event requires not only reference to the physical cause but also a grounding explanation of the sort that structuring causes are supposed to provide, then we need to reject (4) if we want to hold onto (2) and (3).

4.4 Typicality

Since the typicality assumption (3) is not usually part of the exclusion problem as discussed in the literature, there are no standard responses to it. As a starting point for a discussion, let us briefly remind ourselves of the contexts in which typicality came up in earlier sections. We introduced (3) because we did not want to make the overdetermination assumption (5) too strong. In its strongest form, it would claim that there is no overdetermination. In the light of possible overdetermination examples such as the double-assassination case, we opted to merely reject systematic overdetermination. This had us worried that nonphysical events may have an atypical impact such that, if and when they have it, they would overdetermine but not in a systematic way. The typicality assumption was a way to ensure that the causal impact of the nonphysical event would not be atypical. Instead, it would be representative of events that have a causal impact in virtue of a nonphysical property of kind $k$. Any such impact would then yield systematic overdetermination, or so it seems.

Before we introduced the assumption in section 2.2, we drew an analogy that is worth getting back to at this point. To explain why atypical causes are not worrisome, or why typical ones are, we considered a philosophy class in which only students with yellow socks got an A.
This may not be surprising, especially if only one student happened to wear yellow socks all the time. It seems like a mere coincidence. Similarly, if some nonphysical events sometimes have a causal impact on a physical event which already has an adequate physical cause, the overdetermination might simply be due to an accidental arrangement of external circumstances, as we pointed out in section 3.3.1. But if twelve out of twenty students got an A and all of them wore yellow socks and no other students wore yellow socks, we would wonder whether there is something about the grading procedure that connects the color of socks student wear and their assessment by the teacher. Similarly, if nonphysical events of a certain kind and only events of that kind have an overdetermining causal impact on physical events, it may suggest that the overdetermination is not just due to the accidental arrangement of external circumstances. That is, we will begin to suspect that the overdetermination has something to do with the nature of the nonphysical events or perhaps the nature of the causal relation they enter.

Any theory of causation that implies regularities will exacerbate the oddity of this picture. Single-difference theories, however, could simply reject (3). Perhaps there are some kinds of nonphysical event that do have a typical causal impact, but those may indeed be threatened by exclusion qua (4) and (5). But since (5) allows for nonsystematic overdetermination, a proponent of a single-difference theory might claim that nonphysical events that have a causal impact tend to be atypical. For example, there may not be laws that events of their types fall under. While this is an option, we now may wonder what exactly single-difference causes explain. In section 2.2, we considered the example of our carrying an umbrella. The thought that it might rain today caused our carrying an umbrella. If the completeness assumption (4) is true, then there was also a physical event—in our brain, for example—that caused our behavior. Fortunately, we are not facing systematic overdetermination because the thought is supposed to be atypical. But that
amounts to saying that normally, or in other circumstances, it might not have led us to carry an umbrella. But then the reference to the thought loses all explanatory force. So, while the single-difference theory may avoid the threat of systematic overdetermination and thereby avoid causal exclusion, this comes at a high price. The causal explanations involving such atypical causes lose all explanatory force, or so it seems.

4.5 Completeness

In section 2.1.2, we pointed out that some versions of the completeness assumption are too strong. Karen Bennett, for example, phrased it as the claim that every physical occurrence has a sufficient physical cause. Thus stated, it may be false according to our best physical theories, for there may be physical events that have no cause. To make sure that the assumption cannot be shot down easily, we conditionalized the assumption. That is not the only way to do it, of course. Barbara Montero, for example, suggests phrasing it as follows: “Every physical phenomenon that has a sufficient cause has a sufficient physical cause.” She also cautions us not to interpret the assumption of completeness—or closure, as she calls it—as stating that physical effects have only physical causes. In the context of the exclusion problem, this would not only make the nonoverdetermination assumption redundant but also would imply a stronger claim, namely, that overdetermination, even of the nonsystematic kind, does not exist. Our version of the assumption would have such strong implications only in combination with theories of causation such as the naïve regularity theory and some kinds of probability theory. Finally, she warns not to confuse the completeness assumption with the claim that there is no causal

33 Montero (2003, 174).
interaction between the physical and the nonphysical.\textsuperscript{34} After all, completeness is compatible with physical events causing nonphysical events, as long as such causation does not violate some condition for closure, say, conservation laws or the like.\textsuperscript{35}

While our version of the completeness assumption (4) avoids all the pitfalls that Montero cautions against, it does not include what Matthew C. Haug calls the fundamental category completeness assumption: “Physics provides a complete inventory of the fundamental kinds and categories that are sufficient for a realization basis for all of the events that causally impinge on the physical universe.”\textsuperscript{36} According to Haug, this is a distinct completeness claim that physicalists are committed to. While this assumption is not needed for the exclusion problem, we may ask what impact it would have on it. If we are willing to endorse the completeness assumption (4), for example because we have physicalist inclinations, then we should at least be sympathetic towards the fundamental category completeness assumption as well. Let’s call it assumption (7).

If (7) is true and we wanted to endorse the transference theory, we would be bound to restrict the quantities to be transferred to physical quantities. In that case, however, it looks like the impact assumption (2) is threatened, though not necessarily rejected. Remember that we stated in section 3.3.5 that the quantities transferred are not necessarily the properties in virtue of which the transfer takes place. So, we could maintain the causal impact of nonphysical

\textsuperscript{34} Ibid., p. 175.

\textsuperscript{35} Montero hints at this when she discusses the argument for the causal closure from the success of science. Roughly, if the completeness assumption (4) were false, physics would necessarily be incomplete, but since it is not necessarily so, (4) is true. This would be a stronger claim than even physicalists would need to endorse. See Montero (2003, 177f.).

\textsuperscript{36} Haug (2009, 382). There, he explains further that the realization basis “is a set of categories that characterizes the properties/events that realize a given set of phenomena.” In this case, the “phenomena in question are all those that causally affect the physical universe.”
properties, but only by way of a transfer of physical quantities. Essentially, this is a form of inheritance solution. Typically, such inheritance solutions argue by way of a particularly intimate relation between nonphysical events and properties and their physical realizers. But notice that we did not need to specify this relation beyond the fundamental category completeness assumption. The inheritance flowed from a particular theory of causation in combination with that assumption. But if we presupposed mere counterfactual dependence, for example, we would need to spell out the relation between nonphysical events and their realizers more so as to arrive at that view.

When weighing the plausibility of the assumptions of the exclusion problem against each other, we may find that the motivations underlying some of the assumptions come with additional commitments which in turn may be more or less compatible with particular theories of causation.

### 4.6 Overdetermination

Horgan claims that causation is really a four-place rather than a two-place relation. Rather than being a relation between events only, it is a relation between events and particular properties they have:

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 $G$.\(^{37}\)

In our version of the exclusion problem, we specified only the property of the cause that grounds the causal relation between $c$ and $e$. Horgan’s analysis suggest that this picture is incomplete.

\(^{37}\) Horgan (1989, 50).
Such a property does not ground the causal relation simpliciter but only the causal relation with respect to a particular property of the effect. If the completeness assumption (4) is correct, there is a physical event that grounds the relation between $c$ and $e$ in virtue of a physical property, say $F$. And if the causal impact assumption (2) is correct, there is an event that grounds the relation in virtue of a nonphysical property, say $N$. But notice that neither claim specifies the properties of the effect that are supposed to be explained. Assuming that the nonphysical event in (2) is distinct from $c$, say $d$, it could be the case that $c_F$ causes $e_G$ whereas $d_N$ causes $e_H$. But that is an instance neither of overdetermination nor of exclusion.

For this strategy to be a good response to our version of the exclusion problem, we need to assume that being an adequate cause does not entail that $c$ satisfactorily explains all properties of $e$, for then we get both overdetermination and exclusion. That may be a reasonable assumption. Physical explanations usually come with a much higher degree of specificity than explanations in the social sciences, for example. That is, many explanations involve a certain degree of abstraction. Physics might consider a causal explanation of an event satisfactory to the degree that it explains all highly specific physical properties of that event. And if the physical cause referred to in the completeness assumption (4) does that, it should qualify as an adequate cause. Notice, however, that this does put constraints on the kind of property of effects that nonphysical events can causally explain in virtue of their nonphysical properties: either abstract physical properties not captured by the adequate cause or nonphysical properties.

---

38 It could also be identical with $c$ so that $c$’s nonphysical properties explain the occurrence of $e_H$ rather than of $e_G$. As mentioned in section 2.1.2, this distinction might partly depend on which account of events we endorse. A Kimean account might force us to deny the identity of $c$ and $d$ whereas a Davidsonian account could easily capture this.
Abstract physical properties may seem odd from the viewpoint of probability theories. The adequate physical cause would necessitate the probability of the effect’s occurring with all its physical properties. Abstract physical properties are intimately related to the particular properties, perhaps they stand in a determinable-determinate relation to each other. But if they do, why would we think that they would have any impact on the probability of the effect? For the nonphysical event in (2) to be causally related to the abstract-physical effect, it needs to be make a difference to the probability of the effect. But if it indeed does, it is questionable that we were right in considering the physical cause as adequate. If it is not, the completeness assumption (4) is threatened.

Alternatively, consider transference theories. If the quantities transferred must be physical, the physical cause in (4) must satisfactorily explain all the quantities transferred. But then there is no $eH$ to relate to and to explain $dN$. For Horgan’s strategy to succeed, he would have to assume that there are nonphysical quantities that transfer to $e$. If they can convert into physical quantities, this will probably undermine (4). If they cannot convert into physical quantities, nonphysical events could explain only nonphysical properties of events. To avoid some strange parallel between the physical and the nonphysical, it may be most reasonable to assume a different kind of causal relation at work here—counterfactual dependence, for example. That is, we might move towards a Hall-type causal pluralism.

### 4.7 Exclusion

The way the exclusion assumption (6) is phrased is somewhat suggestive. It makes it sound as if any nonphysical event will have to compete with an adequate physical cause. If causation is a two-place relation, and if the first place is already taken by an adequate physical cause, the nonphysical event has no place to occupy. However, we might question this picture and advocate that physical and nonphysical descriptions are relatively autonomous. For example,
neither psychology nor physics may describe the world as it is. Both describe it relative to a
language and its terms and categories. To put it crudely, the language of psychology carves up
the world differently than the language of physics and it does so in such a way that no systematic
mapping of the categories provided by each language is possible.\textsuperscript{39}

For Donald Davidson, that is not a problem. As we saw in section 2.1.1, he thinks that we
use different descriptions of events as a way to refer to them. Even two radically different
descriptions that refer to two distinct sets of properties still may refer to the same event. But
reference to these sets of properties merely serves to pick out the event in question, they are not
part of the identity conditions of an event as they are in a Kimean account. Confusion arises if
we then cite some of the properties to ground causal relations, as we did in previous sections. For
Davidson, an event is not a cause of another event in virtue of any of its properties. Rather, the
event’s standing in particular causal relations to other events is part of the event’s identity
conditions.\textsuperscript{40} What makes events stand in cause-effect relations to each other is the fact that they
fall under strict laws.\textsuperscript{41} However, laws for Davidson are linguistic in nature. As such, events
relate causally only under a specific description.\textsuperscript{42} Since some descriptions differ radically,
events may relate causally under one description but not under another. So, even if a causal
statement such as the completeness assumption (4) refers to physical cause of an event, nothing

\textsuperscript{39} Davidson (1974/1980, 230f.) argues that there cannot be any bridge laws connecting the mental and the physical
partly because of this. Davidson’s arguments and theories are not specifically directed towards the exclusion
problem. Nonetheless, his anomalous monism is often used to establish the autonomy of psychological causal
explanations from physical ones. Jackson (1996, 380) presents Davidson as the main proponent of one of three
important kinds of autonomy solutions that he then criticizes. Notice the similarity between Dupré’s scientific
pluralism which we discussed in section 4.2 and Davidson’s anomalous monism in this respect.

\textsuperscript{40} See Davidson (1969/1980, 179).


\textsuperscript{42} Ibid., p. 215.
seems to follow for possible causes of the same event under a different description, such as a psychological one. In that sense, causal explanations involving the physical and the nonphysical are autonomous. Consequently, Davidson would simply reject the exclusion assumption (6).

Note, however, that this response is semantic and epistemological in nature.\textsuperscript{43} It shifts the focus from what happens in the world to the way we talk about the world and how we explain what is going on.\textsuperscript{44} It remains unclear what such a theory implies metaphysically. We could, for instance, argue that the very concept of causation is as relative to language as the other categories are and maintain that there is no such thing as causation on a metaphysical level, i.e., we could be causal eliminativists. On the other hand, if we are already monists of some sort, we could also argue that, although we distinguish between physical and mental events, the events are really of one kind, neither physical nor mental in nature. But if they are, the problem of exclusion seems to vanish just as it vanished for the identity theorists in section 4.2.

In sections 1.1 and 3.2, we hinted at a more promising strategy to put pressure on the exclusion assumption (6). We mentioned that it is often the conversational context that determines what kind of causal explanation we are after. Menzies takes this very seriously and points out that the context often determines which property values are taken to be default values. These default values may then provide the background against which we specify which properties of the effect we refer to in our causal explanations. Menzies gives the following example:

If someone asserts a statement like “Giving the patient 100 mg of penicillin cured him”, the contrastive focus serves to highlight the fact that only certain dosages

\textsuperscript{43} This is where the similarities between Davidson and Dupré end and where the former’s view starts to resemble Horst’s more.

\textsuperscript{44} This is a point frequently made against Davidson, or at least, if not intended as criticism, it is emphasized that Davidson’s anomalous monism is merely an epistemological position. See, for instance, Yalowitz (20012, section 5.3).
within a range of possible dosages were causally effective. It suggests that we should take the cause variable to be a quantitative variable, which can take various values such as 0, 50 mg, 100 mg, 200 mg, and so on. And that the causal statement is to be understood as saying something along the lines “Giving the patient 100 mg of the drug rather than some other dosage in the given range caused the patient’s recovery rather than non-recovery.”

Since such assumptions often remain in the background of the conversational context, the contrastive structure of causal statements often is obscured. Rather than merely stating that $c$ causes $d$, we should state that $c$ rather than $d$ causes $e$ rather than $f$. Rather than stating that $c$ caused $d$ in virtue of property $P$, we should state that event’s $c$ being $P$ rather than $P^*$ causes event $d$’s being $Q$ rather than $Q^*$. Explicating the contrastive structure of causal statements can avoid confusions and problems.

Consider the pigeon from section 4.3 which was trained to peck at red shapes. According to Menzies, we need to consider two causal statements: (a) The shape’s being red rather than not red made the difference to the pigeon’s pecking rather than not pecking. (b) The shape’s being scarlet rather than not scarlet made the difference to the pigeon’s pecking rather than not pecking. Interestingly, (a) is true whereas (b) is false. Had the shape not been red, the pigeon would not have pecked for it. But had it not been scarlet but crimson instead, the pigeon might still have pecked at it, as long as it was a shade of red. Without having to employ the distinction between causal relevance and causal sufficiency, as Yablo did, and merely relying on the notion of making a causal difference, we can see that determinates and determinables are not in causal competition. We thus may have reasons to doubt the exclusion assumption (6).

---

45 Menzies (2008, 208). Menzies account introduces several elements foreign to our terminological convention. For example, he talks about values rather than properties and about interventions on variables, which he takes to be “an idealized manipulation that sets the value of $X$ by a causal process that is independent of all other possible causes of $Y$”; Menzies (2008, 207). He also talks about properties causing each other, which is not in line with our convention. We will simplify his account by adjusting it to our terminology to make it applicable to our discussion more easily.

46 Ibid., p. 209.
While Menzies relies on the notion of making a causal difference, his strategy may also work in combination with our notion of ‘causal impact.’ According to the causal impact assumption (2), the following contrastive statement may be true: (a) a particular nonphysical event’s nonphysical properties of the kind $k$ rather than properties of the non-$k$ kind (or physical properties) are those in virtue of which that event has a causal impact on a physical event’s properties. Similarly, according to the completeness assumption (4), the following contrastive statement may be true as well: (b) a particular physical event’s physical properties rather than nonphysical properties are those in virtue of which that event has a causal impact on a physical event’s properties. However, to apply Menzies contrastive account properly, we would also have to specify the properties of the effect as well as the contrast properties. The issues here seem reminiscent of the ones we encountered in section 4.6 when we discussed Horgan’s dual-explanandum strategy. It seems we must assume that the properties of the effect in (a) must not be the same as the properties of the effect in (b), or else we the nonphysical cause in (a) faces exclusion. But if the properties to be explained must be nonphysical rather than physical, (2) and (3) might turn out to be false, at least if we understand the reference to the physical event as to mean that the event’s physical properties rather than nonphysical ones are to be explained causally. For that, (b) seems the better candidate. By the same token, if we specify that the properties of the effect in (2) are nonphysical, then this will restrict the properties that adequate physical causes can explain causally.

### 4.8 Chapter Conclusion

The goal of this chapter was twofold. First, we wanted to see how some of the standard responses to the exclusion problem fare as a response to the particular version we developed in Chapter 2. Second, we wanted to check whether theories of causation also affect these responses in some way, especially given that Chapter 3 established that theories of causation indeed affect
the exclusion problem in various ways. Because thousands of articles have been written about the exclusion problem, we were forced to be selective, but we tried to cover some of the most commonly discussed as well as some lesser-known responses to cover a reasonable amount of territory. These were the main lessons:

Eliminativist may deny the distinctness assumption (1), but at least one argument against distinctness does not work as it stands against our particular version of it. If the main reasons for eliminating mental properties are that the theories citing them have not improved much over time and that they do not integrate well into the set of our most successful scientific theories, then many nonphysical properties that play a role in the special sciences remain untouched, for the special sciences have progressed and they integrate better. If those properties are to be eliminated for some other reason, we would have to rethink the status of a large body of theories. Alternatively, the eliminativist can embrace a type-identity theory for nonphysical properties. Adopting such a theory may be based on the hope that all the kinds of events and properties cited by the special sciences will eventually be successfully reduced to physics. The prospect of such unity of science, however, may be dim, especially if there are reasons to think that scientific pluralism is true. Depending on how such pluralism is construed, it may imply that the completeness assumption (4) and the nonoverdetermination assumption (5) are false. While eliminativism and type-identity theories are compatible with all theories of causation discussed in Chapter 3, scientific pluralism may not be compatible with some regularity theories and INUS conditions. Even worse, there may be a tension with probabilistic theories even though proponents of scientific pluralism sometimes explicitly endorse them.

Some philosophers read the causal impact assumption (2) in a weaker way. They respond to the exclusion problem by distinguishing between causal relevance and causal sufficiency and
then claim that nonphysical events are causally relevant to but not causally sufficient for their effects. Unlike causal pluralism, these are not two distinct causal relations. Rather, the former is a grounding relation of the latter. Our version of the exclusion problem, however, specifies that some nonphysical events have a causal impact, which is a production relation of sorts. Because of that, the strategy seems to amount to little more than a rejection of (2) and (3). Since these are the key assumptions of the problem, the strategy is only as good as the reasons for endorsing the suggested distinction between causal relevance and causal sufficiency. It turns out that at least Yablo’s reasons for it are not convincing. The distinction between structuring causes and triggering causes, on the other hand, is more plausible in some respects, but it may lead us to reject the completeness assumption (4). Furthermore, it may be unclear whether counterfactual theories of causation can capture the distinction.

If the typicality assumption (3) is correct, then many nonphysical events that have a causal impact in virtue of their nonphysical properties are overdetermining causes. Any theory of causation that implies regularities will exacerbate the oddity of such a picture. Single-difference theories may point out that assumptions (1) to (3) are compatible with there being atypical nonphysical causes that are not threatened by exclusion qua the nonoverdetermination assumption (5). But the price may be that citing such atypical causes loses all its explanatory force.

The reasons for endorsing the completeness assumption (4) may be the same as those motivating another assumption, which is related to completeness but distinct from (4). That further assumption is not needed for the exclusion problem to arise, but it may have an impact on which of its assumptions we are likely to endorse or reject. In combination with a transference theory of causation, for example, it may imply that the causal impact assumption (2) is false. In
combination with a regularity theory of causation, to consider another example, it may force us to spell out the relation between nonphysical and physical events more to decide whether or not they face exclusion.

Even if the overdetermination assumption (5) is true, the threat of overdetermination might be averted by a dual-explanandum strategy. One such strategy construes causal relations as four-place relations. Once we realize that causes rarely explain all properties of their effects, and once we concede that even adequate causes might not do so without ceasing to be adequate, it is easy to see how some nonphysical events avoid overdetermination and exclusion. They explain different features of effects than adequate physical causes do. Depending on what we think these features are, the resulting view may be incompatible with some theories of causation such as the probability theory. And in combination with the transference theory, the strategy may imply rather implausible views such as parallelism.

The exclusion assumption (6) suggests a picture according to which nonphysical events compete with physical events for their place in causal relations. Autonomy solutions question this picture. They may do it in the form of a dual-explanandum strategy or in the form of scientific pluralism, but they may also do it based on views about language or the theory-dependence of descriptions. If the categories and distinctions that causal descriptions and explanations rely on are relative to frameworks, either of a semantic or of a theoretical sort, then the fact that there is a description according to which a physical event is an adequate cause of an event does not imply that descriptions according to which nonphysical events are causes of the same event are false, for the tension only arises within one framework or theory. However, it is hard to see to what degree this response is ontological in nature, for language and theory dependence seem mostly epistemic issues. The most promising autonomy solutions may thus be
the dual-explanandum strategy and scientific pluralism, for they are both ontological in spirit. Menzies’ contrastive account of causation, however, seems to have more interesting consequences. If we specify the properties of the effect to be explained by the physical and nonphysical event in the right way, we may avoid the causal exclusion of nonphysical events and properties. However, the view does not seem helpful if we insist that nonphysical events sometimes explain some of the same properties of an effect that physical adequate causes explain.
CHAPTER 5
REASSESSING THE EXCLUSION PROBLEM

5.1 Striking a Balance

Chapter 3 and Chapter 4 demonstrated that theories of causation affect both the exclusion problem and some of the standard responses to the problem. Sometimes they affect the strength of particular assumptions; at other times they affect the likelihood of rejecting particular assumptions. Since our version is importantly different from versions often discussed in the literature, as illustrated in sections 2.1.2 and 2.2, we may worry that perhaps it is so strong that it includes a number of unreasonable assumptions. While our analysis serves to show that there is one version of the exclusion problem that is affected by theories of causation and to which some standard responses do not work well, it may not be a version we should endorse. But if we should not, rejecting otherwise promising responses based on it would be misleading as well as unfair to the responses. To address these worries, we will review a stronger and a weaker version of the exclusion problem. These versions involve assumptions that we have addressed in previous chapters. We can then explain why it is reasonable to hold onto our version. Doing so will provide a background against which to honestly assess the strengths and weaknesses of our version. However, we will see that, all things considered, our version seems to strike a good balance between specificity and strength on the one hand and the avoidance of unnecessary ontological commitments on the other hand. The latter is especially important to allow for and to ease the discussion of the exclusion problem among philosophers who differ in their views on causation, completeness, events, overdetermination, and the mind. A version that accomplishes such balance thus seems particularly well-suited for future research. We will end this chapter and dissertation with a brief summary of the main findings and highlight some of the loose ends that may require attention in later projects.
5.1.1 Strong Exclusion

In previous chapters, we encountered various versions of the assumptions of the exclusion problem. They differed in the way they were phrased, in their level of specificity, in their strength, and in their ontological commitments. Let us look at a particularly strong version that we may construct based on some of the assumptions we have looked at.

A. Distinctness: Nonphysical properties are distinct from physical properties.

B. Causal Sufficiency: Sometimes nonphysical events are causally sufficient for physical events.

C. Completeness: Every physical event has a sufficient physical cause.

D. Nonoverdetermination: Events are not overdetermined.

E. Exclusion: If an event has a sufficient cause, it has no further cause.

This version seems much simpler than our version. It is easy to see how it may be used as an argument against (B). If every physical event has a sufficient physical cause, and if events do not have more than one sufficient cause because there is no overdetermination, then nonphysical events cannot be causally sufficient for physical events. The argument also appears to be valid.

While both simplicity and validity are virtues, these virtues do not have much weight unless we have good reasons for endorsing the premises. Unfortunately, there are very good reasons for rejecting several of them.

Assumption (A), a version of which we discussed in section 2.2, is extremely general and because of that does not allow for any nuance. For example, we may endorse a type-identity realization physicalism of the sort discussed in 4.2 for many nonphysical properties cited by various special sciences, but we may want to reject it for particular kinds of mental properties. (A) would force us to treat them in the same way. But why should we accept that? By contrast, our distinctness assumption (1) allows for such nuanced discussion of particular kinds of
nonphysical properties, which seems particularly important in the context of a generalized problem.

In section 4.5, we discussed a version of (B). The first thing to notice is that sufficiency is such a modal notion. Not all theories of causation posit such modal relation between cause and effect. In fact, theories such as the naïve regularity theory in section 3.3.1 were specifically designed to avoid modality. Thus understood, (B) rules out the application of some theories of causation to the exclusion problem. At the very least, (B) makes stronger claims about how cause and effect are related than some kinds of regularity theories, probabilistic theories, and single-difference theories. But if the problem rests on positing a relation that is stronger than these theories posit, what reason would a proponent of such a theory have to endorse it? If he or she has to give up one assumption to ensure consistency among the others, (B) may be a promising candidate. Since probabilistic theories of causation are often considered to be those that are consistent with current science, we should make sure to phrase the exclusion problem in a way that is consistent with the application of such theories. Assuming that a sufficient cause is one that raises the probability of another event to 1.0, i.e., assuming that sufficiency is a special case of probabilistic causation, the problem is not particularly severe from the viewpoint of such a theory, for it does not capture most instances of causation. Our causal impact assumption (2) managed to do that without undermining the validity of the argument and the threat of exclusion. But it required, especially in the light of a weaker nonoverdetermination assumption, the introduction a term of art, ‘causal impact,’ and the typicality assumption (3), which has no equivalent in the strong version. So, we had to trade some simplicity for a higher degree of compatibility with theories of causation.
In sections 2.1.2 and 4.5, we mentioned a particularly strong version of the completeness assumption along the lines of (C). The problem is that even a physicalist is not committed to that. After all, our best physical theories posit that quantum fluctuations are uncaused. Since these are not events in the ordinary sense of the word, they may not count as proper counterevidence. But cosmological theories are compatible with the claim that the big bang was not caused, for example.\textsuperscript{1} By contrast, our completeness assumption (4) accommodates this nicely.

In section 2.1.2, we addressed a nonoverdetermination assumption along the lines of (D). (D) may be true. As we saw in several sections of Chapter 3, several theories of causation struggle with providing a coherent and plausible analysis of genuine overdetermination cases. However, some theories of causation are at least compatible with there being nonsystematic cases. The question, then, is whether we should exclude isolated instances by turning (5) into (D). The answer should not be based on what our stance on overdetermination is but on the role the assumption plays in the problem and in resulting arguments. As a general rule for constructing arguments, “make your premises as strong as necessary for the argument to work, but no stronger” is good advice. For the stronger the assumption, the more likely it is that there are counterexamples. Minimizing our ontological commitments is a good strategy for the very same reason.

Finally, (E) is implausibly strong. Even proponents of INUS conditions point out that causes on their own rarely, if ever, suffice. An event’s having more than two causes is probably the rule rather than the exception. Of course, that does not negate (E), since it is phrased as a conditional. But if we follow the earlier advice that (B) should be modified so that it allows for

\textsuperscript{1} Montero (2003, 174) uses the big bang as an example as well.
probabilistic causation and we should go with a weaker nonoverdetermination assumption (5) rather than (D). (E) needs to be adjusted accordingly.

It seems, then, that we should keep the assumptions strong enough for them to serve as premises in a valid deductive exclusion argument. But they should not be so strong that they face counterexamples, either conceptual or empirical. The assumptions (1) to (6) in our version are weaker than the assumptions (A) to (E) in the strong version. However, we may ask whether perhaps we should endorse yet weaker assumptions.

5.1.2 Weak Exclusion

While adopting weaker assumptions may be beneficial in that they may require fewer ontological commitments, they may also undermine the potential validity of the problem when it is used to construct an exclusion argument. If, for example, we weakened all the sufficiency claims by replacing them with causal relevancy, the problem seems to disappear:

A. **Distinctness:** Nonphysical properties are distinct from physical properties.

B. **Causal Sufficiency:** Sometimes nonphysical events are causally relevant to physical events.

C. **Completeness:** For every physical event, there is a causally relevant physical event.

D. **Nonoverdetermination:** Events are not overdetermined.

E. **Exclusion:** If there already is a causally relevant event to another event, there is no further causally relevant event.

Typically, to say that an event is causally relevant is taken to be a weaker claim than one about its causal impact or causal sufficiency. For example, even if we think that the presence of oxygen is a mere background condition rather than a partial cause of an explosion, we still may say it was causally relevant. Without the oxygen, the explosion might not have happened, after all. But if that is right, the exclusion assumption (E) seems clearly false, for it implies that the cause could not have been relevant in the presence of the background conditions. That makes
little sense. But if that is right, then there seems to be little reason for thinking that nonphysical events cannot be causally relevant whenever there is a physical event that is also relevant. That is, the problem ceases to be one.

One significant difference between both the weak and the strong versions, on the one hand, and ours, on the other hand, seems to be that our version is much more specific. It relies heavily on in-virtue-of talk to specify the properties that ground the causal relations. Notice that in neither the strong nor the weak version is it clear what feature of an event explains that it is a cause of the effect. It may be suggested that it is the physical features of physical events and the nonphysical features of nonphysical events, but that does not follow from anything actually stated. Inheritance solutions sometimes exploit such vagueness in pointing out that nonphysical events may inherit their causal powers from their physical realization base. That may be so, but notice how, once we specify the properties in virtue of which events cause their effects, it becomes much clearer what that strategy amounts to. For nonphysical events and properties to inherit their causal powers from physical events and properties, there has to be a rather tight relation between the nonphysical and physical properties in question. If the strategy is merely to postulate that nonphysical events have a causal impact in virtue of their physical properties, it could not serve as a response to our version of the problem. It seems, then, that trading further simplicity for the sake of explanatory power is reasonable. Doing so will clarify the connections between the assumptions, and it will allow us to explain better how responses to the exclusion problem are supposed to work.

Of course, none of this establishes that our version is the only one available or even the one that best achieves a good balance between specificity and strength on the one hand and the
avoidance of unnecessary ontological commitments on the other hand. But it does seem like our version is at least one of them.

5.2 Conclusion

We began this project because we were puzzled by how few philosophers analyzed and utilized theories of causation in the context of the exclusion problem, even though most of its assumptions include causal claims. It seemed that there is an implicit, and sometimes even an explicit, assumption that theories of causation do not affect the problem at all. While a few philosophers have tried to use the counterfactual theory in particular to develop a response or a solution to the problem, to date there is no systematic survey that would justify the assumption that theories of causation generally do not matter. This dissertation had the rather humble goal to investigate this assumption. Had it turned out that in fact theories of causation have no impact on either the exclusion problem itself or on how we respond to it, we would have provided support for an assumption that seems to underlie the debate. But, fortunately, our investigation revealed quite the opposite. There are many ways in which theories of causation affect the problem.

To the extent that we do not have any obvious reasons to reject a particular assumption of the exclusion problem, we have to weigh their plausibility against each other to decide which assumptions should be rejected in order to resolve the tensions or inconsistencies among them. As it turns out, theories of causation often affect how invested we should be in holding onto particular assumptions. If, for example, we presuppose a regularity theory of causation, we cannot allow for atypical causes. So, the typicality assumption, under such theory, cannot be given up since it is implied by the theory. Single-difference theories, on the other hand, may accept the assumption, but they will not be as invested in it, for they can handle atypical causes quite well.
At other times, a theory may require stronger assumptions than others do. For example, theories that run into inconsistencies when analyzing overdetermination cases, as some kinds of probabilistic theories do, might not only endorse the nonoverdetermination assumption (5) but also may even imply that it is not just systematic overdetermination that is a problem. No overdetermination exists. That in turn has consequences for how to think about the severity of the problem as a whole. As we have set up the problem, it allows for atypical nonphysical causes that nonsystematically overdetermine physical events. That is, the scope of the problem is somewhat limited, for it does not capture a certain set of events. But from the viewpoint of a theory of causation that does not allow for any overdetermination, the exclusion problem covers all cases of nonphysical causation, a significant difference.

As the basis for the discussion of the possible impact of theories of causation on the exclusion problem, we first carefully developed our own version of it, since the versions discussed in the literature were often either too strong or too weak. Once our version was on the table and once it was clear that and how theories of causation affect the exclusion problem, the question whether any of this would impact standard responses to the problem naturally arose. Again, the answer was affirmative. The version of the problem that we developed in Chapter 2 indeed makes it more difficult for some of the responses to succeed. Yablo, for example, endorsed a much weaker version of the causal impact assumption (2). Partly because of that, his response could not convince as a response to our version. Other responses turned out to work particularly well in combination with particular theories of causation but not so well in combination with others. Horgan’s treating causation as a four-place relation, for example, seems to be particularly compatible with counterfactual theories or causal pluralism, but it might not work well with probabilistic theories or transference theories.
In spite of these positive results, there is much we did not accomplish. On the one hand, we tried to discuss a representative set of theories of causation, partly by analyzing the impact of some of the most popular theories of causation as well as the impact of some more obscure theories. The fact that all of them had some kind of impact on the exclusion problem may justify the conjecture that perhaps all theories of causation do. On the other hand, we had to neglect some of the most promising theories of causation available. For example, we pointed out that some of the problems that a simple probabilistic theory of the sort we discussed in section 3.3.4 faces might be dealt with by combining it with another theory or by adding further assumptions to build a more complex theory that can handle problematic cases of overdetermination better. Of course, that is exactly what proponents of probabilistic theories have often done. One reason why we did not pursue this further is that in many cases these theories diverge very far from terminological conventions of the sort we have established. Wesley Salmon’s theory, for example, treats processes as more fundamental than events. But that complicates an application of it to the exclusion problem as we have construed it. One project could thus be to look at such promising theories of causation and analyze their impact on the exclusion problem as well.

Investigating theories of causation not yet covered may have a further benefit. As just pointed out, some of these theories consider causal relata other than events. This may give us the opportunity to test a suspicion we voiced in section 2.1.1, namely, that different causal relata may have an impact on the problem that events do not have. It may seem as if translations of statements involving one type of causal relatum into statements involving another kind are easy to come by. But, as we suggested, such translations may be deceptive. Perhaps other types of causal relata would generate new kinds of problems. In this context, it also is worth reminding ourselves that one shortcoming of our investigation is that, in order to cover a good number of
theories of causation and responses to the problem, we were bound to be superficial in our analysis. Looking at particular theories in more detail might thus reveal further unexpected ways in which different types of causal relata as well as particular theories of causation affect the problem.

In section 3.3.8, we sketched the beginning of a nomological version of the exclusion problem. Since many reductive theories of causation rely on the concept of a law of nature, developing such a nomological exclusion problem in detail should prove fruitful not just for a deeper analysis of the impact of reductive theories of causation on the problem but also for exploring the problem’s implications for the special sciences. In addition, a nomological version might connect better to the problem of free will. Simply put, the problem is whether it is possible for free will to exist in a deterministic world. Determinism is usually introduced as the thesis that, given any state of the universe at a given time and the laws of nature, any subsequent state follows necessarily. To articulate this problem, the notion of causation is not needed. A nomological version of the exclusion problem will thus make it easier to discuss interesting and important connections between the problem of free will and the exclusion problem.

While completeness, laws of nature, overdetermination, and properties played an important role in our analysis, we did not systematically explore how different theories of them may affect the problem. Ultimately, a thorough analysis of the exclusion problem should do so. Thus, this dissertation is best understood as the first step towards a comprehensive primer on the exclusion problem. As of now, no such primer exists, but it should. Regardless of which of these projects we may pursue first, having developed a version of the problem that avoids some of the pitfalls and unnecessary ontological commitments of other versions will serve as a good starting point for future research.
APPENDIX
KANT EXCERPT

The following passage in Immanuel Kant’s *Critique of pure reason* could be interpreted as fusing questions concerning the compatibility of free will and determinism with problems regarding causal exclusion.

[T]hus the difficulty we encounter in the question about nature and freedom is only whether freedom is possible anywhere at all, and if it is, whether it can exist together with the universality of the natural law of causality, hence whether it is a correct disjunctive proposition that every effect in the world must arise either from nature of freedom, or whether instead both, each in a different relation, might be able to take place simultaneously in one and the same occurrence. The correctness of the principle of the thoroughgoing connection of all occurrences in the world of sense according to invariable natural laws is already confirmed as a principle of the transcendental analytic and will [not] suffer violation. Thus the only question is whether, despite this, in regard to the very same effect that is determined by nature, freedom might not also take place, or is this entirely excluded through that inviolable rule? And here the common but deceptive presupposition of the absolute reality of appearance immediately shows its disadvantageous influence for confusing reason. For if appearances are things in themselves, then freedom cannot be saved. Then nature is the completely determining cause, sufficient in itself, of every occurrence, and the condition for an occurrence is always contained only in the series of appearances that, along with their effect, are necessary under the law of nature.  

\[2\]

---

2 Kant (1781/1997, A536/B564). Unfortunately, Guyer and Wood forgot to include the “not” in their translation which we added in brackets. The German version is clear on this.
LIST OF REFERENCES


BIOGRAPHICAL SKETCH

Andreas Falke was first exposed to the exclusion problem as an undergraduate at the Goethe University in Frankfurt, Germany, where he majored in philosophy and minored in historical ethnology and religious studies. He became so intrigued by the problem that he ended up writing his thesis on it. In 2006, he graduated with a Magister degree. Andreas then attended the philosophy graduate program at the University of Florida where he received a master’s degree in 2010, just to get back to studying the exclusion problem in more detail. In his dissertation *Excluding thoroughly: Theories of causation and their impact on the exclusion problem*, Andreas presented the results of his inquiry. He received a doctorate in philosophy in Fall 2017.