

2010

# The Problem of Causal/Explanatory Exclusion

Dwayne Moore  
*Wilfrid Laurier University*

Follow this and additional works at: <http://scholars.wlu.ca/etd>



Part of the [Psychology Commons](#)

---

## Recommended Citation

Moore, Dwayne, "The Problem of Causal/Explanatory Exclusion" (2010). *Theses and Dissertations (Comprehensive)*. 1102.  
<http://scholars.wlu.ca/etd/1102>

---

This Dissertation is brought to you for free and open access by Scholars Commons @ Laurier. It has been accepted for inclusion in Theses and Dissertations (Comprehensive) by an authorized administrator of Scholars Commons @ Laurier. For more information, please contact [scholarscommons@wlu.ca](mailto:scholarscommons@wlu.ca).



Library and Archives  
Canada

Published Heritage  
Branch

395 Wellington Street  
Ottawa ON K1A 0N4  
Canada

Bibliothèque et  
Archives Canada

Direction du  
Patrimoine de l'édition

395, rue Wellington  
Ottawa ON K1A 0N4  
Canada

Your file Votre référence  
ISBN: 978-0-494-68751-2  
Our file Notre référence  
ISBN: 978-0-494-68751-2

**NOTICE:**

The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

---

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

**AVIS:**

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

---

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.



# Canada



# **The Problem of Causal/Explanatory Exclusion**

By  
**Dwayne Moore**

**B.A., Wilfrid Laurier University, 2002**  
**M.A., Wilfrid Laurier University, 2004**  
**M.A., University of Waterloo, 2006**

**A Dissertation**  
**Submitted to the Department of Philosophy**  
**in partial fulfillment of the requirements for**  
**Doctor of Philosophy**  
**in Philosophy**  
**Wilfrid Laurier University**  
**2010**

## **Abstract:**

The problem of mental causation, at least in one of its most basic forms, is how to reconcile two plausible but potentially incompatible intuitions. The first intuition is that the mind makes a difference in the world. For example, I am writing this paragraph for certain reasons, and before long I will stop to eat something because of certain desires for food. Seemingly, these reasons and desires play a role in what happens. The second intuition is that the physical world is causally complete, so everything that happens is the result of the movement of physical particles. For example, the neural turbulence in my head seems to be the actual cause of my hands fluttering across the keyboard in certain ways, whilst certain muscle contractions in my arms cause the food to enter my mouth. What room is there for the mind to play a causal role when everything seems to happen because of the movement of physical particles?

For some time reductive physicalism was the prevailing solution to the problem of mental causation (Place, 1956; Feigl, 1958; Smart, 1959). Reductive physicalism posits a reductive identity of the mental to the physical. In so doing, it endorses physical causal completeness, but achieves mental causation as well, since the mental is identical with the causally efficacious physical. In the nineteen seventies, nonreductive physicalism replaced reductive physicalism as the predominant solution to the problem of mental causation in the nineteen seventies. Nonreductive physicalism solves the problem of mental causation by agreeing that the physical is causally complete, but achieves mental causation as well by supposing that the mental supervenes upon the physical, and thus inherits the causal power of the physical.

In recent years this nonreductive consensus has been threatened. This is partially due to an argument that has been distilled from Jaegwon Kim's principle of causal/explanatory exclusion. One part of this compound principle is the principle of causal exclusion, which states that there can be no more than a single sufficient cause for any given event (Kim, 2005, p. 42). This principle of causal exclusion creates the

following problem: the nonreductive physicalist endorses the causal completeness of the physical, and so she agrees that there is a sufficient physical cause for any given event. The nonreductive physicalist also avoids making a reductive identity between the mental and the physical, so she agrees that the mental is distinct from the physical. Therefore, if a given event has a complete physical cause, and the mental cause is distinct from this complete physical cause, then this supervening mental cause must be excluded. The physical cause does all of the work, so there is no work left over for the mental cause.

In this dissertation I consider and respond to Jaegwon Kim's principle of causal/explanatory exclusion. I conclude that the most promising response to the problem generated from causal exclusion is to endorse what I call structuralism. Structuralism construes mental states as mereological structures, or configurations, of parts. Macro structure plays a role in determining which micro properties its parts will and will not instantiate, so there is a genuine role for the mental to play. The micro properties that are instantiated, however, do all of the causal work, so causal completeness is secured as well. This is a nonreductive position, since the mereological structure of the parts is not identical with the parts themselves. This model avoids the causal exclusion problem by noting that mereological relations are non-causal determinative relations, so mental states can play an important determinative role without contributing any causal power beyond what the causally sufficient micro properties of the parts contribute.

This solution to the problem of causal exclusion affords a solution to the parallel problem generated from the principle of explanatory exclusion as well. The principle of explanatory exclusion states that "there can be no more than a single *complete* and *independent* explanation for any one event" (Kim, 1988, p. 233). I resolve this difficulty by adopting a nuanced form of what is called the dual-explananda reply. Since the above reasoning suggests that mental states are distinct from physical events, we can conclude that mental explanations and physiological explanations do not refer to the same thing, so there is no exclusion pressure between the two explanations.

## **Acknowledgments:**

I want to first of all extend my heartfelt thanks to my wife, Louise, for her support and encouragement throughout this process, without which, I would no doubt still be stuck in chapter three somewhere. Even after many years of study, asking for her hand in marriage remains the best idea I've ever come across.

I would also like to thank my supervisor, Neil Campbell, who has most certainly gone above and beyond the call of duty in helping me to produce this dissertation. His thorough reading of earlier versions of this dissertation have surely improved the end product, and served as a model of philosophical precision for me. I would also like to thank Rockney Jacobsen for bringing his breadth of knowledge to this dissertation, and Andrew Bailey for providing incisive comments, which carried my thinking further. I would also like to thank William Seager and Suzanne Zeller for their valuable feedback.

Portions of this dissertation appear in several different journals, which I would like to acknowledge. First of all, portions of chapter ten appear in an article entitled ‘On Kim’s Exclusion Principle’, co-authored with Neil Campbell, and published in *Synthese*, 169: 75-90. Portions of chapter eight appear in an article entitled ‘The Generalization Problem and the Identity Solution’, which is published in *Erkenntnis*, 72: 57-72. Portions of chapter six appear in ‘Explanatory Exclusion and Extensional Individuation’, which appears in *Acta Analytica*, 24: 211-222. I also thank my commentator, Paul Raymont, and the audience, at a presentation delivered at the Canadian Philosophical Association in Montreal, as they provided excellent feedback on a portion of this dissertation.

Finally, I would like to thank several funding agencies for their generous financial support. These include the Social Sciences and Humanities Research Council of Canada, the Ontario Graduate Scholarships program, and Wilfrid Laurier University.

## Table of Contents:

1. The Problem of Mental Causation.....	1
1.1 – Solutions to the Problem of Mental Causation.....	2
1.2 – Précis of this Dissertation.....	4
2. The Problem of Causal/Explanatory Exclusion.....	10
2.1 – Mental Causation.....	10
2.2 – Causal Completeness.....	18
2.3 – Irreducibility.....	22
2.4 – No Overdetermination/Causal Exclusion.....	25
2.5 – The Problem of Causal Exclusion.....	28
2.6 – The Problem of Explanatory Exclusion.....	29
3. Solutions to the Problem of Causal Exclusion.....	32
3.1 – The Supervenience Solution to the Problem of Causal Exclusion.....	32
3.2 – Problems with the Supervenience Solution.....	35
3.3 – The Emergentist Solution to the Problem of Causal Exclusion.....	38
3.4 – Problems with the Emergentist Solution.....	41
3.5 – The Constitution Solution to the Problem of Causal Exclusion.....	44
4. Causal Exclusion and Overdetermination.....	54
4.1 – Problems with Independent Overdetermination.....	54
4.2 – Dependent Overdetermination.....	56
4.3 – Problems with Dependent Overdetermination.....	61
5. Causal Exclusion and Functional Reduction.....	75
5.1 – Kim’s Functional Reduction.....	75
5.2 – Functional Reduction and Elimination.....	78
5.2.a – Disjunctive Functional Reduction and Elimination.....	79
5.2.b – Conceptualized Functional Reduction and Elimination.....	81
5.2.c – Token Functional Reductionism and Elimination.....	84
5.3 – Epiphenomenalism and Properties of Properties.....	87
5.4 – Explanatory Exclusion and Unnecessary Explanation.....	90

6. Causal Exclusion and Kimian Events.....	98
6.1 – The Property Exemplification Model.....	98
6.2 – Causal Exclusion and Davidsonian Events.....	102
6.3 – Explanatory Exclusion, Events and Descriptions.....	108
6.3.a – Problems with Extensional Individuation.....	108
6.3.b – The Constitutive Property Reply.....	110
6.3.c – Problems with the Constitutive Reply.....	111
7. The Problem of Causal Exclusion and Causation.....	121
7.1 – The Generative Model of Causation.....	122
7.2 – Counterfactual Causation and Causal Exclusion.....	126
7.3 – Causal Sufficiency and Determinative Background Conditions.....	131
7.4 – Causal Heterogeneity.....	135
8. Kim on Mereology.....	140
8.1 – The Generalization Problem and the Higher Level Solution.....	141
8.2 – Problems with the Higher Level Solution.....	144
8.3 – The Generalization Problem and the Identity Solution.....	146
8.4 – Problems with the Identity Solution.....	150
9. Specific Mereological Structure and Non-Causal Mereological Determination.....	163
9.1 – Mereological Determination.....	163
9.2 – Non-Causal Mereological Determination.....	169
9.3 – Mereological Determination and the Co-Dependence Objection.....	173
10. Resolving Causal/Explanatory Exclusion.....	177
10.1 – Structuralism and Mental Causation.....	178
10.1.a – The Structuralist Model of Mental Causation.....	178
10.1.b – Structuralism and Functionalism.....	184
10.1.c – Structuralism and Mental Determination.....	186
10.2 – Structuralism and Causal Exclusion.....	190
10.3 – Structuralism and Completeness.....	199
10.4 – Structuralism and Irreducibility.....	202
10.5 – Structuralism and the Problem of Explanatory Exclusion.....	204
10.6 – Conclusions.....	213
Bibliography.....	216

## **List of Tables:**

Figure 1: Spectrum of Solutions to the Mental Causation Problem.....	2
Figure 2: Supervenient Causation.....	34
Figure 3: The Problem of Causal Exclusion.....	36
Figure 4: Mouse Trap.....	179
Figure 5: Causal Exclusion Solution.....	190
Figure 6: Versions of the Dual Explananda Reply.....	205

## **1 – The Problem of Mental Causation**

The problem of mental causation, at least in one of its most basic forms, is how to reconcile two plausible but potentially incompatible intuitions. The first intuition is that the mind makes a difference in the world. For example, I am writing this paragraph for certain reasons, and before long I will stop to eat something because of certain desires for food. Seemingly, these reasons and desires play a role in what happens. The second intuition is that the physical world is causally complete, so everything that happens is the result of the movement of physical particles. For example, the neural turbulence in my head seems to be the actual cause of my hands fluttering across the keyboard in certain ways, whilst certain muscle contractions in my arms cause the food to enter my mouth. What room is there for the mind to play a causal role when everything seems to happen because of the movement of physical particles?

Perhaps we could avoid the problem by simply rejecting one of these conflicting intuitions. There are those who make such moves, and we will briefly encounter some of them. However, there are some significant arguments grounding these intuitions. Jaegwon Kim, for example, contends that mental causation is required as a foundation for ethics and epistemology, while it simultaneously makes sense of our everyday experiences (Kim, 1993, p. xv; Kim, 1998, p. 32; Kim, 2005, p. 10). He also argues that the causal completeness of the physical world can be inferred from the fact that science has continuously found physical causes for observable effects, and indeed, certain scientific conservation laws rule out non-physical causes (Kim, 2005, p. 154). We will unpack each of these arguments in greater detail in chapter two. For now, let us simply note that these considerations are persuasive enough to convince most people that the best resolution to the problem of mental causation cannot involve the rejection of one of these two intuitions.

### 1.1 – Solutions to the Problem of Mental Causation

In the diagram below, I have charted a very rough spectrum of some of the various resolutions to the problem of mental causation.<sup>1</sup> On the left extreme stands eliminative physicalism which endorses physical causal completeness but does not support mental causation (Churchland, 1981; Stich, 1983). Furthest to the right stands Cartesian dualism, which denies the causal completeness of the physical realm by advocating the strong autonomy of mental causation (Descartes, 1988; Foster, 1989). Eliminativism avoids the problem by abandoning mental causation, whilst Cartesianism avoids the problem by rejecting physical causal completeness. As mentioned, such positions have been largely resisted due to the force of the aforementioned arguments in favour of mental causation and physical causal completeness.

Slightly to the left of centre stands reductive physicalism. For some time this was the prevailing solution to the problem of mental causation (Place, 1956; Feigl, 1958; Smart, 1959). Reductive physicalism posits a reductive identity of the mental to the physical. In so doing, it endorses physical causal completeness, but achieves mental causation as well, since the mental is identical with the causally efficacious physical.

Slightly to the right of centre stands nonreductive physicalism. It replaced reductive physicalism as the predominant solution to the problem of mental causation in the nineteen seventies. This happened because Hilary Putnam and Jerry Fodor persuaded many that mental properties can be realized by various different physical bases<sup>2</sup>, so they

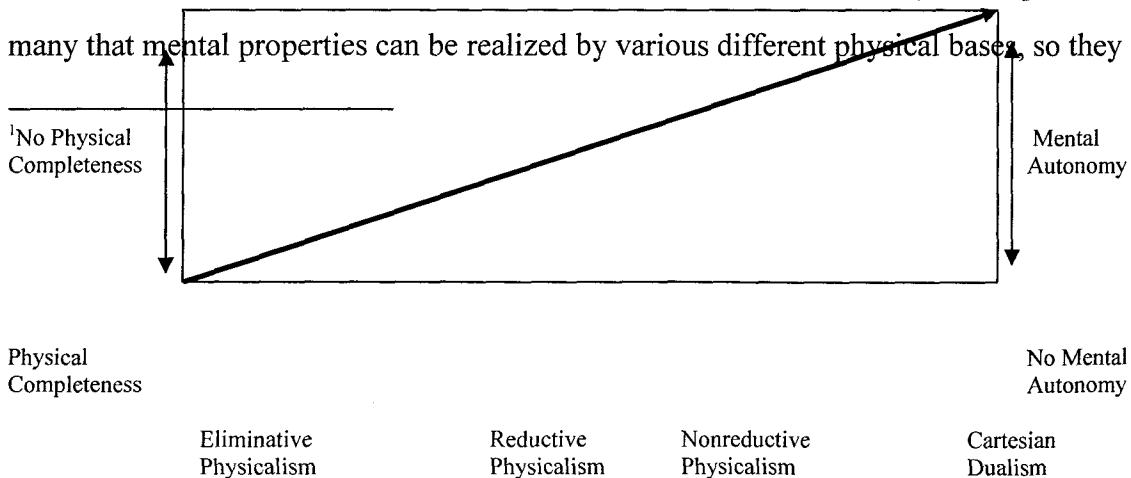


Figure 1: Spectrum of Solutions to the Mental Causation Problem

cannot be identical with these realizing bases (Putnam, 1967; Fodor 1974). Still others were convinced by Donald Davidson that mental vocabulary is irreducible to physical vocabulary because no law can map one conceptual grid onto the other (Davidson 1970; Davidson 1993). Nonreductive physicalism solves the problem of mental causation by agreeing that the physical is causally complete, but achieves mental causation as well by supposing that the mental supervenes upon the physical, and thus inherits the causal power of the physical. Nonreductive physicalism also achieves a certain degree of autonomy and irreducibility to mental causation, since it endorses the supervenience relation rather than the identity relation. I will consider this solution in Section 3.1.

In recent years this nonreductive consensus has been threatened. This is partially due to an argument that has been distilled from Jaegwon Kim's principle of causal/explanatory exclusion. One part of this compound principle is the principle of causal exclusion, which states that there can be no more than a single sufficient cause for any given event (Kim, 2005, p. 42). This principle of causal exclusion creates the following problem: the nonreductive physicalist endorses the causal completeness of the physical, and so she agrees that there is a sufficient physical cause for any given event. The nonreductive physicalist also avoids making a reductive identity between the mental and the physical, so she agrees that the mental is distinct from the physical. Therefore, if a given event has a complete physical cause, and the mental cause is distinct from this complete physical cause, then this supervening mental cause must be excluded. The physical cause does all of the work, so there is no work left over for the mental cause to do. If mental causes are excluded from causal efficacy in this way, then nonreductive physicalism does not deliver mental causation, in which case we need to abandon nonreductive physicalism if we want to retain mental causation.

## 1.2 – Précis of this Dissertation

In this dissertation I consider and respond to Jaegwon Kim's principle of causal/explanatory exclusion. This involves, first of all, an analysis of the aforementioned principle of causal exclusion and the difficulties it poses. I conclude that the most promising response to the problem generated from causal exclusion is to endorse what I call structuralism. Structuralism resembles, to varying extents, the positions that a number of recent philosophers have taken (Kistler, 2009; Gillett, 2006; Murphy 2006; El Hani and Emmeche, 2000; Van Gulick, 1993), and it construes mental states as mereological structures, or configurations, of parts. Macro structure plays a role in determining which micro properties its parts will and will not instantiate, so there is a genuine role for the mental to play. The micro properties that are instantiated, however, do all of the causal work, so causal completeness is secured as well. This is a nonreductive position, since the mereological structure of the parts is not identical with the parts themselves. This model avoids the causal exclusion problem by noting that mereological relations are non-causal determinative relations, so mental states can play an important determinative role without contributing any causal power beyond what the causally sufficient micro properties of the parts contribute.

This solution to the problem of causal exclusion affords a solution to the parallel problem generated from the principle of explanatory exclusion as well. The principle of explanatory exclusion states that "there can be no more than a single *complete* and *independent* explanation for any one event" (Kim, 1988, p. 233). This principle can generate problems for those nonreductive physicalists who think that every event has a complete physical explanation and an independent mental explanation. For if every event has a complete physical explanation, and there can be no more than this complete explanation, then an independent mental explanation will be excluded. I resolve this difficulty by adopting a nuanced form of what is called the dual-explananda reply. Since the above reasoning suggests that mental states are distinct from physical events, we can conclude that mental explanations and physiological explanations do not refer to the same thing, so there is no exclusion pressure between the two explanations.

This dissertation defends the viability of these two related solutions by posing problems for alternate solutions while simultaneously building a case for my own resolutions. To this end, chapter two introduces in detail the arguments for mental causation and physical causal completeness that we have briefly seen above. These two premises are not logically contradictory without the addition of at least two further claims; namely, irreducibility and causal exclusion. These four claims, when combined together, appear to form an inconsistent tetrad. If mental causes are not reductively identified with physical causes, and we can have no more than one sufficient cause for any event, then it is more likely that we have to say that either the physical cause is not sufficient for the event, or there is no mental cause of the event. In this chapter I also spell out the principle of explanatory exclusion and the aforementioned problem that it creates for certain nonreductive physicalists.

In chapter three, I consider various responses to the potentially inconsistent tetrad mentioned above. I consider the traditional supervenience solution, which argues that mental events supervene upon physical events. I conclude that Kim's causal exclusion principle rules out these supervening causes because the subvening bases are sufficient to cause the given effect. I then turn my attention to emergentism, which contends that mental properties emerge out of configurations of certain parts. I conclude that Kim's causal exclusion principle rules out these supervening emergent properties as well, on account of the fact that the configuration of these parts is a sufficient cause for the given effect. I then look to the constitution view, which supposes that mental events are constituted out of physical events, and as such they justifiably inherit the causal powers of their physical constituents. I conclude, once again, that the constituent physical events are sufficient for the effect, and as such the mental events can be excluded on account of the fact that they overdetermine the effect.

Having rejected these popular attempts at reconciling our four competing principles, I spend two chapters considering the viability of rejecting one of these principles themselves. Chapter four considers the rejection of the causal exclusion

principle. I explore an initial attempt at rejecting causal exclusion, called independent overdetermination. It fails because it postulates massive amounts of coincidence, and it renders both the mental and the physical causes individually unnecessary. I then consider a second attempt, called dependent overdetermination. This position avoids the difficulties of the first model, but it amounts to the view that mental and physical causes are individually insufficient. Hence we cannot solve our problem by rejecting causal exclusion. In this chapter I also demonstrate that mentality must play a necessary role, while also noting that mental states can play a determinative role, without being a cause.

In Chapter five, I discuss the possibility of rejecting irreducibility. Kim solves the problem of causal exclusion by reductively identifying mental causes with physical causes via functional reduction. I also outline Kim's solution to the problem of explanatory exclusion, which secures the viability of mental explanations by arguing that the identity guarantees that mental descriptions state the same explanation as physical descriptions, since they refer to the same objective relation between the identified events. Kim's solution to the problem of causal exclusion fails because it raises the specter of mental eliminativism in a number of different ways, which in turn amounts to a loss of mental causation. It does not overcome all of the arguments raised in favour of irreducibility either. Finally, it only offers us reduced mental causation, which Kim acknowledges is not as appealing as the autonomous mental causation that is available when we endorse irreducibility. Kim's solution to the problem of explanatory exclusion also fails. The identity thesis, when combined with the principle of the explanatory completeness of the physical, leads to the conclusion that there is a complete physical description for every event, so mental descriptions are not necessary – a problematic result, to say the least.

After concluding that a resolution will not be found by rejecting one of the four principles of our seemingly inconsistent tetrad, I spend two chapters considering metaphysical solutions to the problem of causal exclusion. Chapter six deals with Kim's model of events. For Kim, an event is the instantiation of a property at a time. A number

of philosophers dodge the exclusion problem by rejecting this model of events. This move, however, amounts to the loss of a certain amount of realism and explanatory power, and it will not convince a Kimian. I also demonstrate that Kim's solution to the problem of explanatory exclusion relies on an extensional model of explanatory individuation, which is highly implausible.

In chapter seven, I consider Kim's model of causation. Kim endorses a generative model of causation that construes causation as an objective causal relation between events. A number of philosophers avoid the causal exclusion problem by adopting a counterfactual account of causation. I argue that the counterfactual account, among other things, counts epiphenomenal events as causes, and as such should not be endorsed as a model of causation. The counterfactual account will not convince a Kimian either. I then demonstrate that Kim thinks a certain event can be a sufficient cause despite the fact that certain background conditions necessarily play a determinative role. This dovetails nicely with the conclusion, reached in chapter four, that an event can be a sufficient cause despite the fact that mental states necessarily play a determinative role. But, in order for a resolution to the problem of causal exclusion to involve necessarily determinative mental states and sufficient physical causes, we will need to demonstrate a distinction between the role that mental states and physical events play in generating the effect. To this end I close this chapter by demonstrating that the distinctness of the mental and the physical implicit in the principle of irreducibility allows us to conceive of mental states in contrast to physical events, and also allows us to conceive of these mental states as playing a determinative role as opposed to the causal role played by physical events. If this heterogeneity is plausible, it is possible for mental states to necessarily play a determinative role while simultaneously securing the sufficiency of the physical cause. What we need to do, therefore, is find a non-causal determinative relation that mental states can play, and we may have a resolution to the problem of causal exclusion.

I spend the next two chapters arguing that mereological relations are such non-causal determinative relations. I begin this argument in chapter eight, with an analysis of the generalization problem. The generalization problem suggests that the causal exclusion problem applies to all higher level (i.e., macro) special science properties, since the lower level (i.e., micro) properties of the parts are causally sufficient. Kim originally argues that there is no generalization problem since causal exclusion pressure applies only between properties of the same object, not between properties of different objects on the mereological scale. According to Kim, the former applies to mental properties, while the latter applies to special science properties. This argument does not work, and in time Kim shifts to solving the generalization problem by reductively identifying macro properties with specific configurations of micro properties. The identity of a macro property with many micro properties and relations does not work either. Moreover, the identity of a macro property with a configuration of micro properties leaves us wondering whether the micro properties and their individual relations are causally sufficient for a given effect or not. If so, we can exclude the configuration of the micro properties. If not, we have an apparent violation of causal completeness.

In chapter nine this dilemma is resolved by arguing that macro properties, which are instantiated as the structure of the parts, play a non-causal determinative role. Mereological relations are symmetrically determinative relations. Micro to macro determination is evident in standard cases of supervenience. Macro to micro determination is evident whenever the structure of a whole triggers or inhibits the conditional micro properties of the parts. Mereological relations are also non-causal on account of the spatial and temporal co-occurrence between part and whole which is not present with causal relations. We therefore have the groundwork for a solution to the problem of causal exclusion that construes mental states as non-causal determinative mereological structures of parts, which leaves the causal efficacy for the properties of its parts. I close this chapter by considering various objections.

In chapter ten I present this model, called Structuralism, as a resolution to the causal exclusion problem. Structuralism suggests that the distinctness implicit in irreducibility implies that mental states may play a different sort of role than physical events. Causal completeness suggests that physical events are sufficient causes, while causal exclusion states that these physical events will exclude any other causes. Thus, if mental states play a role in determining that the given effect will occur, mental states will have to play a non-causal determinative role. The doctrine of mental causation, however, suggests that mental states will play some sort of a role, so mental states must make a non-causal determinative contribution. Mereological relations, however, are one of the few sorts of determinative relations that are not causal, so structuralism looks like an appropriate resolution. I deal with certain further objections, and close by acknowledging that this solution limits the share of mental causation, physical causal completeness, irreducibility and causal exclusion that we can have in important ways. To some degree, however, it allows us to keep all four of these intuitions, and this is perhaps near enough to the goal.

I then argue that structuralism provides the groundwork for a nuanced version of the dual-explananda reply to the problem of explanatory exclusion. Structuralism does not reduce mental states to micro physical states, so mental explanations do not refer to the same thing that micro physical explanations do, so there is no exclusion pressure between the two. I then deal with an objection to this solution. I conclude that the resolutions offered here are the most promising solutions to the parallel problems of causal/explanatory exclusion.

## Chapter 2 - The Problem of Causal/Explanatory Exclusion

In this chapter I outline and motivate four plausible principles that, when taken together, seem to form an inconsistent tetrad: (1) mental causation; (2) physical causal completeness; (3) irreducibility; and (4), the rejection of overdetermination. Taken in isolation each principle is plausible, yet when joined together certain problems arise. As it turns out, we can adhere to any three of the principles, but we may not be able to adhere to all four. One way of characterizing the problem of causal exclusion is in terms of this inconsistent tetrad.

This chapter is split into five sections. In Section 2.1, I outline the arguments in favour of mental causation, while in Section 2.2 I outline the arguments in favour of physical causal completeness. Section 2.3 deals with the arguments in favour of the irreducibility of the mental to the physical, while Section 2.4 spells out certain reasons to reject overdetermination. This rejection of overdetermination is largely rooted in the principle of causal exclusion, so I introduce and analyze this principle as well. In Section 2.5, I demonstrate the tensions that are generated when these four intuitions are conjoined. Finally, in Section 2.6, I introduce the parallel principle of explanatory exclusion.

## **2.1 –The Argument for Mental Causation**

There are four plausible metaphysical assumptions that a number of nonreductive physicalists endorse. There are, of course, several varieties of nonreductive physicalism, not all of which endorse all of these principles as formulated below. However, many nonreductive physicalists do accept these intuitions (Harbecke, 2008, p. 16ff; Bennett, 2008, p. 281; Kallestrup, 2006, p. 459-465; Sturgeon, 1998, p. 413-415). These principles are:

- 1) Mental causation – “Mental properties have causal efficacy - that is, their instantiations can, and do, cause other properties, both mental and physical, to be instantiated” (Kim, 2005, p. 35).

- 2) Causal completeness – “If a physical event has a cause that occurs at  $t$ , it has a physical cause that occurs at  $t'$ ” (Kim, 2005, p. 43)
- 3) Irreducibility – “Mental properties are not reducible to, and are not identical with, physical properties” (Kim, 2005, p. 34)
- 4) No overdetermination/Exclusion – “No single event can have more than one sufficient cause occurring at any given time” (Kim, 2005, p. 42)

Let us spend some time motivating and clarifying each of these principles, beginning with the first.

As noted in Chapter One, Jaegwon Kim offers three reasons to endorse the existence of mental causation, one of which is the argument from common sense. Almost all human beings, hundreds of times each day, perceive that their beliefs, desires and choices influence their daily lives. We want a peach, and so we eat one. We believe it will rain, and so we bring an umbrella. We feel the stove top heating up, so we remove our hand. We choose to go to university, and so we end up there. These common experiences, when combined together, form an impressive body of evidence for the conclusion that our thoughts and desires play a causal role in our behaviour.

This is not to say that these common sense experiences cannot be wrong. Many once thought the earth flat, after all. Rather, it says that we should be hesitant to dismiss these reports as mere folk psychology. Otherwise, we may be tempted to conclude along with Jens Harbecke that “any coherent theory of the world that does not recognize [mental causation] would seem to us as not being very coherent after all” (Harbecke, 2008, p. 18). Not surprisingly, a number of philosophers have deployed this argument from common sense to support the conclusion that mental causation is true (Fodor, 1990, p. 156; Searle, 1984, p. 87-88; Bealer, 2007, p. 47; Elder, 2001, p. 111; Walter, 2009, p. 2; Kim, 1992, p. 119-120).

Ethical grounding is a second argument for mental causation. Immanuel Kant once advanced a fairly intuitive meta-ethical principle: “the action to which the ‘ought’ applies must indeed be possible under natural conditions” (Kant, 1781, A548. See also

Stern, 2004, p. 53-55). For Kant, and for many others, an ‘ought’ implies a ‘can’. For example, if Susan is chained up and locked in a prison in India, and therefore cannot give food to a hungry person in Haiti, she cannot be blamed for not giving food to the hungry person in Haiti. In order for an agent to be morally responsible for not acting, the agent has to be able to act in the first place. Yet if an agent’s mental states have no causal powers, then the agent is literally incapable of acting in any way, be it for good or harm. Thus, after pointing out a number of philosophers who insist that mental causation is necessary to render psychological explanations intelligible (Davidson 1963; Mele 1992), David Robb concludes that, “if psychological explanation goes, so do the closely related notions of agency and moral responsibility” (Robb, 2008; See also Horgan 2007).

Some think that another important component in moral responsibility is that an ‘ought’ implies a ‘cannot’ as well (Aristotle, 1113b6; Kane, 1996, p. 32; Van Inwagen, 1983, p. 68). According to them, it is not enough to say that an ought implies a can, for the rocks that tumble down the hill and roll over the hiker still *can* roll down the hill. This does not make the rocks morally responsible. Rather we must add, at least, that a person could do something different than what they did. Edgar robs the bank, and he could have refrained from doing so, so for these reasons Edgar is morally responsible for robbing the bank. Thus, when Clarence Darrow argued that his clients were pre-determined to murder their victim, so they had no choice to do otherwise, he had some success (Darrow, 1924, p. 65). Similar reasoning is used when making pleas of diminished responsibility or temporary insanity. In these cases the defendant may not be fully capable of acting on alternative possibilities, and so their culpability is called into question. These arguments demonstrate that the ability to not act plays an important role in assigning moral responsibility as well. This additional component is important because, as we will see below, certain varieties of mental causation may not offer this feature in their foundation for ethics.

The epistemic grounding argument is the third reason to endorse mental causation. There are a number of reasons to think mental causation is important to

epistemology, of which I need only mention one. Plato was perhaps the first to argue that true belief is insufficient for knowledge, since we could by chance have a true belief (Plato, *Theatetus* 200d-201c, Plato, *Republic* 506c). For example, I could spin around until dizziness causes me to fall to the ground, point in some random direction, and say “I believe Montreal is in that direction.” This belief could by chance be true, but this hardly seems to constitute knowledge.

Plato then argues that we could by chance have an account, or justification, for a true belief as well (Plato, *Theaetetus* 208b). For example, a child could parrot a memorized philosophical argument, but the many words will still not mean the child has knowledge of the topic. Edmund Gettier also presents a series of examples demonstrating that more than justified true belief is necessary for knowledge, since we could by luck have a justified true belief (Gettier, 1963, p. 121-123). Seemingly, as Duncan Pritchard contends, “proposing an epistemological theory that can eliminate reflective epistemic luck ... is central to knowledge possession” (Pritchard, 2005, p. 181. See also Riggs, 2007).

How can the truth of mental causation help to eliminate luck from epistemology? Consider a computer program that randomly selects three true premises from a database, then puts them in the form of a syllogism. For the most part, unintelligible arguments are generated, such as ‘Spiders spin webs and Brazil is in South America, therefore, the sky is blue’. After several million runs, however, the computer spits out ‘Socrates is human, and all humans are mortal, therefore, Socrates is mortal’. At this point we can ask ourselves whether or not the computer ‘knows’ that Socrates is mortal.<sup>2</sup>

Intuitively, it seems that it does not. Why? The computer has spat out a justification for the true belief, but the propositions still arose due to luck (i.e., a random, a-rational proposition generator), so the computer does not have knowledge. What do we need to add to render it more probable that the computer has knowledge? Seemingly, we

---

<sup>2</sup> Assuming, for the sake of the argument, that a computer is the type of thing that can know.

need to turn the logical relations between these three statements into a causal relation as well. Or, in other words, the proposition ‘Socrates is mortal’ needs to occur in virtue of the logical relation between the two preceding propositions, not in virtue of the random proposition generator.<sup>3</sup>

In contrast, consider a person who apprehends the first two propositions, which then *leads her* to conclude that Socrates is mortal. On this scenario, she thinks Socrates is mortal *because* of the two prior propositions, where this *because* will be the *cause* of her endorsing the conclusion. Or, in Davidsonian terminology, the computer may have reason for its conclusion, but it also needs to come to its conclusion for those reasons. If our epistemology requires that reasons, or beliefs, causally justify other beliefs in this way, then obviously holding beliefs must have causal power. A number of philosophers have come to similar conclusions, though for different reasons (Kim, 1988, p. 389ff; Dretske, 1989; Antony, 1989; Sturgeon, 2000). And, as noted, the primary characters involved in this debate acknowledge something like this argument, as well as the argument from common sense and ethical grounding, so there is no need to further justify the importance of mental causation here.

Before moving on, I want to add two caveats to the doctrine of mental causation. The diagram offered in Section 1.1 demonstrates that mental causation admits of degrees. On the far right of the chart in Figure 1, mental causation entails such things as libertarian free-will and the independence of our mental lives from physical causation. In the

---

<sup>3</sup> Rooted in David Hume’s argument that causes are independent from their effects while reasons are inseparably linked to their acts, the logical connection argument demonstrates that reasons cannot be causes (Ryle, 1949; Melden 1961). Donald Davidson (1963) responds to this argument by insisting that there is a difference between having a reason for acting and acting for a reason. The difference is that someone can have a reason to act, but this not be the cause of the act. On the contrary, someone who acts for a reason has this reason as the cause of their action. If Davidson is correct, as many, including Kim himself (1995, p. 138), think that he is, then reasons are causes and logical relations are causal relations. I follow Davidson’s intuition here, and I hope that my argument here helps to further ground the view that logical relations are causal relations.

middle-right, where the nonreductive physicalist lies, we still have autonomous mental causation. For our purposes, autonomous mental causation will mean that the mental plays a necessary, distinct role in the occurrence of an effect. I will consider reasons for endorsing the necessity of the mental role in Sections 2.4 and 4.1. To say that the mental plays a distinct role means that, to use a lumbering but illustrative term, the mental-as-strictly-mental plays a role. In other words, the decision to climb a tree causes bodily movement strictly in virtue of the decision.

In the middle left, where the reductive physicalist stands, there is reduced mental causation. Since mental properties are identical to physical properties, their causal efficacy is the causal efficacy of physical properties. Kim depicts the situation as follows: “mental events and states have the causal efficacy that they have because their neural/physical realizers have causal efficacy” (Kim, 2007, p. 239). This means we can expect to find mental-as-physical causation. In other words, the decision to climb a tree causes bodily movement in virtue of the fact that the decision is neural activity.<sup>4</sup>

For his part, Kim agrees that nonreductive physicalism and reductive physicalism offer these two different sorts of mental causation (Kim, 2005, p. 159). He also argues that autonomous mental causation is preferable to reduced mental causation:

The best, or the most satisfying, outcome would have been a vindication of mental causation along the lines of nonreductive physicalism; that would have allowed us to retain mentality as something that is causally efficacious and yet

---

<sup>4</sup> There is a possible objection here. Perhaps it is unfair to suggest the reductionist adheres to mental-as-physical causation, as contrasted with mental-as-strictly-mental causation. After all, the identity suggests there is one thing, and that one thing is mental, so it is mental-as-mental causation. Autonomous mental causation, however, is defined as being mental-as-*strictly*-mental causation. It was a choice as strictly the choice, distinct from and not reduced to certain physical properties, which caused the effect. Reduced mental causation, because the mental is physical, cannot provide mental-as-strictly-mental causation. As Slors and Walter contest “the dilemma for physicalism can thus be formulated as follows...how the autonomy of the mental can be secured if mental properties are just physical properties or how this apparent autonomy can be explained away as a mere illusion” (Slors and Walter, 2002, p. 3).

autonomous vis-à-vis the physical domain. But the best outcome, as we saw is not to be had. The next best outcome, in fact our only hope at this point if mental causation is to be saved, is physical reductionism. Physical reduction would save causal efficacy for mentality, at the cost of its autonomy (Kim, 2005, p. 159).

Most nonreductive physicalists not only agree with Kim that autonomous mental causation is preferable, but they also think it is important to keep autonomous mental causation (Jacob, 2002, p. 648; Pineda, 2001).

Autonomous mental causation is important to nonreductive physicalists for a number of reasons. First of all, at the very least, it prevents mental properties from being reduced to physical properties (Walter, 2009, p. 4 of 20; Antony, 2008, p. 164, Jackson, 1996, pg. 10-11). To some it is also important because it enables mental properties to figure in higher level laws of their own which are independent of lower level laws (Fodor, 1997, p. 149; Antony, 2007, p. 155; Kim, 1993, p. 351). Autonomous mental causation is also important because, as will be argued in Section 5.4, it grounds the necessity of mental explanations.

Autonomous mental causation may also be the only form of mental causation that follows from the three reasons given in favour of mental causation in the first place. For example, does the common sense argument lead us to conclude we have autonomous mental causation or reduced mental causation? Plausibly, when people consider their choices, they think that it is their choice as the choice, with at least some degree of autonomy, that causes things to occur. Ned Block, for example, argues that David Lewis' model of reduction only gives us mental causation in a "Pickwickian" or "cosmetic" sense (Block, 1990, p. 164-165. See also Loar, 1981; Greenwood, 1991), since it requires us to re-construe mental causation as physical causation in order to secure mental causation.

Or again, can ethical theory succeed if it is grounded by reduced mental causation? On this model mental events are physical events which invariably follow the laws of physics. Some think that moral responsibility fails if an agent's actions are

entirely determined to occur due to a combination of events in the remote past and the laws of physics, neither of which an agent can control (Ginet, 1990; Van Inwagen, 1975). The success of reduced mental causation as a ground for ethical theory, then, will rest on the arguments of those soft determinists who reject this argument (Lewis, 1981; Frankfurt, 1969). While there is not enough space here to consider these issues in detail, we can at least see how the arguments given in favour of mental causation may appear to be arguments in favour of autonomous mental causation rather than reduced mental causation, or mental causation in general.

Finally, for many the slide away from autonomous mental causation to reduced mental causation is considered benign (Pereboom and Kornblith, 1991, p. 143; Esfeld, 2005, p. 8-9; Shapiro and Sober, 2007, p. 247; Pineda, 2002, p. 31; Bennett, 2007, p. 327). Reduced mental causation is, after all, still mental causation. For others this shift raises some concern. The problem is that since the mental has causal power in virtue of the fact that it is physical, the mental no longer has causal power in virtue of the fact that it is strictly mental. Consider Michael Silberstein's concerns:

But even granting the identity of any given conscious state with a particular brain state, given the preceding argument, conscious states in virtue of being conscious will be epiphenomenal because by hypothesis, only the fundamental physical realizers are truly efficacious (Silberstein, 2001, p. 84. See also Silberstein, 2006, p. 220; Lowe, 1993, p. 631-632; McGinn, 1989, p. 137; Leiter and Miller, 1994, p. 221; Menzies and List, 2010; Walter, 2008, p. 674; Lowe, 1993; Hardcastle, 1996, p. 15; Wilson, 2009, p. 150; Shoemaker, 2001, p. 80).

Autonomous mental causation offers us mental causation as strictly mental causation, while reduced mental causation offers us mental causation as physical causation. On occasion, Kim demonstrates some concern over this shift in the variety of mental causation that is achieved:

Doesn't [reductionism] lead to the conclusion that the mental has no distinctive role of its own, having been entirely absorbed into the physical domain? That

again may seem to some as a form of mental irrealism, and one might think it makes no sense to save mental causation while relinquishing mentality as a distinctive reality (Kim 1998, p. 119. See also Kim, 1999, p. 33).

Although Kim does eventually relinquish the mental as a distinctive reality, others remain concerned that reduced mental causation does literally reduce the robustness of mental causation that is achieved. For these reasons, I want to accentuate Kim's conclusion that it is preferable to arrive at an autonomous form of mental causation where the mental plays a necessary, distinct role.

Kim's text above raises another important prospect as well. Autonomous mental causation may be preferable, but it may not be possible. In order to preserve physical causal completeness, a less autonomous form of mental causation may have to be accepted. This prospect applies to physical causal completeness as well: we may want a version of physical causal completeness which only endorses microphysical causation, leaving no room for special science causation (including psychology), but this may not be possible. In order to preserve mental causation, we may have to accept a nuanced version of physical causal completeness. Kim, for example, rejects eliminativism in favour of reductionism because he wants to preserve mental causation. Beyond this, he eventually settles on a position that is "near enough" to physicalism (Kim, 2005, p. 174), since he also wants to preserve qualitative mental properties.

## **2.2 – The Argument for Physical Causal Completeness**

As mentioned, the principle of physical causal completeness states that "if a physical event has a cause that occurs at  $t$ , it has a physical cause that occurs at  $t'$ " (Kim, 2005, p. 43).<sup>5</sup> Chapter one briefly documented two reasons why Kim considers physical causal

---

<sup>5</sup> The principle of causal completeness, as it stands above, is given in deterministic terms. If it turns out that the completed version of micro-physics is indeterministic in nature we could reframe this principle in terms of a probabilistic model of micro-causation. Karen Bennett suggests the following articulation: every

completeness true. Before considering these arguments I want to clarify several terms. First, the term “physical” can have two meanings. ‘Physical’ can mean ‘microphysical’, as in, physical entities are those microphysical entities referred to by the vocabulary of physics. Or, ‘physical’ can have a broader sense, where anything that appears in time and space can be considered physical. On this broad definition of the physical, macro objects composed of microphysical particles, such as neurons, trees and other special science entities, are considered physical, even though these entities do not explicitly appear in the vocabulary of physics. This distinction will become important in time. For now, notice that Kim (and the others listed in the argument for physical causal completeness below), use ‘physical’ in the latter, inclusive sense when arguing in favour of physical causal completeness.

I want to clarify the term ‘completeness’ as well. The doctrine of causal completeness, as with the doctrine of mental causation, admits of degrees. Causal completeness, in the form outlined above, suggests that every physical event has a sufficient, or complete, physical cause, so the physical world is causally complete. The physical world does not need any external forces to come in and contribute to the cause of certain effects, for the physical events can do all of the causal work themselves. However, as causal completeness stands above, non-physical causes may not be necessary, but this does not bar them from being able to supplement the complete physical cause (Montero, 2003, p 173ff; Kim, 2009, p. 38). One can admit that every physical event has a sufficient physical cause, while continuing to add a mental cause for the event as well.

This opening has been exploited. Eric Marcus, for example, accepts a causally complete physical world, since every event has a physical cause. However, he does not

---

physical event has “its probability fixed by entirely physical antecedents” (Bennett, 2008, p. 281). This prospect does not alter the intuition in any significant way. So, until such a prospect is fully realized, it is simpler and equally viable to speak in terms of events.

think this entails that a mental cause cannot also affect the event (Marcus, 2005, p. 19ff; Crane and Mellor, 1990, p. 206). This strategy, which solves the problem of mental causation by endorsing overdetermination, remains open on the aforementioned model of physical causal completeness. Kim is aware of this vulnerability, and in response he considers strengthening the principle to read “no non-physical event can be a cause of a physical event” (Kim, 2005, p. 50). This principle not only states that non-physical causes are not necessary, but it also states that non-physical causes cannot enter into the physical realm and cause physical events. We can call this the principle of physical causal closure.<sup>6</sup> This amendment would block Marcus from supplementing the physical cause with a non-physical cause. However, Kim rejects this stronger principle on the grounds that it appears to be question begging (Kim, 2005, p. 52): Kim wants to prove that mental causes cannot have causal power, so he cannot start with the premise that mental causes cannot have causal power. Without this option available, Kim turns to the ‘no overdetermination/exclusion’ principle to rule out these potential cases of overdetermination.

That being said, I want to now consider the two arguments in favour of physical causal completeness, beginning with the argument from the success of science. Kim argues that scientists have discovered that neurons, a key correlate to consciousness, fire because of the movement of charged molecules (Kim, 2005, p. 154). This is simply one of many examples where scientists have found physical causes for things once considered independent of physical determination. Thus, it is reasonable to conclude that there will be a physical cause for everything that happens (Kim, 2005, p. 154-155). This

---

<sup>6</sup> Importantly, this terminology is not the same as the terminology deployed by Kim. What I have called causal completeness, Kim calls causal closure. What I have called causal closure, Kim calls strong causal closure. I have shifted the vocabulary, as a number of critics do, because there is nothing in Kim’s original definition which hints at the physical world being causally closed. Rather, it only hints at the physical world being causally complete. There is, however, something in Kim’s second, stronger, proposal that suggests that the physical world is closed to non-physical forces.

assumption has been frequently confirmed through further experimentation, so the conclusion that every physical event has a complete physical cause is reached (Melnyk, 2003, p. 238ff; Levine, 2001; McLaughlin, 1992; Papineau, 2001). David Papineau summarizes this argument from the success of science as follows:

During the first half of the century the catalytic role and protein constitution of enzymes were recognized, basic biochemical cycles were identified, and the structure of proteins analyzed, culminating in the discovery of DNA. In the same period, neurophysiological research mapped the body's neuronal network and analyzed the electrical mechanisms responsible for neuronal activity. Together, these developments made it difficult to go on maintaining that special forces operate inside living bodies. If there were such forces, they could be expected to display some manifestation of their presence. But detailed physiological investigation failed to uncover evidence of anything except familiar physical forces (Papineau, 2001, p. 31).

Scientists have successfully found physical causes for numerous physiological phenomena. Although neuroscientists do not yet have a complete understanding of the brain, it is reasonable to conclude they will eventually find a physical cause for all neural and psychological phenomena as well.

Not only have scientists successfully found physical causes for physical events, but they have been able to construct scientific theories that block the possibility of non-physical forces entering into the world as well. Commonplace laws such as the conservation of momentum and the conservation of energy suggest that the universe is causally closed. These laws have been verified upon numerous occasions, so they cannot be dismissed without good reason. A number of philosophers have used these conservation laws to rule out the existence of non-physical forces (Harbecke, 2008, p. 20;

Searle, 2004, p. 42; Flanagan, 1992, p. 21).<sup>7</sup> Likewise, Kim explains, “we expect the physical world to be causally self-contained” (Kim, 2005, p. 154). This principle rules out all foreign influence, but specifically it rules out the possibility that non-physical minds or properties can break into the causal stream of the world. And, if nothing can break into the causal stream of the world, it certainly stands to reason that the physical is causally complete. Again, the main participants in this debate accept this conclusion, so it is not necessary to pursue these matters further.

### **2.3 – The Argument for Irreducibility**

A third key premise that nonreductive physicalists adopt is that mental properties are distinct from physical properties. Section 1.2 briefly touched on two reasons to suppose that mental properties are distinct from physical properties. I want to flesh these out in greater detail now, and add a third argument. First of all, Saul Kripke has argued that identities are necessary (Kripke, 1972), meaning that they hold across this and all possible worlds. If  $M=P$ , then necessarily, wherever an instance  $p$  is, an instance  $m$  is also.<sup>8</sup> And, wherever  $m$  is,  $p$  is. No doubt, wherever  $p$  is, the supervening property instance  $m$  will be, on account of the dependency that  $m$  has on  $p$ . However, multiple realizability suggests the converse is not true. Multiple realizability is a principle popularized by Hilary Putnam (1967) and Jerry Fodor (1974), and currently endorsed by most people in the mental causation debate. It suggests that mental properties can be realized by various different physical bases. For example, the belief that fish are tasty is realized by a certain neural configuration in John, while it is realized by a slightly different neural configuration in Susan. This belief may be realized by an entirely

---

<sup>7</sup> For further discussion, see Montero (2006), who cites these and many other philosophers as advocates of this argument. For a reply to Montero, see Koksvik (2007).

<sup>8</sup> Following Marras, 2008, I will be using capital  $M$  to represent mental properties, and lower case  $m$  to represent instantiations of mental properties. The same scheme applies to physical properties.

different physical base in a shark. Plausibly, a robot may have this belief as well, even though it is realized by certain silicon chips in the robot. If the instantiation  $m$  can supervene on  $p_1$  in one organism, and it can supervene on a different physical property  $p_2$  in another organism, then  $p_1$  is not present where this  $m$  is, so  $P_1$  cannot be identical to  $M$ . Multiple realizability prevents the identities from holding across this world and other possible worlds, so  $M \neq P$ .

Donald Davidson provides a second reason to refrain from identifying mental properties with physical properties. He argues that mental vocabulary is essentially normative and rational, neither of which has a clear echo in physical vocabulary (Davidson, 2001, p. 231). Furthermore, Davidson argues that mental vocabulary uses intentional idioms (Davidson, 1970, p. 117), while physical vocabulary does not. Physical vocabulary, on the contrary, follows strict laws, but mental vocabulary must remain anomalous in order to preserve freedom (Davidson, 1970, p. 112). The distinctness of these vocabularies renders it impossible to translate or reduce one to the other without something important being lost in the translation (Davidson, 2001, p. 222). Without these requisite psychophysical laws, we cannot smoothly reduce mental vocabulary to physical vocabulary.

This discussion leads to a third reason to avoid identifying mental properties with physical properties. Owing to Leibniz' doctrine of the indiscernibility of identicals, the two items on either side of an equal sign ( $M=P$ ) are typically considered identical when they have only and all the same properties. Thus, if a mental property ( $M$ ) is a physical property ( $P$ ), then the mental property needs to have all the same properties as the physical property does, and vice versa.<sup>9</sup> For his own part, Kim agrees that this criterion must be met before the identity can go through:

---

<sup>9</sup> This manner of framing Leibniz' argument obviously relies on the ability of properties to have properties themselves. I will conclude that properties can have properties in section 4.3. For those not convinced by the arguments advanced

If the identity in the identity theory does not measure up in any way to the law of indiscernibility of identicals, then either the theory is false, or else we do not have an identity theory (Kim, 1972, p. 180).

The question we need to ask ourselves, therefore, is: does a mental property have any features that a physical property does not, and vice versa?

Davidson has suggested certain conceptual differences between the mental and the physical. Although Davidson does not consider these conceptual differences a matter of ascribing properties (for reasons that will become clear in Section 7.2), many, including Kim (Brandt and Kim, 1967, p. 518), do. If these predicates do refer to properties, then we have already seen certain distinct features of the mental and physical respectively. Namely, mental phenomena have intentional properties, rational properties, and normative properties, while physical phenomena do not (McGinn, 1991, p. 23-24; Silberstein, 2001, p. 85; Lowe, 2006, p. 15-16). Mental phenomena are anomalous, free and purposive, while physical phenomena are law-like, deterministic and non-purposive. Plausibly, physical properties are spatial and extended, while mental phenomena are not. Others suggest that mental phenomena have private, or qualitative properties, while physical phenomena are publically observable, or third-personal (Nagel, 1974, p. 437; Chalmers, 1996). Kim eventually uses this last distinction to argue that mental properties are not entirely reducible to physical properties (Kim, 2005, p. 169-170).

These differences have led some to conclude that there is an “intuitive distinction” (Harbecke, 2008, p. 26. See also Rey, 1997, p. 48-57) between mental and physical properties, while it has led others to consider the identity between mental and physical properties “implausible,” or even “inconceivable” (Slors and Walter, 2002, p. 1). In sections 5.2.a and 8.4, I consider additional distinctions between mental and physical properties that may further solidify the irreducibility thesis.

---

there, we could reframe Leibniz’ argument in terms of events: before two events on either side of the equation can be identical they need to have only and all the same properties.

Before moving on, I want to clarify one important point with respect to the principle of irreducibility. I have given three reasons to refrain from identifying mental and physical properties. As I will show in Section 6.1, according to Kim, if a mental property is distinct from a physical property, then a mental event is distinct from a physical event as well. Section 6.2 then shows how this additional step is rejected by many nonreductive physicalists. In that section I will address the viability of rejecting Kim's additional step, as well as the possibility of resolving the causal exclusion problem by rejecting this additional step. For the time being, however, since Kim frames the causal exclusion problem in terms of events, I will grant Kim this conclusion, and assume that irreducibility also entails that mental events are not physical events.

#### **2.4 – The Argument for No Overdetermination/Causal Exclusion**

I have gathered the principle of causal exclusion together with the intuition that overdetermination should be avoided. I have done this because they say largely the same thing. Although they are similar, I do not want to beg the question by assuming that a nonreductive physicalist will endorse the principle of causal exclusion. To alleviate this concern, let us consider overdetermination on its own merits. A murderer is sentenced to death by the firing squad. When his day comes two bullets pierce his heart, but either one is sufficient in itself to cause the death, so only one is necessary. This is a genuine case of overdetermination. How do we know the death is over-determined? Kim typically uses a counterfactual analysis to decipher whether an effect is over-determined or not (Kim, 1998, p. 44-45; Kim, 2005, p. 46-49). Would the death still occur if only one bullet was fired? If the answer is yes, then the death is overdetermined. One bullet is sufficient to cause the death, but in this case two bullets cause the death.

Perhaps mental causation operates in much the same way? The physical cause is sufficient to bring about an effect, but in this case a distinct mental cause also affects the event. We have seen several people endorse such a version of mental causation. Furthermore, we have seen that Kim's articulation of physical causal completeness does

not prevent this solution. Is this a viable solution, then? Although I will consider this issue more deeply in chapter four, there are two initial reasons to think not: first, this solution implies the existence of massive amounts of coincidental overdetermination, and second, this solution implies that neither physical causes nor mental causes, taken in isolation, are necessary causes for the occurrence of an event.

Consider the problem with massive overdetermination first. Mental events purportedly cause physical events hundreds of times a day in each person, multiplied by billions of people in the world. We would have to admit that the world contains massive amounts of overdetermination if we accept it for mental causation. This is problematic because the amount of coincidence involved strains credulity. To think that two divergent causal paths constantly converge is akin to believing that lightning regularly strikes the same neutral venue. It simply defies the odds, so people do not think it will occur. Homicide detectives, for example, after finding a cause for a death, do not wait around to find an anticipated second or third cause. Or again, doctors, after finding the cause behind certain symptoms, do not continue searching for second and third causes for the symptoms. They do not rummage around for additional causes because they will not find any. This being the case, it is not attractive to have to postulate a world where overdetermining causes constantly appear.

As for the second reason, if overdetermination is accepted then neither mental causation nor physical causation is a necessary condition for the functioning of our daily lives (Kim, 1998, p. 44-45; Kim, 2005, p. 48-49). As I will argue in Section 4.1, if the physical cause is sufficient, then the mental cause is by definition unnecessary. Intuitively, however, we think that if we hadn't decided to make chicken for supper our bodies would not make chicken for supper anyway. But on overdetermination our lives would proceed as usual without our mental events playing any causal role (I should note that this is part of the reason why autonomous mental causation requires that the mental play a necessary role). Moreover, if the mental cause is sufficient, the physical cause is not necessary for the functioning of our daily lives either (Kim, 1998, 45). If the

appropriate neural commotion had not occurred then we would still make chicken for supper anyway, since the mental cause is sufficient. This possibility violates causal completeness, and not surprisingly, Kim rejects overdetermination for this reason.

In time I will outline the host of issues surrounding these two arguments, but for now I simply conclude that overdetermination should be avoided if at all possible. Numerous philosophers have agreed that escaping from the problem of mental causation by simply embracing overdetermination is less than ideal (Harbecke, 2008, p. 28; Schiffer 1987, p. 148). And Kim, for his own part, calls the possibility of overdetermination “at best, extremely odd,” (Kim, 1993, p. 247) and at worst, “absurd” (Kim 1993, p. 281).

As mentioned, the principle of causal exclusion is similar to the rejection of overdetermination. Those persuaded that overdetermination should be avoided, therefore, may be convinced that causal exclusion is evident as well. Jaegwon Kim, for example, thinks that causal exclusion is “virtually an analytic truth with not much content” (Kim, 2003, p. 163). One cannot simply assert that causal exclusion is self-evident however, so let us take a look at exactly what it says, and consider how it is established.

The principle of causal exclusion has two parts to it: causation and exclusion. I will consider details of Kim’s view of causation in Section 7.1. What about the principle of exclusion? Exclusion implies that something is left out. What is in danger of being left out? With respect to causal exclusion, once one event is considered sufficient for a given effect, any additional event is excluded as a cause. As Kim explains: “If an event  $e$  has a sufficient cause  $c$  at  $t$ , no event at  $t$  distinct from  $c$  can be a cause of  $e$ ” (Kim, 2005, p. 17).<sup>10</sup> If a certain cause is sufficient to bring about an effect on its own, then no other cause need be invoked, and indeed, no other cause can be invoked. Thus, for example, if

---

<sup>10</sup> Kim supplements this principle with the proviso: “unless it is a genuine case of causal overdetermination”. I will discuss this caveat in chapter four.

an electrical malfunction is a sufficient cause for why the house burnt down, we need to exclude any other distinct potential causes, such as the wood burning in the fireplace. Or, in the relevant case, if neural excitation in the brain is a sufficient cause for Joe waving his hand, then we need to exclude any distinct mental cause, such as Joe's desire to draw attention to himself.

Why would anyone embrace such a principle? Imagine we think that the wood burning in a fireplace is a cause of a fire. We can now ask: would the house still have burnt down if the fireplace was not in use that night? As it turns out, an electrical malfunction is the sufficient cause for the fire, so the answer is yes. This being the case, the burning wood in the fireplace is irrelevant to whether the house burns down or not, and there is no reason to call an irrelevant event a cause of the house burning down (Kim, 1989, p. 82; Kim, 1998, p. 44-45). Beyond this, causal exclusion is supported by the aforementioned arguments against overdetermination.

## 2.5 – The Problem of Causal Exclusion

To recapitulate, the nonreductive physicalist typically adheres to the following four intuitions: (1) mental causation, (2) physical causal completeness, (3) irreducibility and (4) no overdetermination/causal exclusion. These four intuitions, however, possibly form an inconsistent tetrad. We can accept any three, but it is more difficult to accept all four. Indeed, any three propositions put together potentially render the fourth false.

For instance, we can accept causal completeness, irreducibility and causal exclusion. Jointly, however, these three propositions lead to the falsity of mental causation. If every event has a sufficient physical cause, and mental causes are not identical to that sufficient cause nor can we include more than that one sufficient cause, then we are forced to conclude that mental causes do not help generate that effect. This position, known as epiphenomenalism, has a number of adherents (Tammelleo, 2008; Lyons, 2006). However, the loss of mental causation is too steep a price to pay for most to consider this a live option.

We can accept mental causation, irreducibility and causal exclusion. However, these premises together lead to the falsity of causal completeness. If mental events play a distinct causal role in the occurrence of an effect, but this fact does not overdetermine the effect, then it follows that no sufficient physical cause was available for the effect in the first place. There are those who embrace a dualism of this sort (Meixner, 2008; Lowe, 2006). Again, however, the rejection of closure is widely considered too large a sacrifice to pursue this option.

We can accept mental causation, causal completeness and exclusion. This is Kim's position, and he thinks it implies the failure of irreducibility: if every event has a sufficient physical cause which in turn excludes all distinct causes, yet the mental is still a cause as well, it seems as though the mental cause must be identical to the sufficient physical cause. Kim is willing to take this route, but others continue to prefer nonreductive physicalism to what has been called an “indigestible metaphysics” (Loewer, 2002, p. 661).

Finally, we can accept causal completeness, irreducibility and mental causation. This results, however, in the rejection of causal exclusion: if an event has a sufficient physical cause and a distinct mental cause, then we must be willing to allow events to have more than a single sufficient cause. Although I have listed several reasons to endorse causal exclusion above, I will also deal with the problems of taking this route in greater detail in chapter four.

## 2.6 – The Problem of Explanatory Exclusion

Before turning to various possible solutions to the problem of causal exclusion, I want to introduce the parallel problem of explanatory exclusion. Explanatory exclusion states that “there can be no more than a single *complete* and *independent* explanation for any one event” (Kim, 1988, p. 233). If every event has a complete neurological explanation, a mental explanation must either be dependent upon the neurological explanation, or it must be excluded from the explanatory account of the event.

How does Kim defend the translation of the problem of causal exclusion up into the realm of explanation? He does so with the help of the principle of Explanatory Realism. In its barest form, explanatory realism says “*C* is an explanans for *E* in virtue of the fact that *c* bears to *e* some determinate objective relation *R*” (Kim, 1988, p. 226). This objective causal relation *R* between events “grounds” the explanans relation between statements and serves as its “objective correlate” (Kim, 1988, p. 226). When we add this principle of explanatory realism to causal exclusion we get explanatory exclusion: causal exclusion states that there can be no more than one sufficient cause for event *e*, and explanatory realism says that this exclusive causal relation correlates with and grounds explanations (Gibb, 2009, p. 3; McIntyre, 2002, p. 95). Hence there can be no more than one complete and independent explanation for event *e* either.

What reason is there to accept this principle? After all, even Kim acknowledges that “many will consider [it] absurdly strong and unacceptable” (Kim, 1989, p. 79). True to form, numerous respondents have taken exactly this attitude towards it. For example, JeeLoo Liu says “taken by itself, explanatory exclusion seems too strong. And indeed, it has aroused some suspicion that Kim is creating a pseudo-problem” (Liu, 2001, p. 10). Also see McIntyre, 2002, p. 92-93; Worley, 1993, p. 334; Marras, 1998, p. 439; and Sabates, 1996, p. 96). As justification, Kim argues that a certain perplexity overtakes us once there are two different explanations of the same event (Kim, 1989, p. 92). There is, after all, only one world, and if we have already given one complete explanation of everything in the world that caused an event, it would be difficult to understand how a different story of the same event could also be plausible (Kim, 1989, p. 93-94; 96). This being the case, endorsing only one explanation which tracks a simple and unified world generates a simple and unified belief system, which helps to alleviate potential confusion (Kim, 1994, p. 66-69). Kim also says the principle of explanatory exclusion is similar to the revered principle of Ockham’s razor, which also tries not to multiply explanations beyond necessity (Kim, 1989, p. 98).

The principle of explanatory exclusion poses a challenge to many nonreductive physicalists as well. Donald Davidson, for example, argues that explanations are individuated intensionally (Davidson, 1980, p. 171). This means that one explanation is considered distinct from a second explanation if they are intensionally or epistemically inequivalent. A number of participants in the debate about mental explanations adhere to this view (Marras, 1998, p. 443; Fuhrmann 2002, p. 184ff; Campbell, 2008, p. 86).

But psychological explanations are intensionally inequivalent to neurological explanations – Socrates can have a psychological explanation for his behaviour without thereby having the neurological explanation for this behaviour as well. Kim willingly acknowledges this much (Kim, 1988, p. 233; Kim, 1989, p. 81). Intensional inequivalence, however, is the mark of distinctness between explanations. Take Kim's chief target, Norman Malcolm, for example. He argues that psychological explanations are independent from physiological explanations on account of their “logical difference” (Malcolm, 1968, p. 51). There is, therefore, a “collision between the two accounts” (Malcolm, 1968, p. 52) if they are both explanations of the same occurrence. Those who think that the physical explanation is complete, and the mental explanation is distinct from this physical explanation, will have to exclude the mental explanation.

In summary, the problem of causal exclusion arises when we attempt to conjoin four independently plausible propositions. Accordingly, we will either have to find a way to reconcile these four intuitions, or we will have to reject one of them. In the next chapter I consider various attempts at reconciliation, while in the two chapters that follow I consider in detail certain attempts to jettison one of these four principles. In this chapter I also introduced to the partner problem of explanatory exclusion, which will receive an initial resolution attempt in Section 5.1.

### **3 – Solutions to the Problem of Causal Exclusion**

In Chapter two, a tension developed between the combination of the following four principles: (1) mental causation; (2) physical causal completeness; (3) irreducibility; and (4), causal exclusion. Not surprisingly, there are a host of available methods of dissolving this tension. Although I cannot deal specifically with every attempt at reconciliation, there are several prominent schools of thought that are worth investigating.

This chapter is divided into five sections. In Section 3.1, I consider the supervenience solution to the problem of causal exclusion. The supervenience solution argues that completeness is true, while the mental irreducibly supervenes upon the physical. Because of the supervenience relation, the mental in some sense inherits the causal power of its subvening base. Thus, we can have mental causation without overdetermination. In Section 3.2, I outline and endorse Jaegwon Kim's objection to this argument. I also argue that this supervenience solution does not deliver autonomous mental causation for us. In Section 3.3, I consider emergentism as a possible reply to the problem of causal exclusion. In Section 3.4, I once again outline and endorse Jaegwon Kim's objection to emergentism. I also note that emergentism does not deliver autonomous mental causation either. In Section 3.5, I consider the constitutionalist reply and certain variations thereof. I conclude that the constitution reply neither eludes the causal exclusion problem, nor does it provide autonomous mental causation.

#### **3.1 – The Supervenience Solution to the Problem of Causal Exclusion**

The most longstanding, and probably the most popular, method of knitting these four principles together is through mind-body supervenience. Indeed, Jaegwon Kim himself once used supervenience to reconcile causal completeness with the causal potency of unreduced mental properties (Kim, 1984).

Supervenience is a concept normally attributed to G. E. Moore, although Donald Davidson points out that Moore does not use the term himself (Davidson, 1993, p. 188).<sup>11</sup> Moore's proto-supervenience thesis was that moral properties were distinct from natural properties, but still rooted in natural properties. Davidson then popularized the idea that mental properties supervene on physical properties. By this he meant that any two events that were alike in all physical characteristics were alike in all mental characteristics (Davidson, 1980, p. 214). Jaegwon Kim then spent some time analyzing the supervenience relation before eventually defining it as follows:

Mental properties strongly supervene on physical/biological properties. That is, if any system  $s$  instantiates a mental property  $M$  at  $t$ , there necessarily exists a physical property  $P$  such that  $s$  instantiates  $P$  at  $t$ , and necessarily anything instantiating  $P$  at any time instantiates  $M$  at that time (Kim, 2005, p. 33).

Kim argues here that supervenience entails, at the very least, property covariation. That is, everywhere the physical properties are, the mental property will be as well; and everywhere the mental property is, some physical properties are sure to be as well (Kim, 1993, p. 85ff; Campbell, 2000).

There must be more to supervenience than mere property correlation, however, for even parallelists and epiphenomenalists accept this property correlation thesis. Indeed, supervenience also implies that the mental correlates with the physical because the mental asymmetrically depends on the physical, and the physical determines the mental:

I take supervenience as an ontological thesis involving the idea of dependence ... a mental property is instantiated in a given organism at a time *because*, or *in virtue of* the fact that, one of its physical 'base' properties is instantiated by the organism at that time (Kim, 2005, p. 34; See also Kim, 1993, p. 67; Campbell, 2000).

---

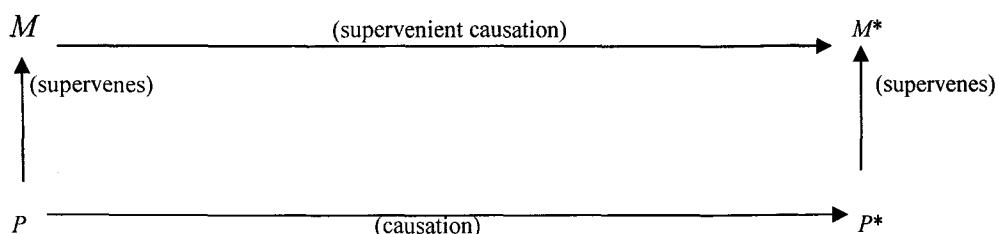
<sup>11</sup> See McLaughlin, 1997, p. 17-19 for a detailed account of the history of the term 'supervenience'.

Mental properties are fixed by physical properties because physical properties determine the nature and existence of mental properties. Thus, for example, the shadow of a person appears whenever there is a person (assuming, of course, daylight, etc...) because the shadow depends upon, and is determined by, the person who casts the shadow (though the shadow is not the person). Or, to use Kim's example, the goodness of a person depends upon, and is determined by, the courage, benevolence and honesty of the person – but is not identical to these traits (Kim, 1993, p. 65).

This notion of supervenience implies a tight enough dependency relation for Kim to have concluded that mental properties inherit causal powers from their subvening physical bases:

If a pain causes the sensation of fear an instant later, this account tells the following story: the pain is supervenient on a brain state, this brain state causes another appropriate brain state, and given this second brain state, the fear sensation must occur, for it is supervenient upon that brain state ... It seems to me that this is sufficient to redeem the causal powers we ordinarily attribute to mental events (Kim, 1993, p. 106-107).

Pain causes fear because pain is so profoundly bonded with a certain brain event, and this brain event causes the occurrence of the brain event that fear is superveniently bonded to. This model of supervenient causation can be diagrammed as follows (where  $M$  stands for a mental property and  $M^*$  stand for a mental property that occurs at a later time, while  $P$  stands for a physical property and  $P^*$  stands for a physical property that occurs at a later time):



**Figure 2: Supervenient Causation**

Mental properties superveniently cause other mental properties to occur. But, this does not prevent physical properties ( $P$ ) from being the sufficient cause of  $P^*$  (thereby retaining causal completeness). Moreover, mental properties remain distinct from physical properties, so irreducibility is retained. Finally, this does not seem to lead to overdetermination, for  $P$  is still the unique cause of  $P^*$ , and  $M$  does not contribute anything over and above  $P$ . This solution, which is echoed by a number of contemporary nonreductive physicalists, apparently allowed Kim to reconcile all four of the aforementioned intuitions.

### **3.2 – Problems with the Supervenience Solution**

Before long, Kim concluded that this supervenience solution is fraught with difficulty. The problem: despite the fact that  $M$  is there, it is  $P$  that is doing all of the causal work. Why? The notion of causal relevance can be of assistance here. In his dispute with Donald Davidson, Kim says that a mental property may be causally relevant without being causally efficacious (Kim, 1993, p. 23). By being causally relevant, Kim means that the presence of the mental property guarantees that an effect will occur, but it is not the mental property that does the causal work. Rather, the physical property does the causal work while the mental property merely comes along for the ride. A moving plane, for example, may always be accompanied by noise, but we can doubt that it is the noise that makes the plane move. Or again, a shadow may always accompany a car moving down the highway, but it is not the shadow that is moving the person down the road.

In the same way, a supervening property may always accompany a causally potent micro-physical property, but it is still the micro-physical property doing all of the causal work, while the mental property, though intimately connected, is merely coming along for the ride:

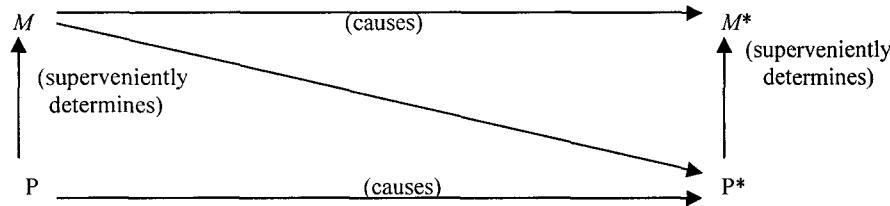
So why not say that  $M$ , though it doesn't quite have the causal status of  $P$  in relation to  $P^*$ , is a derivative cause of  $P^*$  in virtue of its supervenience on  $P$ ? ...

Some years back, I thought that this might be a plausible way of vindicating

mental causation. But it soon began to dawn on me that this was an empty verbal ploy ... this is only a gimmick with no meaning; the facts are [that  $P$  causes  $P^*$ ], and inserting [ $M$  causes  $M^*$ ] and calling it supervenient causation ... does not alter the situation one bit (Kim, 2005, p. 62).

An honest look caused Kim to see the mental was not really doing anything itself. For we can plainly see that  $M$  is not *really* causing  $M^*$ . Rather,  $P$  is actually causing  $P^*$ , which in turn determines  $M^*$ . Moreover, we can see that  $M$  is not *really* causing  $P^*$  either, for  $P$  is causing  $P^*$  to occur.  $M$  is left without any causal power of its own (Kim, 1998, p. 37).

In other words, in order for mental causation to occur, we would need a causal arrow to depart from  $M$ , the mental property. Let us try to add causal arrows departing from  $M$ , and see what happens:



**Figure 3: The Problem of Causal Exclusion**

According to Figure 3,  $M$  now has causal arrows pointing in two different directions. If these arrows can be preserved, then we will have a case of mental causation.

Unfortunately, these arrows must be crossed out. It turns out, in the first case, that  $M$  cannot cause  $P^*$ . Why? From causal completeness we see that  $P$  is a sufficient cause for  $P^*$ . This being the case, there is no work left for  $M$  to do. But perhaps we could still include it as a cause, for posterity's sake, or perhaps just out of nostalgia? Here is where Kim's causal exclusion principle comes in: since  $P$  is a sufficient cause for the occurrence of  $P^*$ ,  $M$  is excluded from being a cause because we cannot admit to the overdetermination of  $P^*$ .

Granted, then, that  $M$  does not cause  $P^*$ ; but maybe  $M$  does still cause  $M^*$ ? As it turns out, however, from the very definition of supervenience,  $P^*$  fully determines  $M^*$  itself. Despite this setback, perhaps  $M$  can still assert itself as a cause of  $M^*$ ? Unfortunately, this is not possible. Since  $M^*$  is fully determined by  $P^*$ ,  $M$ 's impotence is again forced upon us by the exclusion principle. We cannot admit that  $M^*$  is over-determined, so  $M$  has to be left out as a cause. No causal arrow can be forged between  $M$  and a potential effect, so mental causation is once again in jeopardy. As Kim summarizes:

The supervenience/exclusion argument shows that property dualism is not able to explain how mental causation is possible. Instead of saving mental causation, it ends up relegating mental phenomena to the status of epiphenomena.

Nonreductive materialism has been motivated by a desire to save mentality as something distinctive and special, and something that we value. Instead of saving it, it loses it by depriving it of causal powers (Kim, 2005, p. 158).

Notice that Kim also calls the causal exclusion argument by the name of ‘the supervenience argument’ here. This is so because causal exclusion declares that all supervening properties are causally irrelevant in the way we have just seen. The causal exclusion argument, therefore, is an argument against all of those (including Kim, himself) who think the mental supervenes on the physical.

Before moving on, I want to make one note about the supervenience solution. The supervenience solution attempts to offer us autonomous mental causation, which is the stated goal. It attempts to offer us autonomous mental causation by rendering the mental distinct from the physical. However, it fails to provide a distinct role for this distinct mental cause because the physical cause is already doing all of the causal work. Any causal role that the distinct mental cause plays would have to be inherited from the physical cause, and so the mental cause cannot play a distinct role on its own. The supervenience solution therefore fails to offer us autonomous mental causation because

autonomous mental causation requires not only that the mental be distinct, but that the mental play a distinct role as well.

### **3.3 – The Emergentist Solution to the Problem of Causal Exclusion**

Partly due to these concerns, alternative models of reconciling the four central intuitions of the causal exclusion problem have arisen. One recently popular alternative to the supervenience solution is emergentism. There is no need to venture into all of the details of emergentism, rather only several pertinent features are of use.

The central emergentist claim is that novel properties emerge, or arise, out of a base level.<sup>12</sup> The existence of a base level which gives rise to emergent properties implies a layered ontology, which all emergentists endorse (Kim, 1999, p. 19). Although I will spend some time on this layered ontology in chapter eight, a rough outline is necessary here. A layered ontology consists in cutting the world up into mereological levels, beginning with a base level of particles and their microphysical properties. Certain configurations of these basic particles give rise to higher level wholes, such as molecules and their properties. A specific group of molecules in turn give rise to even higher level biological objects and their properties. Although emergentists do not completely agree on how the world is layered, they do all agree that the world is layered.<sup>13</sup> Kim, for his own part, occasionally endorses this layered model (Kim, 1997, p. 291), though I will explore his somewhat checkered relationship with this model later.

---

<sup>12</sup> To be more precise, not all emergentists insist that the emergent entities are properties. Achim Stephan suggests that we can have emergent properties, emergent laws, emergent effects, or emergent events, depending on the exact model (Stephan, 1992, p. 27). But, for our purposes, we will only need to talk about emergent properties.

<sup>13</sup> Lloyd Morgan, for example, posits a base level of atoms, a higher level of molecules, a third level of plants, then of animals, and finally of human beings (Kim, 2002, p. 6 of 20). Samuel Alexander suggests that the bottom level is space-time, followed by matter, molecules, life, mind and finally, deity (Clayton, 2006; p. 23-24). Finally, Charles Broad suggests that the lowest order is physics, then chemistry, biology and psychology (McLaughlin, 1992, p. 79).

With this background in mind, we can now inquire into how an emergent property arises. To start off, there are fundamental particles in the base domain, and they have certain properties. As noted, sometimes these fundamental particles gather together to form a “combination” (Mill, 1843, p. 371) or a “composition” (Mill, 1843, p. 435).<sup>14</sup> Kim, in his analysis, characterizes the same thing in terms of “structured aggregates” (Kim, 1992, p. 123). For the emergentist, then, we have base level particles, and then we have the combination, or structure, of these particles. When these parts are appropriately structured, new and unexpected properties start to be instantiated as well. These new properties are emergent properties, and they are properties of the whole system (i.e., the macro object), and are not possessed by any of the parts of the object (Stephan, 2002, p. 80). For example, sodium has properties such as being silvery and metallic, while chlorine has the properties of being greenish, gaseous and toxic. Neither of them have the property of regulating water content in human bodies. However, when appropriately arranged, the compound sodium chloride “is an essential compound for life, not to mention potato chips” (Rothschild, 2006, p. 153). Salt has the property of regulating the water content in human bodies, whereas the constituents of salt, in isolation, are toxic to the body and do not.<sup>15</sup> Thus, the macro-object (i.e., the salt) has the emergent property of being able to regulate water content in humans.

---

<sup>14</sup> Samuel Alexander describes this gathering as a “constellation or collocation” (Alexander, 1920, p. 45), while others label it an “organization” of particles (Moreno and Umerez, 2000, p. 99; Murphy, 2006, p. 229) or, more simply put, an “organism” (El-Hani and Emmeche, 2000, p. 235; Alexander, 1920, p. 62). McLaughlin, in his outline of British emergentism, suggests that we should consider this a “complex configuration” (McLaughlin, 1992, p. 67).

<sup>15</sup> As another example, at the most fundamental level, some argue that the collapse of the waveform that arises when an electron bounces off of a target is emergent. In isolation, this waveform will continue to spread out, but this does not happen. Rather, due to environmental interactions the waveform decoheres into an electron at a location (Davies, 2006, p. 37; Joos, 2006, p. 63; Humphreys, 1997, p. 15). Thus the macro system has a property that the parts in isolation do not have. To use a less natural example, Olga Markic suggests a bicycle has an emergent property of being a transporter: “Are the bike and its ability that someone can ride it something more than the parts? If we play our own

Emergentists think that these higher level emergent properties exercise causal power. Indeed, Alexander's Dictum, which states that if a property does not have causal power, then it does not exist, is named after the British emergentist Samuel Alexander. But what sort of effects do these emergent properties bring about? The most profitable and common response is that emergent properties cause certain lower level effects to occur.<sup>16</sup> This model, which posits a higher level emergent property exerting a causal

---

mechanic and have all the parts lying disassembled in front of us, there would be no ability for transport, no bike that will carry a rider through the woods or downtown. There will be just a lot of pieces that could be assembled as a bike. The important point is that the parts have to be structurally arranged in the right way to form a bike. So, it may seem that a bike and its property that someone can ride it, emerge from the structure of the parts and can not be explained by the knowledge of the parts alone" (Markic, 2004, p. 66. See also Koslicki, 2008, p. 3). In this case, the same parts can be arranged heap-wise on a factory floor, and these parts can lack the property of being a transporter. However, these same parts can be arranged bicycle-wise on a street, and these parts so configured will have the property of being a transporter. Others suggest the biological property of being able to metabolize is emergent because a human being has this property "by virtue of the particular combination and arrangement of its constituents, but which is not possessed by the constituents in isolation or in other non-living combinations" (Jones, 1972, p. 425. See also Francescotti, 2007, p. 58). The many parts do not have this property in isolation, but rather, once these parts are configured in a certain way the whole system gains this property. With these, and many other, examples of emergence currently being discussed in the literature, Silberstein and McGeever note that "it would seem that emergence is back with a vengeance" (Silberstein and McGeever, 1999, p. 184-185).

<sup>16</sup> As an alternative, some suggest that higher level properties cause other higher level effects (Gibbons, 2006). For example, the force of the baseball causes the window to crash. The benefit of this model is that it may not violate the completeness of physics, for it has no lower level effects. On this model, the lower level properties of the parts of the baseball fully cause the lower level properties of the window to be instantiated, while the higher level force of the baseball itself causes the higher level window to crash. This model faces several difficulties. First of all, it is akin to a Leibnizian parallelism where the higher level interacts with the higher level, while the lower level interacts with the lower level (Harbecke, 2008, p. 402-403). When transported into the mental domain, parallelism seems counterintuitive. We think that our mental properties not only have mental effects, but have physical effects as well, such as the movement of limbs. It is not enough to say that mental properties only cause other mental properties to be instantiated (Paul, 2007, p. 8-9). Secondly, assuming supervenience, Kim's exclusion argument claims that higher

influence on lower level properties is sometimes called “downward causation” (Campbell, 1974). On this downward causal model, the life of the puppy is said to have an effect on the parts that make up the puppy. If the puppy were not alive, the parts would do something different than what they would do if the puppy were alive – perhaps the parts would separate spatially, or at least the parts would not move in the same general direction. Jaegwon Kim argues that this doctrine of downward causation is in fact the *raison d'être* of emergentism (Kim, 2006, p. 548).

From these basic principles we can see an outline of the emergentist solution to the problem of causal exclusion. First of all, we can conceive of mental properties as being emergent, higher level properties of a whole mind. This mind may be made up of lower level neural parts, but since mental properties are not properties of these parts they are not reducible to the properties of these parts. Emergent mental properties secure mental causation by exercising downward causal efficacy on the lower level parts of the body. Emergentism secures physical causal completeness by insisting that higher level properties arise out of lower level physical properties, so they are dependent upon, and determined by the lower level physical properties.

### **3.4 – Problems with the Emergentist Solution**

Integral to the aforementioned emergentist solution to the problem of causal exclusion is the view that emergent properties emerge out of lower level physical properties, and so are dependent upon the lower level physical properties. Indeed, this rootedness often lends itself to the suggestion that emergent properties are supervenient upon their physical parts. For example, Timothy O'Connor defines an emergent property as follows:

---

level properties will only be able to cause other higher level properties by first causing the lower level properties that determine the existence of the higher level property. We will consider this point in greater detail below.

*A*-properties of objects supervene on *B*-properties of their parts =<sub>def</sub> Necessarily, for any object *x* and *A*-property *a*, if *x* has *a*, then there are *B*-properties *b, c, d, ...* (including relational properties) such that (i) some proper parts of *x* have (variously) *b, c, d, ...* and (ii) necessarily, for any things collectively having all of *b, c, d, ...* there is an object of which they are parts that has *a*. This strong form of supervenience is well suited to an account of property emergence (O'Connor, 1994, p. 97. See also Markic, 2004, p. 75-76).

Here we see that a property is emergent if, among other things, it strongly supervenes upon the micro properties of the underlying parts. Warren Shrader suggests that characterizing emergent properties as supervenient properties is essentially the traditional emergentist model endorsed by Broad and Morgan, though not in these exact words (Shrader, 2009, p. 9). For their part, a number of critics characterize the emergentist position as endorsing some form of supervenience as well (McLaughlin, 1997, p. 16; Kim, 2006, p. 549-550).

Framed in this way, emergentism bears a striking similarity to the supervenience solution to the problem of causal exclusion. Indeed, a number of authors argue that emergentism is roughly equivalent to contemporary nonreductive physicalism (McLaughlin, 1997; Shoemaker, 2002; Van Cleve, 1990; Sperry, 1991; Searle, 1992; Humphreys, 1996; Crane, 2001). For our purposes, it is important to note that Jaegwon Kim is likewise convinced of the similarities between the two theories. As he says

Along the way I shall develop a parallelism between emergentism and an influential current position – arguably the orthodoxy of today – on the mind-body problem ... the parallelism is important because it will show that those who accept the current orthodoxy, viz., the nonreductive physicalists, are committed to downward causation just as much as the emergentists and for the same reasons, and their fate, too, is intimately tied to the tenability of downward causation. It will become clear, if I am right, that nonreductive physicalism is a form of emergentism (Kim, 1992, p. 121. See also Kim, 2006, p. 549-550).

Supervenient emergentism believes that certain properties, though irreducible to lower level properties, supervene on lower level properties.

Of course, Kim uses the causal exclusion principle to argue that nonreductive physicalism leads to epiphenomenalism. Since emergentism is similar to nonreductive physicalism, we may not be surprised to learn that Kim uses the causal exclusion argument against emergentism as well. The emergentist acknowledges that a given mental effect is determined by an underlying physical effect. Thus, in order for mental events to cause other mental events, the first mental event will have to cause the underlying physical event of the second mental event. The rest of the problem reads as follows:

So  $M$  causes  $P^*$ . Now,  $M$ , as an emergent, must itself have an emergence base property, say  $P$ . Now we face a critical question: if an emergent,  $M$ , emerges from basal condition  $P$ , why cannot  $P$  displace  $M$  as a cause of any putative effect of  $M$ ? Why cannot  $P$  do all the work in explaining why any alleged effect of  $M$  occurred? ... This appears to make the emergent property  $M$  otiose and dispensable as a cause of  $P^*$ ; it seems that we can explain the occurrence of  $P^*$  simply in terms of  $P$ , without invoking  $M$  at all (Kim, 2006, p. 558. See also, Kim, 1999, p. 32).

The emergentist accepts supervenience, so the first mental event ( $M$ ) also has a physical base ( $P$ ). This underlying physical event seems to be a sufficient cause for the physical event  $P^*$ , so the original emergent mental event is excluded as a cause of  $P^*$ . In this way, the emergentist shares with the nonreductive physicalist the fate of epiphenomenalism. For their own part, a number of critics agree that the causal

exclusion argument poses a significant problem for emergentism in this manner (Bedau, 1997, p. 377; Stephan, 1999, p. 65; Bennett, 2008, p. 282; Crane, 2001, p. 219).<sup>17</sup>

For so long as emergent properties are construed as supervenient properties, emergentism faces the same difficulties that supervenience based solutions to the problem of causal exclusion face. In the same way that supervenient mental events are excluded because the subvening physical causes are sufficient, so are emergent mental properties excluded because the lower level physical causes are sufficient. In the same way that supervenience could not secure autonomous mental causation because it could provide a distinction between the mental and the physical but it could not provide a distinct role for the mental, so emergentism fails to provide autonomous mental causation because it provides distinctness without a distinct role as well.

### **3.5 – The Constitution Solution to the Problem of Causal Exclusion**

A third response to the causal exclusion problem can be called the Constitution view, which, when formulated broadly, is endorsed by a number of philosophers (Pereboom and Kornblith, 1991; Yablo, 1992; MacDonald and MacDonald, 1995; Baker, 1998; Baker, 2000; Antony, 1999; Shoemaker, 2002; Pereboom, 2002; Rueger, 2004; Rueger, 2006; Paul, 2007; Wilson, 2009). Plausibly, the supervenience relation (or the emergence relation) between the mental and the physical is not very tight, and so the mental cause is

---

<sup>17</sup> There are a number of available replies to this argument. For example, a number of emergentists argue that a whole macro object fuses the parts together in such a way as to annihilate the parts (Humphreys, 1997; Toner, 2008). If the parts are destroyed during the fusion into a whole, then there is no causal competition between base properties and the emergent property because there are no base properties anymore, so the emergent property is free to do all of the work. A number of objections have been raised to this view (Welshon, 2002; Shaffer, 2003). For example, Rex Welshon argues that not all fusions appear to be destructive. Thus, the existence of a water molecule does not destroy the existence of the hydrogen atom or the oxygen atom, it simply means that they are temporarily bound together. They could be detached and continue their independent existence. It seems more probable to conclude that they continue to exist while bonded, than that they are annihilated and then are restored upon separation.

screened off due to the causal sufficiency of the physical cause. The constitutionalist suggests that we need to find a tight enough relation between the mental and the physical that the closeness of this relation will guarantee that the mental cause is not screened off. This relation, which is nearly as intimate as identity, is called the constitution relation.

The most commonly invoked relation of constitution is mereological constitution – though this is not without exception (Baker, 1998; Baker 2000). Derk Pereboom and Hilary Kornblith are typically considered the first to attempt a resolution to the problem of causal exclusion using an appeal to constitution (Pereboom and Kornblith, 1991; Pereboom, 2002). In the later work, Pereboom argues a token whole cannot be identical to its token parts in virtue of their distinct modal and temporal parts. For example, he argues that Theseus' ship both could, and across time, does, remain the same even though the parts constituting Theseus' ship might not, and across time do not, remain the same (Pereboom, 2002, p. 503).<sup>18</sup> For this reason, we cannot posit a token identity of the ship with its parts, either at this time, or across time. In the same way, mental tokens are not identical to their constitutive neural parts (Pereboom, 2002, p. 503). For example, John's belief that chairs are sturdy remains the same for many years in John, even through the potential replacement of some of the neurons that make up John's belief that chairs are sturdy.

---

<sup>18</sup> Pereboom points out, but does not reply to, the objection that temporal and modal properties do not disprove the identity, since Theseus' ship, here and now, may still be identical with the current parts of Theseus' ship, regardless of what happens later, or what could have happened differently (Pereboom, 2002, p. 503). Harbecke also thinks this objection prevents the constitutionalist from pointing to distinct temporal properties to prevent the identity, though he thinks distinct modal differences render the identity impossible (Harbecke, 2008, p. 169). Presumably, however, if there is an identity then there is one thing. And, it is not possible for one thing to both cease to be itself while simultaneously continuing to be itself. If Theseus' ship is identical to its parts, however, this is the scenario that arises. For example, Theseus' Ship is parts WXY at time t, and Theseus' Ship is parts WXZ at time t+1. In this case Theseus' Ship at t remains Theseus' Ship at t+1. However, WXY (which is Theseus' ship) is now WXZ at time t+1, which is not WXY.

The constitution relation between part and whole, though not amounting to identity, is very tight. As Jens Harbecke summarizes, “if object  $x$  composes object  $y$ , then  $x$  is linked to  $y$  in a very intimate way. No thing can be discrete from its parts ... [there is] merely a mild distinctness” (Harbecke, 2008, p. 168). So tight is the constitution relation that it is able to prevent the mental token from being excluded by its causally sufficient parts. How is this possible? First of all, let us note that Pereboom agrees that physical token  $P$  is sufficient for the effect  $P^*$  in the above Figure 3, and that  $P^*$  is sufficient for the occurrence of  $M^*$  (Pereboom, 2002, p. 510). According to the causal exclusion problem diagrammed above, there should now be some pressure to exclude  $M$  as a cause of  $P^*$  and as a cause of  $M^*$ . Pereboom, however, argues that there is no such exclusion pressure:

Just as Kim claims that no competition between explanations arises in the case of reduction and identity, I propose that no competition arises in the case of mere constitution either. For if the token of a higher-level causal power is currently wholly constituted by a complex of microphysical causal powers, there are two sets of causal powers at work which are constituted from precisely the same stuff (supposing that the most basic microphysical entities are constituted of themselves), and in this sense we might say that they coincide constitutionally (Pereboom, 2002, p. 505).

The argument here is a bit unclear, as Kim has noted (Kim, 2005, p. 61), but I want to parse it out. First of all, Pereboom is saying that in the same way that an identity secures physical causation and mental causation, so the constitution relation secures physical causation and mental causation. How does the identity secure mental causation? The absolute closeness of the mental and the physical implicit in the identity (indeed, there is only one thing) secures mental causation. In the same way, the almost absolute closeness of the constitution relation secures mental causation.

Notice that Pereboom also states that the two causes “coincide constitutionally” in virtue of the fact that they are constituted from the same stuff. This seems to mean that

the mental and physical causal power tokens share the same spatial parts. Thus object  $x$ , made of  $H$ , instantiates  $p_1$  of property  $P$ , while object  $y$ , made of  $H$ , instantiates  $p_2$  of property  $P$ . For example, a root of a tree shares certain spatial parts with the tree itself; namely, this root. This root has the causal power of being able to absorb water. However, the tree also has the causal power of being able to absorb water, and this is partially so in virtue of sharing the root as a spatial part. The tree's ability to absorb water is not identical to this root's ability to absorb water since the tree's ability to absorb water can outlast the life of this particular root. Therefore, the tree's causal power of absorption coincides with, but is not identical with, the root's power of absorption. Since these two causal powers originate in the same parts, however, there is no causal competition between them. As Alexander Rueger summarizes, "since parts and wholes do not stand in competition for causal sufficiency, the analogon of the problem of mental causation does not arise" (Rueger, 2004, p. 2; Rueger, 2006, p. 340-341; Clapp, 2001, p. 133).

With regard to this issue of mental causation, John's belief that chairs are sturdy is constituted by certain neural parts. These neural parts have the causal power of disposing John to sit when chairs are present. John's belief also has the causal power of disposing John to sit when chairs are present. The causal power of John's belief is not identical with the causal power of the neural parts since we can lose or replace some of the neural parts without changing the causal power of John's belief. However, since these two causal powers arise from the same stuff, we can dismiss any pressure to exclude the mental cause.

Let us see how this model fares against the two difficulties raised earlier; namely, the problem of causal exclusion, and the problem of providing a necessary and distinctive role for the mental token. First of all, it is not clear whether we can have a relation that is short of identity but is still tight enough to secure mental causation. As Alyssa Ney points out, the supervenience relation is tight as well, but this tightness does not ward off exclusion pressures (Ney, 2007, p. 489). Indeed, as some have argued, so long as there is

any distinction between the mental cause and the physical cause, we will be able to exclude the mental cause because the physical cause is still sufficient and the mental cause is distinct from this sufficient physical cause (Harbecke, 2008, p. 172). Obviously, though, this intuition is not universally shared.

For the sake of the argument, let us say that the constitution relation is tight enough to ward off exclusion pressures. What does this leave us with? Pereboom suggests that the physical token is sufficient for the occurrence of the effect. When Pereboom asks whether this leads to the exclusion of the mental token, he responds:

I do not believe so... suppose that  $M$  instantiates an irreducibly mental causal power, and that the activation of this irreducibly mental causal power brings about  $P^*$ . Accordingly,  $M$  will cause  $P^*$ , and  $M$ 's causing  $P^*$  will not be identical with  $P$ 's causing  $P^*$ , and thus there is a significant sense in which  $M$  is indispensable as a cause of  $P^*$  (Pereboom, 2002, p. 511).

Although the physical cause is sufficient for  $P^*$ ,  $M$  also causes  $P^*$ , so we can conclude that the mental cause is indispensable. Or, in other words, the physical cause is sufficient, but as it turns out the mental cause brings about the effect as well, so we cannot simply ignore this causal relation that does obtain. This move overcomes the causal exclusion problem by suggesting that the effect is determined to occur by a sufficient physical cause and another mental cause. This amounts to an acceptance of overdetermination (Baker, 2009, p. 114; Ney, 2007, p. 487). I will discuss the feasibility of overdetermination in the next chapter and so will return to this issue later.

How about our concern about autonomous mental causation? As with supervenience and emergentism, the constitution solution retains autonomy in the sense that the mental token is distinct from the physical token. But does it posit a distinct role for the mental token? In the sense that the mental token is a distinct cause, the mental token does play a distinct role. However, this mental token does not play a necessary, distinct role. If the mental cause did not occur, the effect would still occur due to the sufficiency of the physical cause. Indeed, Pereboom states that in contrast to

emergentism, where the higher level properties can produce novel effects, constitution does not give mental states the “power to produce deviations from the ordinary microphysical laws” (Pereboom 2002, 510-11). If the physical cause is sufficient, and therefore can generate the effect without help from the mental cause, then we cannot say that the mental cause plays a necessary role. This being the case, the constitutionalist model does not give the mental a necessary distinct role, so autonomous mental causation is not preserved.

Laurie Paul offers a slightly modified version of the constitution model. Where Pereboom posits two objects made of the same stuff that have different instantiations ( $p_1$  and  $p_2$ ) of the same property  $P$ , Paul argues that there are two objects made of the same stuff which share the same instantiation  $p_1$  of the same property  $P$ . As Paul explains:

In the cases of constitutive overdetermination we've considered, the causally important or relevant property instances are shared, we can see why causal responsibility is shared, not overdetermined. When I, my constitutive sum of cells, constitutive sum of molecules, etc. hit the tennis ball, a *shared* part causes the tennis ball to bounce off my racquet at 100 mi/hr (Paul, 2007, p. 24).

If Pereboom’s argument is vulnerable to the causal exclusion argument because the sufficiency of the physical cause screens off the marginally distinct mental cause, Paul dodges these exclusion pressures by arguing that the mental cause is the physical cause. However, Paul still fails to provide us with a necessary, distinct role for the mental. Although the mental state is distinct from the neurons that constitute it, the mental state and the neurons share the same neurons which instantiate the same causal property, so there is no distinct mental causation (Ney, 2007, p. 492).

A number of philosophers endorse a slightly different version of this constitution model, which yields a different response to the problem of causal exclusion (Yablo, 1992; Shoemaker, 2002; Wilson 1999; Wilson, 2009). These philosophers still maintain that the causal exclusion problem is resolved by looking to the part-whole relation.

However, where Pereboom conceives the part-whole relation as one of spatial coincidence, these writers conceive of the part-whole relation as involving subsets.

Stephen Yablo (1992) was the first to suggest something like this strategy. He marks a distinction between determinables and determinates, where the property ‘red’ stands in a determinable relation to the property ‘scarlet’. Scarlet, on the contrary, is a determinate of red. Now, imagine that a bird is taught to peck at red things. A red triangle is presented to the bird, and it pecks. What is it about the triangle that makes the bird peck? The redness of the triangle. However, in this case the triangle is a particular shade of red, namely, scarlet. If the scarletness of the triangle is causally sufficient for the pecking, then should we exclude the redness of the triangle (Yablo, 1992, p. 257)? Yablo says no because “determinates do not contend with their determinable for causal influence” (Yablo, 1992, p. 259. See also Shoemaker, 2001, p. 81; Wilson, 2009, p. 153). Now, due to multiple realizability, it is plausible that mental properties are determinables, whereas one particular physical property is a determinate of this mental property on one occasion (Yablo, 1992, p. 254). If this is true, then it stands to reason that mental properties are not in competition with their determinate physical properties either.

Yablo, followed by Sydney Shoemaker (2001), uses this subset model to offer a nonreductive solution to the problem of causal exclusion. He does not identify the determinable with the particular determinate that appears on an occasion. However, he still insists that a determinable cannot be excluded from causal efficacy at the hands of its determinate due to the tightness of the relation between them. Some have argued that Yablo’s model, despite his protestations, remains susceptible to the problem of causal exclusion:

Stephen Yablo has held that even though determinable events are not identical to their determinate correlates, these events do not causally exclude one another. Yet, he provides no explanation for this non-exclusion. As such, it appears to be an *ad hoc* exception to exclusion principles (Funkhouser, 2006, p. 21).

In order to sort out whether Yablo's model falls to the causal exclusion problem, let us consider the bird one more time. Why does the bird peck at the triangle? There seem to be three available replies (1) the pigeon pecks in virtue the triangle's scarletness being scarlet; (2) the pigeon pecks in virtue of the fact that the triangle's scarletness being red; or (3), the pigeon pecks in virtue of the fact that the triangle's scarletness being both red and scarlet.<sup>19</sup>

Consider the first option. Let us grant, with Yablo, that there is an important sense in which the triangle's scarletness is red. In fact, since the bird is only trained to peck at red things, let us say that it is in virtue of the redness of the triangle's scarletness that the bird pecked. Is the redness of the triangle's scarletness sufficient to cause the bird to peck? If not, then we can include other properties of the triangle's scarletness. But this will not help us in the mental causation debate since those involved in the mental causation debate assume that physical properties are sufficient for the given effect. Thus, we will have to say yes, the redness of the triangle's scarlet is sufficient to cause the bird to peck. In this case, we can exclude the scarletness of the triangle's scarlet, since we already have a sufficient cause.

Similarly, if we say that the bird pecks because of the scarletness of the triangle's scarlet, then we will be able to exclude the redness of the triangle's scarlet. Seemingly, Yablo has in mind the third option, which states that both the scarletness of the triangle's scarlet and the redness of the triangle's scarlet are causally relevant. Again, in order for the parallel to the mental causation debate to work, we need to assume that the scarletness of the triangle's scarlet is causally sufficient. Thus, we have a sufficient cause, and another cause on top of this. This amounts to overdetermination. As I indicated earlier, I will explore, and ultimately reject such appeals to overdetermination in the next chapter.

---

<sup>19</sup> This way of framing our options obviously relies on the plausibility of the property instance of scarlet having the property instance of redness. This assumption that property instances can have property instances is endorsed by those who use the subset strategy, and it will be discussed in Section 5.3.

For now, it is also important to note that Yablo's model does not provide a necessary, distinct role for the mental cause. Although Yablo maintains that the mental cause is distinct from the physical cause, he also thinks the physical cause is sufficient in itself, so the mental cause is not really necessary. For these two reasons, among others,<sup>20</sup> a resolution to the problem of causal exclusion rooted in the determinable/determinate relation is as problematic as the previously mentioned constitutionalist solutions.

Jessica Wilson (1999; 2009) offers a slightly modified version of Yablo's model. She says that scarletness has more causal powers than redness. For example, another bird named Alice may have been trained to peck at scarlet things, but not other red things (Wilson, 1999, p. 48; see also Shoemaker, 2001, p. 78). In this case, Alice would peck at the scarlet triangle as well, but not because it is red, but rather because it is scarlet. So, the scarlet "evidently has *more* causal powers than the property of being red" (Wilson, 1999, p. 48). This being the case, she concludes that redness is a subset, or a part, of scarletness.

Jessica Wilson resolves the causal exclusion problem by arguing that the properties of scarlet and red are different, since they have different causal powers. However, the determinate instance of scarlet that causes this bird to peck is identical to the determinable instance of red:

The reason why 'determinables and their determinates are not causal rivals' – is because it is plausible, in the case of determinables and determinates, that each causal power of the determinable is identical with a causal power of its determinate (Wilson, 1999, p. 48; See also Wilson, 2009, p. 153; MacDonald and MacDonald, 1986; Ehring, 1996).

---

<sup>20</sup> Yablo's model faces certain other difficulties as well. For example, a number of philosophers (Ehring, 1996; Funkhouser, 2006; Walter, 2007) suggest that the relation between mental and physical properties cannot be understood as a relation of determinable/determinate. Wilson (2009) has offered up a response to this concern, but Wilson's reply has been met by a reply from Haug (forthcoming). Beyond this concern, others have suggested that determinable properties do not exist, since we ought to accept a relatively sparse ontology (Gillett and Rives, 2005; Crane, 2008).

Construing scarlet as a set and red as a subset of scarlet leaves an overlap where there will be instances that are both red and scarlet. Rather than arguing that these red instances are distinct from these scarlet instances, she posits an identity between the two instances. Thus, we cannot say that the scarletness of the triangle caused the bird to peck, so the redness is excluded, for the scarletness of this triangle is the redness.

Although Wilson's model may have more success in dodging causal exclusion pressures, it does not give us autonomous mental causation. Consider Shoemaker's reason for not identifying the instance of one set with the instance of a subset:

If what is in fact an instance of a mental property causes something, but does so in virtue of being an instance of a physical property rather than in virtue of being an instance of that mental property, then the causal efficacy of the mental does not seem to have been adequately vindicated (Shoemaker, 2001, p. 80).

If a determinable mental property instance is a specific determinate physical property instance, then this mental property instance has causal power in virtue of being a physical property instance. This is not mental-as-strictly-mental causation, nor can the mental property instance play a distinct role, for it is not distinct from the physical instance.

To summarize, I considered three of the more popular methods of responding to the problem of causal exclusion. The supervenience model, emergentism and the constitutional model all fail. This is not to say that there are no other responses; rather these three models were chosen because of the method found in their failures. All of the models considered here have at least one of the following three consequences: (1) they posit a distinct mental cause beyond a sufficient physical cause, and as such the mental cause can be excluded; (2) they posit an identity of the mental cause with the sufficient physical cause, and as such fail to provide a necessary distinct role for the mental as required by our preference for autonomous mental causation; and/or (3), they lead to overdetermination, a consequence I will reject in the next chapter. Clearly, what we need is a mental state which plays a necessary distinct role, but which simultaneously is not excluded by the sufficient physical cause.

## **Chapter 4 - Causal Exclusion and Overdetermination**

To this point, I have argued that most nonreductive physicalists adhere to four individually plausible, but collectively problematic principles. Attempts at reconciling these principles have not been successful. While considering some of these attempts, however, we were introduced to the possibility of resolving this tension by simply endorsing overdetermination. Indeed, as we have seen, if overdetermination is accepted, then physical causal completeness, mental causation and irreducibility can all be preserved. An increasingly large number of philosophers suggest this approach (Sider, 2003; Bennett, 2003; Melnyk, 2003; Pereboom, 2002; Loewer, 2002). In this chapter I explore the advantages and disadvantages of relying on overdetermination to resolve the problem of mental causation.

This chapter is divided into three sections. Section 4.1 briefly revisits the version of overdetermination we saw in chapter two, which is called “independent overdetermination”, and its accompanying problems. In Section 4.2, I introduce the nuanced version of overdetermination briefly discussed in Section 3.5, which is called “dependent overdetermination”. Dependent overdetermination successfully eludes the problems that befall independent overdetermination, and as such breathes fresh life into the overdetermination strategy. Section 4.3 draws out some of the consequences of the dependent overdetermination position.

### **4.1 – Problems with Independent Overdetermination**

In Section 2.4 I considered two initial problems with overdetermination. Here I explore these issues in greater detail. I want to first of all get clear on a definition of overdetermination. Consider the example from section 2.4: a criminal is sentenced to death by the firing squad. When his day comes, two bullets pierce the criminal’s heart at the same time. We say that the criminal’s death is overdetermined. Why do we say this?

There are two independent and individually sufficient causes for this event.<sup>21</sup> How do we know this? Kim typically uses a counterfactual analysis to decipher whether or not an event has more than a single sufficient cause (Kim, 2005, p. 46-49; Kim, 1998, p. 44-45). Would the death still occur if only bullet *A* fired? Our answer is yes, so this is a sufficient cause for the death. Would the death still occur if only bullet *B* fired? Our answer is also yes, so this is a sufficient cause as well.

Importantly, if the firing of bullet *A* is a sufficient cause for the death, then the firing of bullet *B* is not a necessary cause. Similarly, if the firing of bullet *B* is a sufficient cause, then the firing of bullet *A* is not a necessary cause either. We see this demonstrated in the counterfactual scenario where bullet *A* does not fire, but the event of the death still occurs, thereby proving that bullet *A* is unnecessary. We also see this when considering the following typical definition of sufficient causation: “a sufficient cause is any cause that will bring about the event alone and by itself” (Campbell, 2004, p. 153; See also Marras, 2007, p. 319). If a cause is sufficient, *ceteris paribus*, it guarantees that the effect will occur all by itself, so no other causes are necessary for the effect to occur.<sup>22</sup> With this in mind, we arrive at a general characterization of overdetermination:

---

<sup>21</sup> The firing squad example is commonly used in the literature as an example of what is called independent overdetermination here. There is, however, a sense in which the two bullets are not independent, as the firing of the two bullets both have their origin in the same order issued by the same commander. We could avoid this problem by modifying the example to read that two random murderers decide to shoot the same person at the same time. .

<sup>22</sup> Some prefer to say that *x* is a sufficient cause of *y* if the presence of *x* necessarily implies the presence of *y*. This definition does not work. As we will see below a shadow or an epiphenomenal mental event may be present and necessarily imply the presence of effect *y*, but we would not say these are sufficient causes of this effect. Perhaps these are not causes, however, so they do not claim to be sufficient causes, and so they do not disprove this definition. We could alter our definition to rule out these cases as follows: *x* is a sufficient cause of *y* if the presence of the cause *x* necessarily implies the presence of *y*. This does not work either, for it is possible that the mental cause *x* necessarily implies the presence of *y* because the mental cause *x* is determined by the sufficient physical cause *p* which is the actual sufficient physical cause for the effect.

Independent Overdetermination: cause  $P$  is sufficient for effect  $e$ , thereby rendering cause  $M$  unnecessary, but cause  $M$  occurs anyway. At the same time, cause  $M$  is sufficient for effect  $e$ , thereby rendering cause  $P$  unnecessary, but cause  $P$  occurs anyway.

Perhaps mental causation operates in much the same way as the firing squad example. A physical cause is sufficient to bring about an effect, but a distinct mental cause also generates the event. Moreover, the mental cause is sufficient to bring about an effect, but in this case the distinct physical cause brings about the event as well.

We have seen the two reasons to think this will not work. First of all, this solution implies the existence of massive amounts of overdetermination, which is problematic. Secondly, it leads to the conclusion that neither mental causation nor physical causation is a necessary condition for the functioning of our daily lives. For these reasons, the consensus opinion is that we cannot solve the problem of mental causation by endorsing this model of independent overdetermination.

#### 4.2 – The Dependent Overdetermination Solution

Independent overdetermination solves the problem of mental causation by positing two individually sufficient causes for a given effect. This position led to massive amounts of overdetermination and the mental and physical causes being individually unnecessary for a given effect. A number of nonreductive physicalists, however, have crafted a more nuanced version of the overdetermination solution. This nuanced version overcomes both of the aforementioned objections, and as a result breathes fresh life into the overdetermination option.

The type of overdetermination we just rejected had the following definition: cause  $A$  is sufficient for effect  $e$ , thereby rendering cause  $B$  unnecessary, but cause  $B$  occurs anyway. At the same time, cause  $B$  is sufficient for effect  $e$ , thereby rendering cause  $A$  unnecessary, but cause  $A$  occurs anyway. There is reason to think that this is not the type of situation the nonreductive physicalist has in mind when considering the

problem of mental causation. One of the reasons is that nonreductive physicalists think mental events depend on, and are determined by, physical events. Consider Barry Loewer's analysis of what happens when two assassins both shoot the same victim at the same time:

In the two assassins case the two causes are metaphysically (and nomologically) independent. In the latter case  $M$  depends on  $N$  since it is metaphysically (or physically) entailed by it. Cases of the first sort of overdetermination are indeed rare but that provides no reason to think that cases of the second sort are unusual. The reason that we think that the first kind of overdetermination is generally not plausible is that when we posit two independent causes  $C1$  and  $C2$  for  $E$  we raise the question of why both  $C1$  and  $C2$  and why the events that occurred led to the same effect. But in the case of [Nonreductive Physicalism] we don't face these questions since  $N$  and  $M$  are not independent (Loewer, 2002, p. 657-658; See also Segal, 2009, p. 83; Walter, 2008, p. 678; Thomas, 2008; Bontly, 2005; Melnyk, 2003, p. 168; Crisp and Warfield, 2001, p. 135).

On the original model, there were two independent causes for the same event. But in the case of mental causation, the mental depends on the physical and is entailed by it. Other nonreductive physicalists have taken the same position due to their acceptance of a strong supervenience relation which implies that supervening mental events are dependent on their subvening physical bases, while these subvening bases necessitate the occurrence of the supervening mental events (Walter, 2008, p. 678; Marras and Yli-Vakkuri, 2008, p. 125; Kallestrup, 2006, p. 473; Macdonald and Macdonald, 2006, p. 566). Because of the supervenience relation between mental and physical events, "if  $P$  occurs,  $M$  will necessarily also occur" (Kallestrup, 2006, p. 473). And, because of the dependence of the mental on the physical, which is implicit in the supervenience relation, if some  $M$  occurs, some  $P$  will necessarily occur as well. In Section 3.5 we saw that those who endorse a constitution view, or a determinable/determinate view, also propose a similar dependence of the mental on the physical.

This is clearly a different sort of overdetermination than our original case. For, according to our original definition, cause  $P$  was sufficient so cause  $M$  is unnecessary. Within the mental causation debate, however, cause  $P$  is sufficient, but this physical cause necessarily guarantees that event  $M$  will occur as well. Similarly, on the original model, cause  $M$  was sufficient, so cause  $P$  was unnecessary. Within the mental causation debate, however, causal completeness and the asymmetric dependence of the mental on the physical guarantees that whenever some cause  $M$  is present, some underlying cause  $P$  is necessary. With these considerations in mind, here is a summary of the version of overdetermination available to the nonreductive physicalist in the debate on mental causation:

Dependent Overdetermination: some cause  $P$  is necessary and sufficient for effect  $e$ , but this cause  $P$  renders it necessary that cause  $M$  occurs, so this cause  $M$  is necessary for this effect  $e$  as well.

For his own part, Jaegwon Kim acknowledges the unique status of this version of overdetermination:

In standard cases of overdetermination, like two bullets hitting the victim's heart at the same time ... each overdetermining cause plays a distinct and distinctive causal role. The usual notion of overdetermination involves two or more separate and independent causal chains intersecting at a common effect. Because of supervenience, however, that is not the kind of situation we have here. In this sense, this is not a case of genuine causal overdetermination (Kim, 2005, p. 48).

See also Walter, 2008, p. 678).

Our standard case of overdetermination suggested that the two causes were independent, rendering them individually sufficient, and individually unnecessary. This new form of overdetermination suggests the mental cause depends on the physical cause, rendering the physical cause necessary and sufficient, while the supervening mental cause will necessarily occur as well.

Considering the drastic differences in types of overdetermination, it is helpful to separate them by invoking some new terminology. As hinted at in the previous passage, Kim distinguishes between genuine cases of overdetermination, and (presumably) non-genuine cases of overdetermination. Genuine cases of overdetermination are cases where two independently sufficient causes both bring about one effect, thereby rendering the other cause unnecessary. Non-genuine cases of overdetermination, on the other hand, are cases where two causes bring about one effect, but one cause depends upon, or is entailed by, the other, which makes the occurrence of both causes necessary. Kim's terminology is not followed consistently in the literature.<sup>23</sup> I will use the following vocabulary to highlight the important distinction between the two: independent overdetermination is the sort of overdetermination where two independent causes overdetermine an effect. By contrast, dependent overdetermination is the sort of overdetermination where two causes, one dependent on the other, the other necessitated by the other, overdetermine an effect.

This new model of dependent overdetermination is able to overcome the difficulties that undermined the previous model. The first problem was that independent overdetermination implied massive amounts of overdetermination, which was problematic because of the degree of coincidence involved. On the model of dependent overdetermination, there is still a massive amount of overdetermination, but this is no longer problematic. There is no problem because the appearance of the second overdetermining cause is no longer coincidental. Since the second cause is dependent on the first, it is neither coincidental nor far-fetched that the two causes would appear at the same time (Walter, 2008, p. 678; Raymont, 2003, p. 232; Funkhouser, 2002, p. 338; Block, 1990, p. 159). Imagine we reconstruct our initial firing squad example in the

---

<sup>23</sup> Sven Walter, for example, contrasts "genuine overdetermination" with what he calls "systematic overdetermination" (Walter, 2008, p. 678). Alyssa Ney, on the other hand, talks of "redundant overdetermination" versus "non-redundant overdetermination" (Ney 2007, p. 488). Ney then points to those who speak of "pejorative" versus "nonpejorative overdetermination" as well (Ney, 2007, p. 488). Finally, Jens Harbecke speaks of "exotic overdetermination" versus "pervasive overdetermination" (Harbecke, 2008, p. 28).

following way: a murderer is sentenced to death by fire. A fire is lit and the toxic smoke kills the man along with the searing heat. In this case, the smoke and the flame both kill the man together. But, it is not a large surprise or coincidence to see the death occur because of the smoke and flame this time. The smoke is dependent on the flame, and this intimate relationship makes their co-occurrence very much expected.

As for the second objection, we rejected independent overdetermination because neither the mental nor the physical cause was necessary. On our new model of dependent overdetermination, the mental cause necessarily accompanies the physical cause, so it is necessary for the effect. On the other hand, some physical cause is necessary due to the constraints of causal completeness, and because it necessarily generates the mental cause. Both causes being necessary, we are no longer faced with the worry that the mental could still cause the effect without the physical, or vice versa.

Since dependent overdetermination is benign and circumvents the aforementioned difficulties, some think we can successfully press it in to service in the mental causation debate. Sven Walter, for example, states the case:

However, the kind of overdetermination nonreductive physicalists typically have in mind is not like typical cases of genuine overdetermination at all, because the causes in question are not independent. First, the co-occurrence of the two causes is systematic and thus not coincidental, because supervenience ensures that the presence of the physical cause necessitates the presence of the alleged mental cause. Second, neither of the co-occurring causes is dispensable in the same sense as, say, the two assassins' shots are dispensable as causes of the victim's death: given supervenience,  $M$  will be present as long as  $P$  is present ... what prevents nonreductive physicalists from treating mental-to-physical causation as a case of systematic overdetermination? (Walter, 2008, p. 678; See also Kroedel, 2008, p. 126; Crane, 1995; Mellor, 1995, p. 103)

According to Walter, nothing prevents the nonreductive physicalist from treating mental causation as a matter of dependent overdetermination. When someone objects that the

acceptance of overdetermination implies an unusual world, we can respond by saying that in fact the overdetermination is neither surprising nor difficult to imagine. When someone objects that the physical cause makes the mental unnecessary, we can respond by saying the dependency relation between  $M$  and  $P$  renders the mental cause necessary after all.

#### 4.3 – Problems with Dependent Overdetermination

The key move that allows the dependent overdeterminist to dodge the problems with independent overdetermination is to endorse the necessity of the mental and the physical cause for this given effect. If the mental cause does not necessarily occur when the physical cause occurs, then the overdetermination will look coincidental once again. And, if the mental event does not necessarily cause the given effect, then we can argue that the mental cause is unnecessary for the effect in virtue of the sufficiency of the physical cause for the effect. Similarly, if the physical cause does not necessitate the mental cause and the physical effect, then we have lost causal completeness, and our notion of supervenience. Unfortunately, the necessity of the physical cause and the necessity of the mental cause also create several problems for dependent overdetermination.

Consider the difficulties that occur when we argue that the mental cause is necessary first. Since the mental cause is necessary whenever the physical cause occurs, there may be no possible worlds where the physical cause occurs without the mental cause. There are, after all, a number of dependent overdeterminists who suggest that the mental cause is metaphysically necessary whenever the physical cause occurs (Ezquerro and Vicente, 2000, p. 16; Kallestrup, 2006, p. 459-460; Bennett, 2003, p. 479-480; Loewer, 2002, p. 658). The absence of worlds where the physical cause occurs without the mental cause renders it difficult to prove the dependent overdetermination thesis true. For example, if there are no worlds where the physical cause acts in isolation and the effect still occurs, then we cannot point to a given world to demonstrate that the physical cause alone is sufficient. But the causal sufficiency of the physical cause is an important

component of dependent overdetermination, for it secures physical causal completeness.

Jaegwon Kim articulates a second worry:

In considering the claim that  $M$  and  $P$  are each a sufficient cause of  $P^*$ , however, we need to be able to consider a possible situation in which  $M$  occurs without  $P$  and evaluate the claim that in this possible situation  $P^*$  nonetheless follows. If such is not a possible situation ... what significance can we attach to the claim that  $P$  and  $M$  are each an overdetermining sufficient cause of  $P^*$ , that in addition to  $P$ ,  $M$  also is a sufficient cause of  $P^*$  (Kim, 2005, 46).

Here Kim argues that if we cannot locate a world where the mental cause occurs without the physical cause, we cannot be certain that the mental cause is sufficient to bring about the effect. Presumably, the causal sufficiency of the mental cause is important for the dependent overdeterminist.

We may be able to get around this problem by merely endorsing the nomological necessity of the dependency relation. Indeed, a number of people do not insist that the physical cause metaphysically necessitates the mental cause, but rather only insist that the physical cause nomologically necessitates the mental cause (Block, 2003, p. 136; Crisp and Warfield, 2001, p. 314). If the physical cause only nomologically necessitates the mental cause, then we will be able to find worlds with different laws that allow the physical event occur without the mental event. In some of these worlds causal completeness will still hold, so the physical cause will have to occur in order to generate the effect. However, the mental cause will not occur there, so we have a demonstration that the physical cause is sufficient on its own.

There are a number of problems with this response. First of all, in the same way that this distant world proves the physical is individually sufficient it also proves that the mental is unnecessary. After all, we now have a world where the physical event occurs in isolation but the effect still occurs, so we do not really need the mental cause. Kim makes a similar argument when he uses the mere nomological necessity of the mental cause to locate a world where the mental event occurs without an accompanying physical

event. Under our laws the physical cause necessitates a mental cause, but in a nomologically distinct world where supervenience no longer holds, the mental cause occurs without a subvening physical cause. Since the mental cause brings about the physical effect without there being a physical cause of the physical effect, causal completeness fails in this world (Kim, 2005, p. 49-50). According to Kim, this discovery brings trans-world implications with it: if there is a possible world where the mental exists without the physical, then the mental does not really need the physical, and that means causal completeness is in danger.

There is a way to overcome Kim's argument here. Sven Walter responds to Kim's concern by arguing that there is no reason to be worried about the things that transpire in such modally distant worlds (Walter, 2008, p. 680). This world operates under a set of laws that necessitates the mental event whenever the physical event occurs, so we cannot apply the discoveries found in nomologically distinct worlds to our debate here. In order for this response to successfully refute Kim's argument, however, we will need to agree that what transpires in nomologically distinct worlds has no repercussions for the discussion we are having in this world, under our laws. But if we grant this, then we cannot appeal to a nomologically distinct world where the physical cause is sufficient to demonstrate that the physical cause is sufficient here. Perhaps the physical cause is sufficient in that world, under those laws of nature, but that does not prove that the physical cause is sufficient here, under our laws. Without recourse to a world where the physical event causes the effect in isolation, it seems like there is no way to be certain that the physical cause is individually sufficient.

This argument does not prove that the physical cause is individually insufficient, however, it merely demonstrates that we cannot prove that the physical cause is individually sufficient. In what follows I construct an argument that leads to the stronger conclusion that the physical cause is individually insufficient. The argument proceeds as follows: if a physical cause of a given effect necessitates a mental cause of this effect, then the physical cause is not individually sufficient for this effect. The physical cause of

this given effect does necessitate a mental cause of this effect. Therefore, the physical cause is not individually sufficient for this effect.

I want to begin by motivating and clarifying the second premise, that the physical cause of a given effect necessitates a mental cause for this effect. Three reasons have been given to come to this conclusion so far. First, to avoid massive coincidence, the physical cause must necessitate the occurrence of the mental event. Second, to render the mental cause necessary (and hence not excludable) for a given effect, the physical cause must necessitate a mental cause of this effect. Finally, the strength of the dependency relation between the two causes requires that the mental cause necessarily occurs whenever the physical cause occurs.

To be clear, the second premise does not state that a mental cause is always necessary for a given effect. Rather, it makes the more modest claim that whenever a physical cause of a given effect necessitates a mental cause of this effect, the mental cause is necessary.<sup>24</sup> The second premise, therefore, only makes the mundane claim that

---

<sup>24</sup> This is important to note because there are several scenarios where the mental cause in general turns out to be unnecessary. For example, we can conceive of a world where there is another sufficient physical cause ( $P_2$ ) of the effect, but this sufficient physical cause does not necessitate the appearance of the mental cause. In this world, the effect occurs without the mental cause, indicating that the mental cause is not necessary here. Alternatively, we can conceive of a world where our original physical cause ( $P_1$ ) does not necessitate the appearance of the mental cause, although it does still necessitate the effect. Here again, the effect occurs, but the mental cause does not, so the mental cause is unnecessary. On both scenarios the mental cause is not necessary, so if we were trying to make the stronger claim that a mental cause is always necessary, it may be proven false. There are a number of available replies to this argument. First of all, we have to appeal to very remote worlds in order to prove that the mental event is not necessary. In both worlds there are pertinent laws which are different from the laws present in our world. In our world, the law ‘whenever  $P$  occurs,  $e$  occurs’ is in effect, whereas in the first alternate world the law ‘whenever  $P_2$  occurs,  $e$  occurs’ is in effect while the law ‘whenever  $P$  occurs,  $e$  occurs’ cannot be in effect. Moreover, in our world the law ‘whenever  $P$  occurs,  $M$  occurs’ is in effect, but in the second world this law is not in effect. As we’ve seen above, it is not clear that what transpires in worlds that are so drastically different in such pertinent ways can have any bearing on the debate that is happening in our world, under our laws. For our purposes here, however, we can grant the existence of such worlds,

whenever a physical cause of a given effect necessitates a mental cause, in those cases the mental cause is necessary. This seems like virtually an analytic truth. By definition the mental cause is necessary whenever we stipulate that the mental cause is necessary. And this is, of course, entirely the point. The dependent overdeterminist ought to accept the necessity of the mental cause in these situations, for it is simply a re-articulation of their position.<sup>25</sup>

Granting that whenever a physical cause necessitates a mental cause, that mental cause is necessary, what reason is there to think this renders the physical cause individually insufficient? This follows from the definition of sufficiency and necessity that was outlined earlier. If the mental cause is necessary for the effect to occur, then by very definition, the physical cause is individually insufficient for making the effect occur.

---

and we could even conclude that the results found there are pertinent to the debate at hand. The results of these arguments simply state that a mental cause is not always necessary for a given effect, while we do not need this stronger claim here.

<sup>25</sup> Note that we cannot appeal to those cases where the mental cause is not necessary to help us out of the problem articulated below. The reason for this is that those of us in the mental causation debate are only concerned with cases where the mental cause is necessary. For example, we are not concerned with the aforementioned cases where in some distant world, which may not even be conceivable, and with different laws than ours, the mental cause is not necessary. We are not even concerned with more commonplace examples where the mental cause is not necessary. For example, consider the case where neural stimulation  $P_1$  is normally the cause of John's hand raising. Neural stimulation  $P_1$  necessitates the mental state  $M_1$ , which is John's desire to raise his hand. On this occasion, however, John is sleeping and is strapped to a contraption which raises his hand whenever a lever is pulled. Sue pulls the lever, so the effect of John's hand raising does occur, but the mental cause does not occur, which means that the mental cause is not necessary. This is trivially true, but it is of no concern to us because the relevant physical cause  $P_1$  is not even necessary in this situation. There is no possibility of mental causation in this situation, nor is there the possibility of a physical cause excluding a supervening mental cause. This is simply not an instance of the tension between mental and physical causes which interest those of us in the mental causation debate. Rather, we are interested in those cases where a physical cause of a given effect necessitates a mental cause of this effect. And, in all of these scenarios, the mental cause is necessary. For this reason, we can conclude that the second premise in our deduction is true.

That is, if one says that they certainly need the mental cause, then they cannot also say that the physical cause does not need anything else (i.e., that they certainly do not need the mental cause). If the physical cause is individually sufficient for the effect, then no other cause needs to occur as well. As it stands, however, the mental cause does need to occur as well. The physical cause, therefore, is not individually sufficient. Consider Ausonio Marras' analysis:

But then, if neither  $M$  nor  $P$  can *occur* in the absence of the other, neither can *cause*  $P^*$  in the absence of the other:  $P$  is causally sufficient for  $P^*$  if and only if  $M$  is. Thus neither  $M$  nor  $P$  qualifies as an independently sufficient cause of  $P^*$  in the required exclusionary sense (Marras, 2007, p. 319).

If, as the dependent overdetermination thesis states, the physical cause necessitates the occurrence of the mental cause, then this renders the physical cause insufficient on its own. Thus, dependent overdetermination implies that the physical cause is not individually sufficient for the effect, but causal completeness states that the physical cause must be individually sufficient for the effect. Dependent overdetermination, therefore, appears to violate causal completeness.

There are a number of objections that can be raised against this line of reasoning. Perhaps the most obvious objection is that a mental event may be necessarily present, but this does not mean that the underlying physical cause is insufficient for the effect.<sup>26</sup>

---

<sup>26</sup> Importantly, this objection poses an additional problem for dependent overdetermination as well. Recall that dependent overdetermination avoids the pitfalls of independent overdetermination by insisting that the physical event necessitates the mental event. However, the necessary presence of the mental event does not guarantee that the mental event is necessarily a cause of the effect. On epiphenomenalism, for example, the mental event is necessary, but does not make a causal contribution to the given effect. Therefore, dependent overdetermination does not necessarily secure the intuitive assumption that the mental plays a role in making the effect occur, it merely secures the conclusion that the mental event must be present. In order to secure the necessary causal contribution, dependent overdetermination would have to insist that the mental event is a cause of the effect. Unfortunately this additional move, as we will see below, is what causes the physical event to be causally insufficient on its own, for there is now another cause that is necessary.

Consider the following example: a metal heats to a certain temperature, which causes the piece of newspaper that it is touching to start on fire. This metal, however, necessarily changes colour when it is heated to this temperature. Does the fact that the metal necessarily changes colour somehow demonstrate that the heat of the metal is insufficient to cause the piece of paper to ignite on its own? Obviously not. In the same way, the mental event may necessarily appear, but the physical event can still be sufficient to cause the effect on its own. Surely the appearance of a mental event cannot detract from the sufficiency of the physical cause.<sup>27</sup>

This objection misfires. To see this, let us return to the distinction often drawn between causal relevance and causal efficacy (Sosa, 1984; Davidson, 1993; Marras, 1997; Jackson and Pettit, 1990). A mental event is causally relevant if a change to the mental event leads to a change in the effect. For example, the colour of the metal is causally relevant because if there had been no colour, then the paper would not have ignited. Causal relevance, as Kim argues, does not give us causal efficacy (Kim, 1993b, p. 23). The colour change can be causally relevant in the sense that without the colour change the fire would not have occurred. However, the colour change is not causally efficacious, because the heat of the metal did all of the work. Kim even argues that an epiphenomenalist can achieve the causal relevance of mental events, for without the mental event the physical event would not be there to do anything. But even though the mental event guarantees that a physical event is about to do some causal work, this does not change the fact that the physical event does all of the causal work itself.

With this distinction in mind, does the dependent overdeterminist want to achieve the causal relevance of mental events, or the causal efficacy of mental events?

---

<sup>27</sup> It is possible that the necessary appearance of a mental event renders the physical cause individually insufficient. This can be argued for as follows: if the physical cause is individually sufficient for an effect, then nothing else needs to occur. However, the appearance of a mental event (no matter if it is epiphenomenal) is necessary as well. In so far as the physical cause needs the mental event to appear, the physical cause is not sufficient on its own, for it also needs the mental event to appear. This is a much stronger argument than the one I am making here, and it is probably false.

Presumably, since the dependent overdeterminist wants to hold on to mental causation, they will opt for causal efficacy. Otherwise, they could simply endorse epiphenomenalism and acquiesce to the apparently inconsistent tetrad by rejecting mental causation in the first place. The only reason that the dependent overdeterminist goes to the lengths she does is to reconcile mental causation with the other three principles.

Since the dependent overdeterminist wants to achieve the causal efficacy of mental events, she cannot simply say that the mental event is necessarily present, but not causally active. Rather, she will have to say that the mental event is necessarily present, and is necessarily a cause. This ties in well with our definition of autonomous mental causation. I suggested that autonomous mental causation involves the mental playing a necessary, distinct role. It is not enough for the mental to be distinct, and to necessarily occur, but not play a role. Rather, the mental needs to be distinct and play a necessary role. This is where the analogy breaks down in the case of the metal and the colour. On this example the heat of the metal does all the work while the colour of the metal is epiphenomenal. Colours, after all, do not generally make papers burst into flame. This is not a satisfactory model for the dependent overdeterminist. The supervening event must also be a cause. The better analogy is the one involving the fire and the smoke. The fire necessarily produces the smoke, but the fire and the smoke are both causally efficacious. The death occurs because of both the fire and the smoke.

This leads to another objection, however. Isn't it the case that the fire and the smoke are both sufficient for causing a death, but the fire simply pre-empts the smoke from causing the death? If this is the analogy, then it seems like the physical cause is individually sufficient, despite the fact that another cause is present. This objection does not work either. According to dependent overdetermination, we not only need the mental event to be present, and potentially causal, but we also need the mental event to actually make a causal contribution to the effect in question. If the mental cause is present, and potentially causal, but does not actually contribute anything because it gets pre-empted by the physical cause, then the mental cause does not actually exercise any causal power,

so there is no mental causation. For this reason, the mental cause needs to be part of the cause for the effect.

To return to the fire and smoke analogy, if we want to argue that the smoke is actually a cause of the death on this occasion, then we need to say that the fire did not do all of the work itself, for the smoke had to do some of it. Perhaps on some other occasion, where the victim has a gas mask on, the fire could have done all of the work. Or perhaps, on some other occasion, where the victim is wearing a fire retardant suit with a hole in the nose area, the smoke could have done all of the work. But in our situation, the one where we have assumed that both the fire and the smoke make a causal contribution, neither of these could do all of the work by themselves because we have already said that they do not. In the same way, if we stipulate that the mental necessarily makes a causal contribution in our case, then we cannot also say that the physical cause does all of the work by itself. It does not, for we have already insisted that the mental cause does some work.

We have seen that it is not enough for the mental event to merely be present, but it must be causal as well. Nor is it enough for the mental event to be potentially causal, but to have not made its contribution, rather it must make a contribution as well. Similarly, the mental event must not simply make a causal contribution, but it must make a causal contribution to the same effect that the physical cause brings about. The mental causation problem only arises when there are two competing causes of the same event. For example, the problem only arises when we think a physical event causes John to eat spaghetti, while we also think that a mental event causes John to eat spaghetti. Given these clarifications of the necessity condition and the causal condition, it seems clear that under these circumstances we cannot say that the physical cause still does all the work itself. After all, if we insist that the mental cause does some work for the occurrence of

the effect, we cannot simultaneously say that the physical cause does all of the work for the occurrence of the effect.<sup>28</sup>

Having arrived at this conclusion we can close off the open line of reasoning that remains from Section 3.5. Namely, upon occasion the constitution reply turned to overdetermination to resolve the problem of causal exclusion. The closeness of the constitution relation may have enabled the mental to remain a cause despite the causal sufficiency of the physical cause. We now see, however, that as soon as we insist that a mental cause is a part of the cause of our effect the physical cause is no longer sufficient on its own. The constitutionalist appeal to overdetermination, therefore, leads to the view that the physical is not sufficient, which breaches causal completeness. For this reason, as well as the possible violation of the principle of exclusion and the failure to achieve autonomous mental causation, we can close the book on the constitution reply.

Returning to the topic at hand, perhaps we can object that the physical cause is sufficient since the physical cause is sufficient to generate the mental cause and the effect all by itself. In this sense, the only thing necessary for the effect to occur is the physical cause, since the physical cause will guarantee that the mental cause will appear as well, and that the effect will occur. This model bears certain similarities to the view that the first step in a causal chain is the sufficient cause for the third effect since the first step is sufficient for the second step which is in turn sufficient for the third step. Using this reasoning we could say that the initial singularity at the beginning of cosmic time is the

---

<sup>28</sup> This is not to say that the physical cause could not do all of the work itself in bringing about a similar though different effect. A fire is sufficient to cause a similar though slightly different death, and the physical cause is sufficient to bring about a similar though slightly different effect. This is to say, however, that the physical cause does not do all of the work in bringing about this effect, which is by definition caused by the physical cause and the mental cause. The fire is not sufficient to bring about this death which is by definition caused by burning and smoke inhalation. Similarly, the physical cause is not sufficient to bring about this effect which is by definition caused by the physical cause and the mental cause.

sufficient cause for our effect. While this may be true, we generally separate out these necessary background conditions in an attempt to determine the immediate sufficient cause of an effect. The immediate sufficient cause of our effect cannot be the physical cause alone, because the mental event is an immediate cause as well.

Having considered the difficulties that arise when the physical cause necessitates a mental cause, let us now briefly consider the difficulties that occur when the physical cause is necessary. On dependent overdetermination, the physical cause is necessary due to causal completeness and the fact that mental events are determined by physical events. The physical cause is necessary, so the mental cause is not individually sufficient, for many of the same reasons outlined above. Also, the mental cause could not bring about an effect without the occurrence of the physical cause, for the mental cause would not have occurred, so the mental cause is not sufficient on its own.

Perhaps it is acceptable to reject the individual sufficiency of the mental cause. After all, causal completeness and supervenience require that some physical cause needs to be present in our situation. However, token identity, type identity and Cartesian dualism all offer mental sufficiency. On the identity views, mental causes are the sufficient physical causes, so it is true that the desire to make chicken for supper is sufficient to make a person start to make chicken. On Cartesian dualism, mental causes have so much autonomy from physical causes that they can produce effects on their own, especially other mental effects. Thus, although nonreductive physicalism is not jeopardized by the failure of individual mental sufficiency, these competing theories have an advantage in this area.

Assuming that these difficulties arise whenever mental and physical causes are individually necessary, what options does this leave for the dependent overdeterminist? First, she can simply retract her claim that the mental cause is necessary for the occurrence of the effect. This would resolve the tension, but it comes at a price. Recall that the necessity of the mental cause arose in an effort to refute the suggestion that the physical cause could do all of the work itself, thereby excluding the mental cause.

Rejecting the necessity of the mental cause now would render us vulnerable again here. Similarly, she can retract her claim that the physical cause is necessary. But this implies the failure of causal completeness and the supervenience thesis.

The dependent overdeterminist can also accept a slight modification to causal completeness as a result of our discussion. Ausonio Marras is willing to consider this move (Marras, 2007, p. 319). He argues that this stance leads to a neat solution to the causal exclusion problem as well. Since causal exclusion says that there can be no more than a single sufficient cause for a given effect, and the physical cause is not sufficient in the absence of the mental cause, there is no pressure to exclude the mental cause. Rather, there is reason to include the mental cause, for once it is included we gain the required causal sufficiency. This position, along with this solution, requires us to reformulate both causal completeness and the definition of dependent overdetermination.

First of all, causal completeness would no longer state that all physical effects have a sufficient physical cause. Rather, causal completeness would now state that all physical effects have a sufficient physical cause, except for those physical causes that necessarily subserve mental causes, for in this case the physical cause together with the supervening mental cause will be a jointly sufficient cause for the given physical effect. As for dependent overdetermination, the following model more accurately reflects the situation:

Dependent Determination: some event  $P$  is necessary for this effect  $e$ , but this cause  $P$  renders it necessary that event  $M$  occurs, so this cause  $M$  is necessary for this effect  $e$  as well, and  $P$  together with  $M$  is jointly sufficient for this effect  $e$ .

As we have seen, dependent overdetermination leads to the conclusion that neither the physical nor the mental is sufficient in itself for a given effect, so we need no longer worry about overdetermination, nor do we need to use that label. Rather, it is better to use the term ‘determination’. Moreover, since both causes are necessarily present, and neither cause is individually sufficient in itself, it is best to continue to use the term ‘Dependent’, for it hints at the fact that both of these causes are necessary. If we are

willing to accept the modifications that this view brings, we can also receive the merits of this view with respect to preserving mental causation and irreducibility while overcoming the causal exclusion problem. However, since I have set out to reconcile without rejecting any of our four competing intuitions, I will not take this route.

There is a third possibility. Some of the reasoning that led to the dependent determination solution can be followed, but the position can then be modified in such a way as to not violate causal completeness. This can be accomplished as follows: the above reasoning suggests that one way to secure both the sufficiency of the physical cause and the necessity of the mental event is to argue that the mental event is only necessarily present, but is not causal. If the mental event is only necessarily present, but has no causal power, then the physical cause must do all of the causal work itself, so the physical cause is sufficient. The option was rejected because it led to epiphenomenalism and could not secure autonomous mental causation. The mental event is distinct and necessarily present, but it cannot play a role due to the requirement of physical causal completeness, so the mental event does not necessarily play a role.

But there is a similar, though nuanced, strategy available. Namely, the mental state<sup>29</sup> can be necessarily present (for all of the reasons listed above), and also necessarily play a non-causal determinative role. If the mental state plays a non-causal determinative

---

<sup>29</sup> As evidenced in certain passages quoted in chapter two (Kim, 1993, p. 106-107; Kim, 2007, p. 239), Kim seems to freely interchange the word “event” and the word “state”. For him, to call something a mental event or a mental state is largely to say the same thing. In this dissertation I will mark a slight distinction between physical events and mental states. By ‘physical event’ I mean the same thing that Kim means by an event. Namely a physical event is an occurrence or an instantiation of a property in a thing at a time. By ‘mental state’ I also mean an event, in the sense of an occurrence, or the instantiation of a property at a time. However, as will become clear, my position suggests that a mental event remains instantiated for the duration of time it takes for one relevant physical cause to bring about a second physical effect. Seen from this perspective, the mental event has the appearance of being a stable state of affairs, so I will refer to it as a mental state.

role, then we have in some sense overcome epiphenomenalism and achieved the goal of mental causation. But does this mean we are violating causal completeness? The answer here is no. The mental state plays a determinative, but non-causal role. If the mental state is not causal, the physical cause will once again be forced to do all of the causal work in bringing about the effect itself. This guarantees that the physical cause is sufficient, which secures causal completeness for us. This nuanced strategy will be considered in detail in the chapters that follow.

In summary, independent overdetermination fails because it implies that mental and physical causes are not necessary for the occurrence of the effect. Dependent overdetermination resolves this difficulty by rendering the physical cause necessary, and the mental cause necessary whenever the physical cause occurs. However, in so doing, dependent overdetermination implies that physical causes and mental causes are not sufficient in themselves. The idea of dissolving the problem of causal exclusion by endorsing overdetermination, therefore, is not particularly viable. However, the discussion led to the consideration of a strategy which construes the mental as playing a necessary non-causal determinative role.

## **Chapter 5 - Causal Exclusion and Functional Reduction**

The problem of causal exclusion could not be resolved by endorsing the models presented in chapter three, nor could it be resolved by rejecting the fourth principle, that of causal exclusion. In this chapter I will consider the possibility of rejecting the third principle, that of irreducibility. Although some reason has been given to posit a distinction between the mental and the physical, the tension that is created by assuming the mental is irreducible has also been evident. For this reason, I will now consider the viability of Jaegwon Kim's reductive solution to the problem of mental causation.

This chapter is divided into four sections. In Section 5.1, I outline Kim's model of functional reduction and his solution to the problem of causal/explanatory exclusion. In Sections 5.2 and 5.3 I provide various reasons to conclude that it is difficult to secure mental causation with Kim's reductionism. Finally, in Section 5.4, I argue that on Kim's reductionism, mental descriptions are eliminable as well.

### **5.1 – Kim's Functional Reduction**

Kim's favoured method for overcoming the problem of causal exclusion is functional reduction. He gives specific details about how his model of functional reduction is supposed to work in a number of places (Kim, 1998, p. 97; Kim, 1999, p. 10; Kim, 2005, p. 101; Kim, 2006, p. 280; Kim, 2008, p. 104; Kim, 2009, p. 46). According to this model, if we want to reduce mental property *M*, we must first of all prepare it for reduction by providing a functional definition of it. As Kim says, “property *M* to be reduced is given a functional definition of the following form: having *M* = <sub>def.</sub> having some property or other *P* such that *P* performs causal task *C*” (Kim, 2005, p. 101). Call this functional definition *h*. Thus, for example, the desire for a Big Mac (*M*) is defined as being in whatever state is caused by viewing McDonald's billboards and causes the body to wind up close to this type of hamburger (*h*).

After we have given *M* a functional definition *h*, we can now “find the property, or mechanism, that does the causal work” (Kim, 2006, p. 280). In other words, we need

to locate, or “ascertain” (Kim, 2008, p. 104), which physical properties realize the specified causal role. Let us say, for the sake of the argument, that dopamine release in the dorsal stream ( $P_1$ ) is caused by viewing McDonald’s billboards and causes the body to be close to hamburgers. This means that  $P_1$  is the realizer of  $h$  on this particular occasion. We should add that a number of different physical systems can realize the specific causal task defined by  $h$ . Conceivably, in robots, electrical signals flowing through a certain circuit ( $P_2$ ) is caused by the presence of McDonald’s billboards and causes the robot to be close to a hamburger ( $h$ ).

After we have located the realizer ( $P_1$ ), we now need to develop a theory that explains how the appropriate physical properties perform the causal task specified in the definition  $h$  (Kim, 2008, p. 104). This is a matter for scientific research. We can let the scientists figure out how and why dopamine is released in the dorsal stream whenever McDonald’s billboards are seen, and why this release of dopamine causes the body to walk near to hamburgers. We need not worry about these scientific matters, we only need to conclude that mental property  $M$  gets defined as  $h$ , which is in turn realized by some physical property,  $P_1$ , or  $P_2$  or  $P_3$  or ...  $P_n$ .

We do not yet have a complete reduction, however, since we have not yet explicitly seen  $M$  reductively identified with anything. As it turns out, there are a number of possible ways in which the reductive identity can be carried out. We will look at three such ways (as labeled by Kim): disjunctive functional reductionism (Kim, 2008, p. 108-109), functional property conceptualism (Kim, 2008, p. 111-112), and token reductionism (Kim, 2008, p. 106-107). Kim has oscillated between these models for some time now (see, for example, Kim, 1992, p. 24-25). A number of critics think that Kim currently endorses functional property conceptualism (Rueger, 2004, p. 11; Khalidi, 2005, p. 3). Most recent evidence suggests that Kim endorses functional property conceptualism mixed with token reductionism (Kim, 2008). Out of an abundance of caution, however, I want to consider the consequences of all three models.

For now, let us simply assume that Kim identifies the mental with the physical in some manner. How would such an identity solve the causal exclusion problem? If we can identify the mental with a causally potent physical realizer, then the mental regains the causal power that was once threatened by the exclusion argument. Kim explains:

Mentality has causal effects in the physical world; however, the physical world is causally closed; therefore, mentality must be part of the physical world, and specifically, mental states are identical with brain states (Kim, 2005, p. 125).

Kim once diagramed the problem of mental causation in terms of  $M$  and  $M^*$  supervening above and outside of the causal stream from  $P$  to  $P^*$  respectively, and this blocked off  $M$  as a second potential cause. Now Kim fuses the  $M$  to  $M^*$  line into the  $P$  to  $P^*$  line, leaving only one ostensive cause, thereby alleviating exclusion pressures. This solution falls into the reduced mental causation camp, which was not our preferred model, but at least it was one such model of mental causation.

How does the identity help to solve the problem of explanatory exclusion? As we saw in Section 2.6, Kim grants that psychological and physiological descriptions are intensionally distinct, so there are two descriptions. However, as we've now seen as well, they both refer to one and the same underlying causal process. If Kim adopts a model whereby explanations are not differentiated conceptually, but rather explanations are considered distinct only if they have different external content, then Kim can dodge the problem of explanatory exclusion.

This is exactly what Kim does. Kim argues that “the content of the explanation consisting of  $C$  and  $E$ ” is that  $c$  and  $e$  are related by objective relation  $R$  (Kim, 1988, p. 226). Here  $C$  is an explanans statement and  $E$  is an explanandum statement, while  $c$  is the cause of the effect  $e$ . The objective relation  $R$  between  $c$  and  $e$  “is what the explanation says” (Kim, 1988, p. 226), so if two nonequivalent descriptions refer to this same causal relation between events, then they state the same explanation. And, if they state the same explanation, there is little reason to consider them distinct:

As for explanation, at least in an objective sense, there is one explanation here, and not two. The two explanations differ only in the linguistic apparatus used in referring to, or picking out, the conditions and events that do the explaining; they are only descriptive variants of one another. They perhaps give causal information about E in different ways, each appropriate in a particular explanatory context; but they both point to one objective causal connection, and are grounded in this single causal fact (Kim, 1989, p. 87. See also Kim, 1988, p. 233; Kim, 1989, p. 80; Kim, 2005, p. 114-115; Kim, 2005, p. 144-146).

Since explanations are individuated in terms of events, and the mental event is the physical event, there is only one explanation. It is worth noting that if this “explanatory externalism” (Kim, 1994, p. 57) proves untenable, as I will argue it does in Section 6.3, then Kim will no longer have a solution to the problem of explanatory exclusion. Rather, he will be left with two intensionally distinct descriptions of one event, which welcomes explanatory exclusion pressures back again.

### **5.2 – Functional Reduction and Elimination**

Does Kim’s solution adequately resolve the causal exclusion problem? In this section I argue that the answer is no. Why not? First of all, Kim’s model does not secure autonomous mental causation for us. Autonomous mental causation requires that the mental cause play a distinct role from the physical cause. As Kim identifies the mental cause with the physical cause, the mental cause will not play a distinct role. For this reason, we should not prefer Kim’s model of reduced mental causation. However, perhaps autonomous mental causation is out of our reach, as Kim thinks. In this case, if we endorse reduced mental causation at least we will have some form of mental causation. In this section I argue that there is in fact reason to believe that Kim slides past reductionism and right into eliminativism. If this is the case, then Kim’s model not only fails to provide us with autonomous mental causation, but it fails to provide any type

of mental causation. I will begin by considering what happens with mental causation on all three of the aforementioned models of functional reduction.

### **5.2.a – Disjunctive Functional Reductionism**

The disjunctive model of functional reduction appears in *Mind in a Physical World*. Here Kim rejects the Nagelian bridge-law model of reduction between mental and physical properties ( $M \leftrightarrow P$ ) in favour of reduction via property identity ( $M = P$ ) (Kim, 1998, p. 97). If reduction via property identity is successful, then it will be the case that “ $M$  and  $P$  are ... in fact one and the same property” (Kim, 1998, p. 98). Kim arrives at this property identity through functional reduction. On this model, functional reduction begins by preparing  $M$  for reduction by “construing  $M$  as a second-order property defined by its causal role” (Kim, 1998, p. 98). In other words, we turn  $M$  into a second-order functional property with a certain causal profile. Call this functional property  $H$ . We should note here that not only is  $M$  given a functional definition  $h$ , as we’ve seen above, but it is also turned into the functional property  $H$ , where mental property  $M =$  functional property  $H$ .

As it turns out, however, there will be some property  $P$  that is “exactly the property that fits the causal specification” (Kim, 1998, p. 98) given for this particular functional property  $H$ . We can therefore identify property  $H$  with property  $P$ . Thus, if mental property  $M =$  functional property  $H$ , but functional property  $H =$  physical property  $P$ , then mental property  $M =$  physical property  $P$ . As Kim summarizes:

This grounds the identification of  $M$  with  $P$ .  $M$  is the property of having some property that meets specification  $H$ , and  $P$  is the property that meets  $H$ . So  $M$  is the property of having  $P$ . But in general the property of having property  $Q =$  property  $Q$ . It follows then that  $M$  is  $P$  (Kim, 1998, p. 98-99).

The reductive identity between  $M$  and  $P$  is forged here, but we should note that the process is slightly more complicated than this. Recall that  $h$ , and hence functional property  $H$ , is multiply realizable. This means that there will be a number of physical properties that realize  $H$ . As we’ve seen, the desire for a hamburger ( $H$ ) was realized by

dopamine release ( $P_1$ ) in humans and electrical firing ( $P_2$ ) in robots. It could also be the squishing of green goop ( $P_3$ ) in aliens, etc... Thus, it is fair to conclude that  $H$  will be identical to a rather large disjunctive set of physical properties:  $H = [P_1 \text{ or } P_2 \text{ or } P_3 \text{ or } \dots P_n]$ . This is why Kim calls this the disjunctive model.  $M$  is identical with  $H$ , which is in turn identical to a disjunction of physical properties  $[P_1 \dots P_n]$ .

There are a number of problems with this disjunctive model. Immediately after characterizing functional reduction as a matter involving second-order functional properties, Kim wonders whether the mental can be a second-order property or must merely be a conceptual predicate. He describes the problem as follows: “how could  $M$  be identical with  $P$ ? Isn’t it incoherent to think that a property could be both first order and second order, both extrinsic/relational and intrinsic, both a role and its occupier” (Kim, 1998, p. 103)? The identity is strained because this property would be higher order than itself; though there is only one property, it would entirely be a role in need of an occupant while also being entirely the occupant of the role; though it is by definition intrinsic, it would be defined extrinsically. Kim agrees that the identity seems incoherent, and suggests that we may not want to think of the mental as a property, but rather as a concept alone (Kim, 1998, p. 103). On this matter, a number of critics agree with Kim that the identity in question here is difficult to maintain (Horgan, 1997, p. 175; Jackson, 2002, p. 646; Block, 2003, p. 145).

A second worry stems from the fact that property  $H$  is multiply realizable, and as such would be a disjunctive property. Disjunctive properties, according to some, are not projectible (they do not pick out natural kinds, and cannot figure in laws), so it is unlikely that they are real properties (Kim, 1992; Kim, 1998, p. 104-110; Kim, 1999, p. 17-18; Kim, 2002, p. 672; Kim, 2009, p. 47). The raison d’être of the property is therefore undermined. There is no good reason to keep functionalized mental property  $H$  around if it cannot figure in laws and cannot provide any useful explanations. Parsimony seems to require us to eliminate it from our ontology and our explanatory practice. A number of

critics think Kim is on the right track here in abandoning disjunctive properties (Pineda, 2002, p. 31-33; Gozzano, 2003).

Sven Walter has noted a third problem with the disjunctive property model (Walter, 2008, p. 691). Presumably,  $P_1$  is a property in its own right. If this is so, then we need to ask whether the property of being  $P_1$  has sufficient causal powers in its own right to determine what an instantiation of  $P_1$  can do. Seemingly, the answer is yes. If the answer is yes, then what work is left over for the disjunctive property of ‘being  $P_1$  or being  $P_2$  or being  $P_3$ ’ to do? There is none. The individual property excludes the disjunctive property from doing any work, so the disjunctive property would be epiphenomenal at best.

All things considered, it seems reasonable to conclude that disjunctive functional reductionism will not work. For his own part, Kim occasionally hints at retaining this model (Kim, 2002, p. 678-679; Kim, 2005, p. 58; Kim, 2009, p. 47), but he rejects it for the most part (Kim, 1998, p. 104). He goes so far as to conclude that, “the reasons for rejecting [disjunctive functional reduction] are quite compelling” (Kim, 2008, p. 112).

### ***5.2.b – Conceptualized Functional Reduction***

We have seen that the disjunctive model of functional reduction does not secure mental causation. Fortunately, Kim has an alternative model. As hinted at above, Kim’s suggested alternative to the disjunctive approach is to conceive of the mental as a concept alone. Kim has frequently endorsed this model (Kim, 1999, p. 10; Kim, 2005, p. 101; Kim, 2006, p. 280), and in his most recent work on the subject he continues to deploy it as part of his strategy (Kim, 2008, p. 111-112). According to this approach, property  $M$  gets defined in terms of its causal role  $h$ . This causal definition  $h$  does not refer to disjunctive property  $H$ , however. Rather, disjunctive property  $H$  does not exist, and this leaves  $h$  as a concept alone. The causal task defined in  $h$  has various realizers:  $P_1, P_2, \dots, P_n$ .  $M$  is then identified with whichever particular realizer is relevant on a given occasion. The strengths of this move are that it avoids the difficulties with identifying

second-order properties with first-order properties, and it does not require us to add disjunctive properties to our ontology.

Conceptualized functional reduction comes at a price, however. Namely, the mental now exists as a predicate alone, it does not exist independently of language, as part of our ontology. A number of people have rejected this conceptual model on account of this apparent eliminativism of the mental. Pierre Jacob, for example, says:

Kim entertains the radical view that one ought to trade higher-order functional properties for higher-order functional concepts. This proposal, I think, amounts to giving up reductionism and embracing eliminativism (Jacob, 2002, p. 253. See also Sabates, 2002, p. 666-667; David, 1997; Marras, 2003, p. 255; Walter, 2008, p. 40).

Functionalized mental properties no longer exist when the mental is understood as existing conceptually alone. The remaining functionalized mental predicates are not part of Kim's ontology, so there is a very real sense in which they do not exist.

In various places, Kim acknowledges that this move seems to eliminate mental properties:

This option may sound like a form of eliminativism, and perhaps rightly so. *M* as a property is gone; it has been eliminated. It remains true that, in Smart's idiom, *M* as a property is nothing "over and above" its realizing properties for the trivial reason that *M* is nothing. If mental properties are functionally reduced, we may well have to live with mental eliminativism and irrealism. (Kim, 2008, p. 111-112. See also Kim, 2005, p. 160; Kim, 2002, p. 674).

In fact, Kim goes so far as to call the position that exchanges mental properties for mental concepts by the name of "eliminativism" on occasion (Kim, 1999, p. 17; Kim, 2002, p. 643).

On most occasions, however, Kim is reluctant to accept the elimination of the mental. There are two avenues he takes to avoid these eliminative consequences, both of

which are lumped into Kim's response to Marcelo Sabates' objection that conceptual reductionism leads to eliminativism:

A mental predicate, or concept, on my account, is true of its physical realizers.

When I say ' $x$  is  $M$  at  $t$ ', where ' $M$ ' is a mental predicate, its 'truth-maker' is the fact that  $x$  has  $P_1$  at  $t$ , where  $P_1$  is one of  $M$ 's realizers and, as a physical property, it is as 'real' as anything else. Does that seem like eliminativism about  $M$  or about the fact that  $x$  is  $M$  at  $t$ ? It does not to me. Moreover, I have been torn between the conceptual approach recommended in my book and the disjunction approach also discussed in the book (Kim, 2002, p. 678).

Notice, first of all, that Kim suggests mental concepts are real because they have truth-makers. Kim makes this point in his discussion in *Mind in a Physical World* as well (Kim, 1998, p. 105; see also Robb, 2008; Pineda, 2002, p. 31).

This insight may be helpful in the study of truth, and perhaps epistemology, but it is not clear that it helps in our study of causality. There are two reasons for this. First of all, consider Donald Davidson, who makes a similar maneuver. As we will see in greater detail in Section 6.2, Davidson claims that mental predicates truly describe a token event, much as Kim is doing here (Davidson, 1993, p. 12-13). Kim rejects this move because "the issue has always been about the causal efficacy of properties" (Kim, 1993, p. 21). In other words, it is not sufficient for mental causation that an event can be truly described using mental predicates, it only matters whether there are causally potent mental properties or not. Davidson's view suggests there are not, so Kim accuses him of epiphenomenalism. In this present context, therefore, we can justifiably complain that Kim's mental concepts, though they may have truth-makers, still do not themselves cause anything (Jacob, 2002, p. 653). For, presumably a pre-requisite for having causal power is to actually exist. If the mental doesn't exist as a property, as part of our ontology, it cannot have causal efficacy.

Secondly, as we will explore in Section 6.1, Kim endorses what Terrence Horgan calls the quausal model of causation (Horgan, 1989). This means that events

cause *qua* a specific property, or in virtue of a specific property. According to Kim, we achieve mental causation if an event has causal power in virtue of its mental properties. Not only have we seen Kim acknowledge this point already, but Kim has recently taken an even clearer stance on this issue:

The fact that properties  $M$  and  $P$  must be implicated in the identity, or nonidentity, of  $M$  and  $P$  instances can be seen from the fact that ‘An  $M$ -instance causes a  $P$ -instance’ must be understood with the proviso ‘in virtue of the former being an instance of  $M$  and the latter an instance of  $P$ ’ (Kim, 2005, p. 42).

A mental event implicates the property  $M$ . Recall, however, that on the conceptual model of functional reduction there is no mental property  $M$ . Thus, mental causation requires an event to cause in virtue of a mental property, but on the conceptual model there is no mental property, so the event cannot cause in virtue of a mental property, so mental causation is not achieved. As we see in Kim’s response to Sabates above, these troubles are so damaging to conceptualized functional reduction that Kim is tempted to revisit his discarded model of disjunctive functional reduction.

### 5.2.c – *Token Reductionism*

The conceptual model threatens the loss of mental causation since the mental does not exist as a property. However, the final model that Kim discusses may be able to secure mental causation against these difficulties. According to token reductionism,  $M$  is once again defined in terms of a specific causal task  $h$ . This causal definition  $h$  still does not have a corresponding disjunctive property  $H$ . However, the task defined in  $h$  is realized by a specific physical property instance  $p_1$ . Thus, an instance of  $M$  can be identified with an instance of  $P_1$ . Kim explains this move as follows:

If  $x$  has  $M$  at  $t$ , it follows that  $x$  instantiates at  $t$  a realizer of  $M$ , say  $P_k$ . I believe that the identification of this instance of  $M$ ,  $x$ ’s having  $M$  at  $t$ , with the instance of the realizer involved on this occasion,  $x$ ’s having  $P_k$  at  $t$ , is all but compelling ... these considerations point to a form of token identity thesis: this instance of  $M$  is

identical with the instance of the realizer of  $M$  on this occasion. In general, then, if mental properties are functionally reducible, instances of mental properties are identical with the instances of their physical realizers. That, at any rate, is the proposal. We may call it token reductionism (Kim, 2008, p. 106; See also Kim, 1998, p. 110-111; Kim, 2005, p. 58; Kim, 2009, p. 47).

Perhaps, for the sake of clarity, we can express this view as follows:  $m_1 = p_1$  and on another occasion  $m_2 = p_2$  and later on  $m_3 = p_3$  and so on and so forth. A mental property instantiation (lower case  $m_1$ ) is identical to a certain physical property instantiation on one occasion (lower case  $p_1$ ), and another mental property instantiation ( $m_2$ ) is identical to a different physical property instantiation on a different occasion (lower case  $p_2$ ). If the identity occurs at the level of instances we avoid the problems associated with the disjunctive model, while simultaneously securing mental causation due to the identification of the functionalized mental property instance with the causally potent physical property instance.

Although Kim endorses token reductionism on certain occasions, he also calls it into question at other times (Kim, 2005, p. 58), while leaving the other two options open on other occasions as well. There are, after all, several significant issues that it faces. First of all, this is still close to the Davidsonian token identity position, which we will see Kim reject in Section 6.2. Kim rejects Davidson's model because, as we will see, Kim endorses a quausal model of causation. According to the quausal model, we cannot only say that a token event has causal power, rather we have to add that this token event caused in virtue of a property. This being the case, we need to ask ourselves whether a token event occurs in virtue of a mental property on token reductionism.

In response, we need to remind ourselves that there is still no overarching functionalized mental property  $H$  on token reductionism. Token reductionism avoids the problems of the disjunctive model by identifying a functionalized mental instance with a physical instance. Thus, if we want to say that an event occurred in virtue of

functionalized mental property  $H$ , we would not be able to, for this property does not exist.

Perhaps we can agree that  $H$  does not exist as the all-encompassing property type envisioned on the disjunctive approach, but still maintain that the mental property is identical to whichever physical property was instantiated on this occasion. This would allow us to preserve the quausal model, for we can still say that an event caused in virtue of mental property  $M$ , since mental property  $M$  is whichever physical property that was instantiated on this occasion, say  $P_1$ . This option faces certain difficulties as well. Let us consider what happens if the fact that an instance  $m$  is an instance  $p_1$  on one occasion implies that the property  $M$  is the property  $P_1$ . Now of course,  $P_1$  is not the same property as  $P_2$ .  $M$  is, however,  $P_1$ . This means that  $M$  is not the same property as  $P_2$ . Similarly,  $M$  cannot be  $P_3 \dots P_n$ . This move, therefore, limits the future appearance of property  $M$  to systems that instantiate  $P_1$ . For if  $M$  is identical to  $P_1$  then wherever  $P_2$  occurs,  $P_2$  cannot be  $M$ , since  $P_2$  is not  $P_1$ . Therefore, identifying  $m$  with  $p_1$  with the additional caveat that  $M$  is  $P_1$  so we can secure mental quausation leads to limiting the locations where the instance  $m$  can be, which calls multiple realizability into question.

The remaining option for token reductionism seems to be to argue that the identity occurs at the level of instances ( $m = p_1$ ), but this token identity does not imply the property identity of  $M = P_1$  as well. We may be able to say that this preserves mental causation, since causation occurs at the level of instances, where we do have the identity (Kim, 2005, p. 58). However, we cannot say that this model secures mental quausation. Recall that according to Kim's quausal model, he needs an event to occur in virtue of a property. It is not enough for the mental event to be causally efficacious as the token. Rather, as we have seen, the mental event must be efficacious in virtue of being the mental property.

Perhaps the token reductionist can preserve quausation without making the identity between mental property  $M$  and physical property  $P_1$ . This can be achieved by pointing out that this event still occurs in virtue of a property, namely, it occurs in virtue

of the physical property  $P_1$ . This move would preserve quausation, as it invokes the property in the analysis. However, it would not preserve mental quausation, since it states that this event occurs in virtue of the physical property, which is not identical to the mental property.

The remaining option seems to be to resort to the brute token identity in an effort to save mental causation, without retaining mental quausation. As mentioned, and as we will see in Section 6.2, this is essentially Davidson's position – a token identity without worrying about mental causation in virtue of mental properties. Kim, however, insists on the importance of the property in the causal process, and this token reductionism does not pass this strict requirement. He would have to reject the quausal model, thereby securing mental causation while simultaneously claiming that we do not need to add anything else, to achieve a satisfactory picture of mental causation. Until such a time as Kim rejects the quausal model, it seems that his version of token reductionism will not satisfy his own quausal criterion. The Kimian version of token reductionism has the additional conceptual problem of positing instances of the mental property  $M$  without the existence of mental property  $M$  in the first place. It is difficult to see how we can have an example of a property without having a property in the first place, but this appears to be the case on Kim's model of token reductionism. On all three models of functional reductionism, therefore, it is difficult to secure mental causation.

### **5.3 – Epiphenomenalism and Properties of Properties**

There is a second, independent reason to think that Kim's model of functional reduction actually ends up in eliminating a causal role for the mental. This reason is discernible when we analyze a dispute between Kim and Davidson about the nature of events. Again, we will consider this dispute in detail in Section 6.2. For now, however, we once again only need the fact that events cause as events, they do not cause in virtue of certain properties of events, because events do not have causally efficacious properties. Kim thinks Davidson's view is too coarse. Kim then endorses a fine-grained model which

suggests that we can analyze events into their component parts and see that events cause in virtue of certain causal properties (Kim, 1993, p. 21). The important point: Kim is not content with the coarse view, rather he has a desire for greater precision.

This discussion can be transposed one step further in the following way: does a property have various aspects to it, or should we be content to rest with the coarseness of saying that it is simply one austere property? Presumably, Kim's convictions would lead him to say that we can break the property down into its aspects, if a property has aspects/properties of its own. Let us ask this last question first, then: does a property itself have aspects/properties of its own? In Section 3.5 we saw a number of philosophers endorse the view that scarlet can have redness as a property. The property of being scarlet seems to have other properties as well. Scarlet has the property of being instantiable, the property of being the same as itself, the property of being different from the property of being green, the property of being a colour, the property of being the colour of very ripe tomatoes, the property of causing bulls to get angry, the property of being composed of wavelengths of light of approximately 700 nm, and so on and so forth.

For his own part, there are several places where properties of properties make an appearance in Kim's overall work. First of all, Kim's functional reduction makes explicit use of second-order properties, which seem to be construed as properties of properties. On the functional model, Kim often takes a mental property and redefines it according to its causal properties. Pain, for example, is defined as being in a state caused by tissue damage, and causing winces and groans (Kim, 2006, p. 280). Notice that the intrinsic feel of pain is not eliminated. Indeed, Kim thinks that this qualitative aspect remains unreduced (Kim, 2005, p. 174). So, our mental property has a qualitative aspect to it, and a causal aspect to it.

According to Kim, this causal aspect of the property is a second order property in its own right. David Armstrong, from whom Kim received the functional model (Kim, 1998, p. 98), argues that second order properties are properties of properties (Armstrong,

1978b, p. 133-143). As a typical example, a sweater may have the property of being red, while this redness itself has the property of being a primary colour. As for Kim's functional model, he states that second order properties are the "property of having a property..." (Kim, 1998, p. 98). This means that the second order property itself is had by a property. Further, Kim states that these second order properties quantify "over a given domain of properties" (Kim, 1998, p. 103). This being the case, it seems that Kim is following Armstrong in concluding that second order properties are properties of first order properties. Our mental property, then, has a qualitative aspect to it, a second order causal aspect to it, and a first order physical realizer aspect to it. One property has three aspects (or properties) to it.

The second place that properties of properties make their appearance is in Kim's discussion of the generalization problem. We will save detailed discussion of this issue for Section 8.1, but for now it suffices to point out that Kim introduces the notion of micro-based properties. These micro-based properties are properties of a whole, but they themselves have further properties as well. To use Kim's example, water has the micro-based property of being composed of hydrogen and oxygen in a certain bonding relationship (Kim, 1998, p. 84-85). This micro-based property in turn has certain properties, such as the ability to douse flames.

Having established properties can have properties, we can return to the previous question: should we be satisfied with the coarse conception of properties that does not analyze properties according to their own aspects, or should we opt for the fine grained route? There are two reasons why Kim may choose the fine-grained route. First, all things being equal, it is best to remain consistent. He endorses the fine-grained route against Davidson, so he may want to follow the same model now. Secondly, if he abandons the fine-grained account, then what is there to stop Davidson from insisting that events are coarse-grained themselves? Kim rejected Davidson's account because events could be more finely analyzed, but if Kim rejects the fine-grained account after all then his argument against Davidson loses some weight.

Assuming, therefore, properties have properties, and properties ought to be analyzed in terms of these further properties, we can ask the following question: is it in virtue of the role aspect of the physical property, or is it in virtue of the realizer aspect of the physical property, that the physical property has causal efficacy? In other words, is it in virtue of the mental aspect, or the physical aspect, that the property has causal efficacy?<sup>30</sup> Importantly, one of the times that Kim uses properties of properties is in his model of functional reduction. So it is no stretch to talk as though this one complex property has a role aspect and a realizing aspect to it. These aspects may belong to the same property, but this does not mean that the role aspect of the property is identical to the realizing aspect of the property. Presumably, when push comes to shove, there is some reason to conclude that the realizing aspect of the property is doing all of the causal work. If this is the case, then it turns out that the role aspect of the property, also called the mental aspect, does not make a causal difference. Since the physical aspect of the property wholly takes up one hundred percent of the causal efficacy of this property, there is no work left for the mental aspect of this property, despite the fact that it is the same property.

#### **5.4 – Explanatory Exclusion and Unnecessary Mental Explanation**

Not only are there these reasons to doubt that Kim's reduction delivers mental causation for us, but as it turns out there is some reason to consider mental explanations eliminable as well. Recall that Kim solves the problem of explanatory exclusion by appealing to one cause that underlies both explanations. Given this model, let us ask the following

---

<sup>30</sup> Noordhof asks much the same question in a discussion he has with Robb: "My house burns down. It is quite legitimate to ask which aspect of air was responsible. The answer is that the air was causally relevant in virtue of being part oxygen. So it seems that complex properties do have aspects concerning which one can ask 'was that responsible?'(Noordhof, p. 223) Someone's house burns down, and this happened in virtue of the presence of air. But what aspect of air? The aspect of being part oxygen. For further discussion, see Sawyer, 1996, p. 243; Armstrong, 1989, p. 105; Gibb, 2004, p. 475; MacDonald, 2006, p. 18, p. 22 of 31; Whittle, 2007, p. 66-67.

question: since there is one cause, what need is there to give two descriptions (one mental and one physical) of that cause? One possible response: the physical description, in itself, is not complete, so we also need to provide the mental description. This response does not fare very well. For, given causal completeness, we can accept the plausible assertion that every event has a complete physical description as well. The question then becomes more poignant: given that there is a complete physical description for a given event, why do we need to provide a mental description as well? Andrew Melnyk considers, without endorsing, the same point:

Suppose you are also a physicalist, so that you believe (at least) that every event, without exception, just is some fundamental physical event. It seems to follow that every event, without exception, has a fundamental physical explanation. But if every event, without exception, has a fundamental physical explanation, then every event, without exception, has an explanation. What, therefore, is the point of the explanations apparently supplied by the special sciences, i.e., the sciences distinct from fundamental physics? They seem, indeed, quite needless, since they explain nothing that is not already explained. But if, like explanations citing phlogiston, they are explanatorily dispensable, surely we should dispense with them, just as we have dispensed with the phlogiston citing explanations ... People might raise an awkward question: why keep it around? (Melnyk, 1996, p. 185; See also Melnyk, 2003, p. 166-169; Horgan, 1991; Jacquette, 1994, p. 34-36).

In the case of phlogiston, after a complete chemical description of oxidation presented itself, we no longer needed phlogiston, and it was thereby eliminated. In the case of mental events, we can also have a complete physical description of a given event, so we also seem to no longer need mental descriptions, and they can thereby be eliminated.<sup>31</sup>

---

<sup>31</sup> Perhaps we could reply to this by saying that we do not yet have a complete physical description of many complex neural events, so mental descriptions are still necessary. This may be true, but this means that mental descriptions are

Perhaps we can reply that that the completeness of the physical description renders the mental description unnecessary, but we can still keep the mental description around because we want it. There is, however, reason to think that if mental descriptions are truly unnecessary, we should not keep them around. Recall that Kim appeals to Ockham's razor to discourage the multiplication of explanations beyond necessity. He is even open to using exclusion to eliminate multiple explanations as well (Kim, 1989, p. 101; Kim, 1989, p. 106). These same values of parsimony and simplicity suggest that we should not multiply descriptions beyond necessity either. A fuller application of Ockham's razor, therefore, seems to be that once a complete physical description is given, we should not continue to provide a mental description as well.

In order to avoid eliminating the mental description due to Ockham's razor, we will need a reason for retaining mental descriptions. Several possible reasons are available. First of all, we could simply point out that mental properties do exist, in virtue of them being (identical with) existing physical properties. Since mental properties exist, it would not be fitting to leave out mental descriptions. This may be what Kim is hinting at when he addresses the issue in *Physicalism or Something Near Enough*:

Some will say that the reductionist option is hardly distinguishable from eliminativism ... there is an honest difference between elimination and conservative reduction. Phlogiston was eliminated, not reduced; temperature and heat were reduced, not eliminated (Kim, 2005, p. 160. See also Kim, 2006, p. 276).

There is no phlogiston and there are no witches. These things never existed, so descriptions invoking them can be eliminated. On the contrary, there is water and there are mental properties. These things really do exist, and have always done so, so we can continue to talk in these terms.

---

simply a stop gap until complete physical descriptions are available. The argument would then read that once complete physical descriptions are available, then mental descriptions will no longer be necessary, so they can be eliminated.

Justin Schwartz uses this ontological identity between water and H<sub>2</sub>O to secure a conservative reduction of water. He says:

Now if Lavoisier established that water is identical to H<sub>2</sub>O, then H<sub>2</sub>O will have all and only the properties of water. This is rather far from saying that there is no water! Showing that water is not just correlated with H<sub>2</sub>O but is identical to it eliminates the separateness of water but not its existence" (Schwartz, 1991, p. 209-210. See also Sachse, 2007, p. 7).

Water exists because it is ontologically identical to H<sub>2</sub>O, which exists. On the contrary, presumably phlogiston does not really exist, and so descriptions involving phlogiston can be eliminated. Since water exists, we can continue to offer descriptions in terms of water. Similarly, since mental properties exist, we ought to continue to offer mental descriptions.

The matter, however, is not so simple. First of all, as we have seen, it is not clear that the mental property does still exist on Kim's model of reductionism. If the mental property is eliminated, then Kim may not be able to use this argument for keeping the mental description. For the sake of the argument, however, let us assume that mental properties are identical to physical properties on Kim's model. But if mental descriptions are a necessary component for a full account of a given cause, doesn't this mean that physical descriptions are not complete descriptions of this cause anymore? If the physical description is complete, then we do not need a mental description as well. However, if we need the mental description, it seems like the physical description is incomplete. In order to secure explanatory completeness, we need to conclude that the physical description is complete, so we cannot need an additional mental description as well.

Finally, consider the following problem: a log starts to flame in an enclosed space while two people are standing nearby. The first, Joseph Priestly, points to the air and says the log is burning because the air is dephlogisticated. Before long, the log stops burning and Priestly explains that this is because the air is now too phlogisticated. The

second man, Antoine Lavoisier, points to the same air and says that the same log is burning because of the oxygen present in the air. When the log stops burning, Lavoisier explains that the oxygen is gone from the air, so the log cannot burn any longer. Both people see the same effects (i.e., the log burning), and they both point to the same region of space (i.e., the air), indicating they are talking about the same series of events and objective relations. Does this mean that they stated the same explanation?

Some have supposed that this could theoretically be the case (Putnam, 1994, p. 15; Patricia Churchland, 1986, p. 282; Kitcher, 1978, p. 534). More importantly, some of Priestly's followers have supposed so as well. William Odling, for example, posits an identity between dephlogisticated air and oxygenated air:

Cavendish showed that when inflammable air or hydrogen, and dephlogisticated air or oxygen, are exploded together in certain proportions, 'almost the whole of the inflammable and dephlogisticated air is converted into pure water' ... When Lavoisier spoke of red lead as being metallic lead combined with oxygen, he meant that the matter or stuff of the red lead consisted of the matter or stuff of lead plus the matter or stuff of oxygen. But when the Stahlians spoke of metallic lead being burnt lead combined with phlogiston, they had the same sort of idea of combination in this instance (Odling, 1871, p. 257. See also, Kitcher, 1978, p. 534; Kitcher, 1992, p. 100; Gyung Kim, 2008).

Here Odling equates dephlogisticated air with oxygen, and he equates various combinations involving oxygen with various combinations involving phlogiston. Using the same logic, since oxygenated air exists, and dephlogisticated air is oxygenated air, then dephlogisticated air continues to exist, and ought not be eliminated (Melniky, 1996, p. 189). This shows us that it will not work to simply assert that an entity being considered for elimination or reduction actually does (or does not) exist. For, what we are trying to determine is whether this entity under consideration does (or does not) exist. We cannot beg the question by presupposing that a given entity does (or does not) exist. This is especially poignant in the case of mental descriptions. For, there are a number of

people who think that the mental ought to be eliminated (Feyerabend, 1981; Churchland, 1988; Rorty, 1970; Dennett, 1988; Wilkes, 1995), while there are also a number who think the mental ought to be conservatively reduced (Kim, 2005, p. 160). We cannot assume that mental descriptions ought to be preserved, for this very question continues to be considerably contested, and it is the question we are trying to answer.

What we need is a non-arbitrary way to sort between entities fit for the reduction pile and entities fit for the elimination pile. We need to be able to say that witches and phlogiston can be eliminated for this reason, while water and mental properties can be conservatively reduced for that reason. There is a possible litmus test. Perhaps we can say that whenever two competing descriptions cut the world up in largely different ways, then one of them has to be eliminated. Paul Churchland, for example, argues that temperature can be conservatively reduced to mean kinetic energy because the way they cut up the world is similar enough (Churchland, 1998, p. 68). However, phlogiston theory says the world is doing something radically different than oxidation chemistry, so it must be eliminated:

Phlogiston emerged, not as an incomplete description of what was going on, but as a radical misdescription. Phlogiston was therefore not suitable for reduction to or identification with some notion from within the new oxygen chemistry, and it was simply eliminated from science” (Churchland, 1988, p. 44).

The assumption that one theory cannot be conservatively reduced to another if it is widely divergent is quite common (Popper, 1957, p. 29; Hooker, 1981, p. 49; Bickle, 1998, p. 29; Churchland, 1989, p. 50; Poirier, 2006, p. 478). Indeed, a thrown pebble follows a parabolic trajectory for Galileo, while the path is elliptical for Newton. The pebble cannot travel in both a parabolic and an elliptical manner at the same time. Since both concepts say the world is doing divergent things, a conservative reduction cannot be posited, and so Galileo’s theory has to be eliminated.

Though this model may have some appeal, there are several reasons to think it is not helpful in this particular case. First of all, Jaegwon Kim’s model of functional

reduction does not include such a test for banning improbable reductions. Consider what happens when Joseph Priestly plugs phlogiston theory into Kim's functional apparatus. Dephlogisticated air is first of all defined as being whatever causes wood to burn. As it turns out, the presence of oxygen is what causes wood to burn. Dephlogisticated air is therefore conservatively reduced to the presence of oxygen. Kim would have to add some sort of mechanism that discerns when two descriptions are sufficiently incongruent, and then ban one of them. But he has, as of yet, no such mechanism, so he cannot yet sort between reduced and eliminated entities in this way.

Beyond this, as noted above, numerous people think that mental theories are different from their base theories, so they may not pass the criterion up for consideration (Feyerabend, 1981; Stich, 1983; Nosek, 1995; Churchland, 1998). According to these people, folk psychology does not smoothly map onto neurophysiology, and so identities are unlikely, and elimination is forthcoming. For his own part, we have seen that Kim thinks there is some divergence between mental and physical properties. Although he still thinks the reduction can go through, it may not be because mental and physical properties smoothly map onto each other. Rather it may be because he does not share the conviction that reduced properties must be able to smoothly map onto their bases.

Perhaps an instrumental criterion for discerning between entities fit for reduction and elimination would have more success. Using this criterion, it is desirable to eliminate such theories as witchcraft and ether, but it is not desirable to get rid of classical mechanics or folk psychology. Schwartz argues in a similar vein when he contends “we may not need [classical mechanics] … We nonetheless have [classical mechanics] whether we need it or not” (Schwartz, 1991, p. 208). Without any reason to eliminate the partially incorrect theory, the threat of elimination becomes “an idle bugaboo” (Schwartz, 1991, p. 219), for both theories can be used. But then, who decides when a theory should only be reduced, rather than eliminated? Seemingly a witch or a psychic would use the instrumental criterion to refrain from eliminating their beliefs. And more importantly, an

eliminativist will use the instrumental criterion to eliminate the mental, while a reductionist will use the instrumental criterion to conservatively reduce the mental.

We can summarize our results as follows: Kim's reductionism solves the problem of explanatory exclusion by maintaining that mental and physical descriptions state the same objective causal relation between events. However, since there is one causal relation between events, and this one causal relation between events has a complete physical description, we can exclude any additional mental descriptions of this same event. As Louise Antony notes, "but if mental events just are physical events, then there are no specifically psychological properties at work, and no need for - indeed no possibility of - a specifically psychological taxonomy or science" (Antony, 2007, p. 145). So, if we want to keep mental descriptions around, and presumably we do, we will not be able to endorse Kim's solution to the problem of explanatory exclusion.

In this chapter I considered Kim's reductionism, which circumvents the apparently inconsistent tetrad by rejecting irreducibility. As it turns out, however, Kim's reductionism does not offer us autonomous mental causation, nor does it offer us reduced mental causation. Kim's reductionism also renders mental descriptions unnecessary. This gives us reason to search out a new solution to the problem of explanatory exclusion, and it also gives us another reason to prefer autonomous mental causation to reduced mental causation. Namely, it was reduced mental causation that posited one causal relation between events, which in turn rendered mental descriptions unnecessary. If we have two causes, as we do on autonomous mental causation, then we can have a physical description of the physical cause, and a mental description of the mental cause.

## **Chapter 6 – Causal Exclusion and Kimian Events**

To this point we have seen three failed attempts at resolving the problem of causal exclusion. We have also seen that we cannot reject the principle of causal exclusion or the principle of irreducibility. For the reasons outlined in chapter two, we cannot reject the principle of mental causation or physical causal completeness either. There are, however, at least two metaphysical responses to the problem of causal exclusion that we should consider. The first of these, rooted in the role that Kim's model of events plays in generating the problem of causal exclusion, will be discussed here.

This chapter is divided into three sections. In Section 6.1, I outline Kim's property exemplification model of events. In Section 6.2, I consider the Davidsonian solution to the problem of causal exclusion which relies on a different model of events than Kim endorses. I conclude that Davidson's solution may work, but it relies on a model of events Kim will not endorse, and it requires us to accept nominalism about properties which diminishes the amount of realism we can achieve in two ways. In Section 6.3, I return to Kim's solution to the problem of explanatory exclusion. I consider an objection to Kim's extensional solution to the problem of explanatory exclusion, and a proposed response rooted in the property exemplification model. I conclude that certain other features of the property exemplification model disable this response, so we are left with further reason to find a different solution to the problem of explanatory exclusion.

### **6.1 – The Property Exemplification Model of Events**

Kim sets forth his model of events in a number of papers (Kim, 1969; Kim, 1973; Kim, 1976). There he argues that “an event is a structure consisting of a substance, a property, and a time” (Kim, 1993, p. 34). More specifically, an event is the instantiation of a

property by an object at a time.<sup>32</sup> Objects are involved because there has to be something that undergoes events or changes. Times are involved because events are occurrences

---

<sup>32</sup> It is worthwhile to notice that an individual event is “constituted” (Kim, 1991, p. 643) from at least three parts: the constitutive object, the constitutive property, and the constitutive time – as well as any other extrinsic properties the event has. Events, therefore, turn out to be “compounds” (Kim, 1993, p. 11) or “complexes” (Kim, 1993, p. 8). Importantly, this does not mean that an event is identical to these parts. As Kim says “the account so far presented is not an ‘eliminative’ or ‘reductive’ theory of events; that is, it does not attempt to show that events are in some eliminative sense ‘reducible’ to substances, properties, and times” (Kim, 1993, p. 36; Thalberg, 1980; Feldman and Wierenga, 1979). This raises a difficult question: are the micro-parts of the event sufficient in themselves to make an effect occur? If so, and if an event is not identical to these micro-parts, then don’t we need to exclude events from causal efficacy? According to causal exclusion an event is supposed to be a sufficient cause for a given effect, but it seems like the instantiation of the property, the object, and the time, is sufficient, so the event lacks causal efficacy. There are at least two solutions to this problem. First, we can assume that Kim comes to endorse the reductive identification of events with their constituents, thereby securing causal efficacy for events. Trenton Merricks supposes that Kim follows this path: “suppose that the causal powers of an event supervene upon the powers of its constituent substance and property; by Kim’s principles, then, this event does not exist over and above the substance and property” (Merricks, 1995, p. 161). Although Kim does turn towards reductionism in other cases of supervenience, he has not explicitly done so in this particular case. Moreover, he provides reason to refrain from the identity: “notice that  $[(x, t), P]$  is not the ordered triple …  $x$ ,  $t$ , and  $P$ ; the triple exists if  $x$ ,  $t$ , and  $P$  exist; the event  $[(x, t), P]$  exists only if  $x$  has  $P$  at  $t$ ” (Kim, 1993, p. 9). The event is not identical to its three parts because they have different conditions for existence. The three parts could exist in isolation without the event existing, so the event is not identical to these parts. This argument still remains, so it is not likely that Kim takes this route. Alternatively, Kim frequently suggests that events are “structured” complexes (Kim, 1993, p. 34; Kim, 1993, p. 8). He even goes so far as to state that “an event is a structure consisting of a substance, a property and a time” (Kim, 1993, p. 34). This structure binds all three of these parts together, so all of the parts are included within one event. Accordingly, Kim says that an event is “the exemplification of property  $P$  by an object  $x$  at time  $t$ ” (Kim, 1993, p. 8; Kim, 1993, p. 34). And, as Kim hints at, as the words ‘by’ and ‘at’ knit the words ‘property’, ‘object’ and ‘time’ together, so the structure of an event knits these parts into one event. Indeed, Thalberg uses Kim’s statement that events have a similar structure to language to mean that Kim has something like this in mind (Thalberg, 1978, p. 7-8). Kim says “events … have something like a propositional structure” (Kim, 1973, p. 8). These considerations are important for a number of reasons. First of all, they furnish us

and occurrences happen at specific times. Properties are involved because changes imply that an object has gained or lost a property (Lombard, 1985, p. 111; Kim, 1976, p. 33). Since these three parts constitute or make up an event, Kim calls these the constitutive object, constitutive property and constitutive time of an event (Kim, 1973, p. 9).

Although the notion of a constitutive object and constitutive time deserves analysis, we need only concern ourselves with the constitutive property here. According to this model every event has one constitutive property, which is a causally relevant property picked out “relative to [a] theory, in terms of which lawful regularities can be discovered” (Kim, 1976, p. 37). Since theories offer us generalizations, it is not surprising to see that the instantiation of a constitutive property will generate a repeatable event which Kim calls a generic event. In similar circumstances in other times and other places, this same constitutive property will be instantiated, and the same type of event will occur.

Events have further properties themselves, but these are not constitutive properties, and this is so for two reasons. First, they are not properties of the object which generated the generic event, but properties of the event itself (Kim, 1976, p. 43). Relatedly, they are extrinsic properties of the event, as opposed to the intrinsic nature of the constitutive property itself (Kim, 1976, p. 42). This means that they involve external factors and that they need not have been instantiated in order for the event to occur.

---

with an example of, and a reason for, how the structure of parts is not reducible to the parts – a lesson Kim will abandon in his mereology (Section 8.3). Secondly, we can see how structure plays an important role. The three parts, taken in isolation, may not form an event, and will not have the same causal power as the event. Hence the appropriate structuring of these parts in a certain way is necessary. Finally, it demonstrates how Kim’s model of events is in jeopardy. Kim’s model of events suggests that an event is not reduced to its parts, but Kim’s causal exclusion argument suggests that since the parts are sufficient the event can be excluded. I argue that this tension is resolved in the same way that the broader causal exclusion problem is resolved. Namely, we take an event to be the structure of the three parts, but this structure plays the role mentioned, while this structure is not reducible to the parts.

Consider, to use Kim's example, the event 'Brutus stabbed Caesar in 79 AD'. The stabbing is a property of Brutus, and as such is the constitutive property for the event. The stabbing is also intrinsic, so if 'being a stabbing' had not been instantiated, the event would not have occurred.<sup>33</sup> This event also has the property of 'occurring outside of Rome', but this is not a constitutive property. This is so because it is not a property of Brutus, nor is it intrinsic to the nature of the event for it mentions external factors (i.e., Rome), and if it had not occurred outside of Rome, the stabbing could have still occurred.

Kim calls this the property exemplification model of events, and it comes with the following conditions for event existence and event identity:

*Existence Condition:* Event  $[x, P, t]$  exists just in case substance  $x$  has property  $P$  at time  $t$ .

*Identity Condition:*  $[x, P, t] = [y, Q, t^*]$  just in case  $x = y$ ,  $P = Q$ , and  $t = t^*$  (Kim, 1976, p. 35)

It is worthwhile to note that the existence condition does not state that the event  $[x, P, t]$  exists if the triple  $x, P$  and  $t$  exists (Kim , 1973, p. 9). Rather, only the specific structure of  $x$  having  $P$  at  $t$  guarantees that the event exists.<sup>34</sup>

---

<sup>33</sup> It is possible that 'being a stabbing' must be extrinsic and relational, since there must be another person present in order for there to be a stabbing. Possibly, Kim has in mind the stabbing motion, where it does not matter if a person or merely the wind is stabbed at, it is still a stabbing motion.

<sup>34</sup> Of note as well is the identity condition  $P = Q$  requires an analysis of the circumstances under which we can conclude that two potentially identical properties are in fact only one. For Kim, properties are identical if they have the same causes and effects (Kim, 1998, p. 105). 'Being a stabbing' has different causes and effects than 'being a killing', for not all stabbings cause the heart to stop beating, but all killings do (Kim, 1976, p. 42). So, we can conclude that 'being a stabbing' is a different property than 'being a killing'. Katz also suggests that co-extensivity is a requisite for Kimian property identity (Katz, 1976, p. 437). Since stabbings occur in places that killings do not, we can conclude that these properties are not identical (Kim, 1976, p. 36-37). We should be careful here, however, for Kim appears to favour the former criterion in his later works.

## 6.2 – Davidsonian Events and Causal Exclusion

Kimian events, as outlined here, are typically contrasted with Davidsonian events. We have already alluded to Davidsonian events on a number of occasions, but we should consider them in greater detail, as they offer a possible response to Kim's exclusion argument. To get us into the nature of Davidsonian events, consider Davidson's doctrine of anomalous monism. Donald Davidson argues that the doctrine of anomalous monism arises naturally out of the combination of the following three tenets:

- (1) Mental events are causally related to physical events
- (2) Singular causal relations are backed by strict laws
- (3) There are no strict psycho-physical laws (Davidson, 1970, p. 132; Davidson, 1993, p. 187)

At first glance these premises seem to form an inconsistent triad. If causal relations are nomological, and there are no laws between the mental and the physical, how could mental events causally interact with physical events? If there are no strict laws between the mental and the physical, and the mental causally interacts with the physical, then how can causal relations be backed by strict laws? Finally, if mental events causally interact with physical events and causal interaction is nomological, how can there be no strict psycho-physical laws?

Davidson, however, fineses through the seeming inconsistency to generate a token identity theory between the mental and the physical. Suppose, he argues, that a mental event causes a physical event. Premise two contends that this interaction will have to be describable in a way that is amenable to the formulation of a strict law. This law, however, must be strictly physical, since premise three contends that a strict law cannot be found between the mental and the physical. Thus the original description of the mental event causing the physical event will also have a description in law-like physical terms. Since there is both a mental description and an underlying physical description of the same event, the identity between the two is forged (Davidson, 1970, p. 133; Davidson, 1980, p. 231). Due to premise three, however, the mental description

cannot collapse into the physical description. Thus an identity between mental and physical types cannot be generated. The resulting picture is one in which only a token identity between mental and physical events is possible.

Numerous critics question the viability of Anomalous Monism. The most common concern is that the mental as mental does not seem to have causal powers. How exactly would this consequence come about? Honderich explains:

The mental event as physical causes the action. To give this answer is of course to cast a new light on the first claim [of Davidson's Anomalous Monism], that the mental interacts causally with the physical. It becomes the claim that the mental as physical interacts causally with the purely physical. What is important, however, is that the resulting picture seems not to account for a conviction that lies behind acceptance of his first claim when it is naturally understood, as the claim that the mental as mental causes the physical. This is the conviction of the efficacy of the mental, already mentioned (Honderich, Pg. 159. See also Kim, 1984, p. 267; Sosa, 1984, p. 277; Kim, 1993b, p. 21).

Since premise two states that causal relations are law-like, and premise three claims that law-like relations could only be applicable to physical events, the mental event as a mental event can never be causally potent. At best, it is the mental event subsumed under the description of being a physical event that has causal power. Numerous critics find this conclusion inescapable, and therefore charge anomalous monism with epiphenomenalism. Moreover, if this charge is true, premise one is false, which in turn renders anomalous monism internally inconsistent.

Davidson responds to his various critics in a paper entitled 'Thinking Causes' by introducing the Extension Reply. He acknowledges that many of his critics contest that Anomalous Monism leads to epiphenomenalism: "But the complaints have most often been summed up by saying that anomalous monism makes the mental causally inert" (Davidson, 1993, p. 187). Rather than conceding defeat, however, he thinks that his

critics come to their conclusions because they fail to appreciate his view that causation is an extensional relation between two events. As Davidson explains:

If causality is a relation between events, it holds between them no matter how they are described ... but we cannot say that an event caused another only as described. Re-describing an event cannot change what it causes, or change the event's causal efficacy. Events, unlike agents, do not care how what they cause is described (Davidson, 1993, p. 191. See also Evnine, 1991, p. 175).

Davidson is here capturing the intuition that causes between events in the world will do what they do whether or not humans talk about them. Whether one describes the temperature in Celsius or Fahrenheit, it will still feel the same on the skin. For this reason, it makes no sense to say that under a mental description the cause will not occur, but under a physical description it will occur. Causation, as an extensional relation between events, will happen no matter what humans say about it. As Davidson summarizes, "Nature in its causal doings is indifferent to our supply of concepts" (Davidson, 2004, p. 142; See also Davidson, 1993, p. 198; Levin, 1977).

By using this model of events, Davidson overcomes the exclusion problem from within a nonreductive framework. The exclusion problem states that there can be no more than one sufficient cause for any given event. Davidson posits an ontological identity, so there is only one sufficient cause. However, this event can be described using mental and physical vocabulary. Kim, like many others, thinks that Davidson's solution fails because it does not give us mental causation in virtue of the mental property. But Davidson responds by stating that events cause as events, not in virtue of their properties, so we can apply a physical and a mental description to the event without difficulty (Crane 1995: p. 226–9; Gibb, 2006; Campbell, 2008, p. 63; Marras and Ylu-Vakuri, 2008, p. 116ff).

Should we conclude that Davidson's solution successfully avoids the exclusion problem? To answer this question, first of all notice that Davidson's solution reveals his nominalism about properties. It is not as though nature causes things in virtue of

language independent properties that can be apprehended and described in terms of corresponding predicates. Rather, nature causes things, and humans describe it how they wish. This means that humans may describe events by use of certain predicates, but this does not mean that there are language independent properties corresponding to these predicates. He acknowledges as much in places. For example, he states “people have mental properties, which is to say certain psychological predicates are true of them” (Davidson, 1996, p. 231). Here we see that Davidson uses the term ‘property’ to denote predication, he does not have something language independent in mind.

Davidson’s solution to the problem of mental causation relies upon this nominalism as well. He argues that properties do not exist as language independent entities, so events can hardly involve property exemplifications. If events do not involve property exemplification, then they can be coarse and unrepeatable particulars. Since they are bare particulars, the event will cause as the event itself. Since the event causes as the event itself, we cannot complain that Davidson loses mental causation because the physical property does all of the work.

In order to accept Davidson’s solution, then, we will need to endorse nominalism about properties. Should we do so? For the bulk of the twentieth century the logical positivist movement, and the ensuing linguistic turn, rendered the existence of properties dubitable. More recently, to borrow a term from Martin and Heil, we have seen a sort of ontological turn (Martin and Heil, 1999). Part of this ontological turn includes a turn away from nominalism about properties towards adopting an ontology that includes properties (Strawson, 1985; McGinn, 1989, p. 13; Poland, 1994, p. 10-44).<sup>35</sup> Chris Swoyer summarizes the sentiment: “thirty years ago the reputation of properties had hit

---

<sup>35</sup> This ontological turn is not solely isolated to the property/predicate debate. For example, within the scientific realism/anti-realism debate Arthur Fine once argued that “realism is well and truly dead” (Fine, 1986, p. 112). More recently, however, Stathis Psillos lists over a dozen recent books or papers advocating scientific realism (Psillos, 2000, p. 705).

rock bottom. Their stock has risen sharply since, however, and although nominalists remain, most philosophers can now invoke properties without guilt or anxiety" (Swoyer, 1996, p. 243).

There are a number of reasons for this turn towards properties. First of all, within the mental causation debate, a number of authors have suggested that events must cause "qua," or, "in virtue of" a property of the event. Terence Horgan calls this the quausal model (Horgan, 1989), and he cites Fred Dretske and Ernest Sosa as examples of people who use the quausal model. To use Dretske's example, a meaningful sound causes some glass to break. Is it in virtue of the high pitch of the sound, or is it in virtue of the meaning of the word, that the glass breaks (Dretske, 1989)? It is coherent, at least, and more precise and explanatorily complete, at best, to think the event causes not only as the event, but in virtue of a certain property of the event as well.

Secondly, Davidson's nominalism implies a rejection of the traditionally conceived correspondence theory of truth; and this move diminishes the amount of realism that is achievable. Briefly, the correspondence theory suggests that true statements correspond to objective facts. Within a Kimian framework, as demonstrated below, these facts are complex events involving an object, a property and a time, such as the electron's negative charge at noon. Thus, for my purposes here, true statements correspond to complex events. For example, 'the electron's negative charge at noon' is true if there is an event of the electron's negative charge at noon. On this model, language independent properties such as 'having negative charge' play an integral role, since predicates correspond to, or refer to, these properties in an important sense (Boghossian, 1990, p. 161; Swoyer, 1996; Lewis, 1983). Thus, in the words of David Mellor, "the simple predicates we use in our law statements ... correspond to properties" (Mellor, 1999, p. 36; See also Armstrong, 1997, p. 164-166).

If Davidson's nominalism is correct, then language independent properties do not exist and there can be no correspondence between predicates and these properties. If 'negative charge' is merely linguistic, then it, of course, does not refer to a language

independent property. This nominalistic model cannot achieve as much realism as the correspondence model can. On property nominalism, the electron exists independently of language, but its negative charge does not exist independently of language. On the contrary, on the correspondence theory, the electron exists independently of language, and so does negative charge as a chemical property, so the electron's negative charge exists independently of language. This approach enables our sentences to track, or mirror, reality very closely. Not only do we know that an electron exists independently of conceptualization, but we know that a negatively charged electron exists independently of conceptualization. This enables our language to scrape closer to reality and our understanding of reality to be more precise. As Patricia Marino explains:

The nature of predicate denotation plays an important role in debates over realism and anti-realism. Some such debates concern the question of whether the predicates in the given domain denote real properties. If they do not, from the correspondence point of view, the statements in question cannot be taken at face value or taken to be true (Marino, 2008, p. 88. See also Mellor, 1997, p. 266).

If objects and events do not have mind independent properties, then the world is not finely structured, and our statements are mere conceptualizations of objects. If, on the contrary, objects and events do have mind independent properties, then the world is finely structured and our statements are able to capture and articulate the nature of these objects in intricate detail.

With these things considered, should we endorse Davidson's solution to the problem of causal exclusion? There are three reasons to prefer not to turn towards a Davidsonian solution. As noted, Davidson's solution to the problem of causal exclusion relies on his nominalism about properties. This nominalism about properties seems to diminish the amount of realism we can achieve in two key ways. Properties enable us to achieve a more detailed explanatory analysis of how and why the causal relation between events occurred. Properties also provide a foundation upon which our language can closely describe reality. Davidson's solution does not offer us autonomous mental

causation either. Davidson posits a distinction between mental and physical vocabulary, but he argues that the mental cause is not distinct from the physical cause, so the mental cause cannot play a distinct role. We should also note that Kim endorses the existence of properties. To solve the causal exclusion problem by simply rejecting the background assumptions that Kim uses to create the problem would hardly convince Kim.

This last point works in the opposite direction as well. Kim's insistence that there is a danger of mental property exclusion will hardly convince a Davidsonian either. For the Davidsonian, the mental event is the physical event, and this guarantees mental causation in a similar way that Kim does. Kim's objection that mental properties of events are causally inert because the physical property can do all of the work will not be persuasive since neither mental properties nor physical properties exist. We can conclude, then, that if we are willing to sacrifice some realism and explanatory precision, and if we are content with reduced mental causation, we may wish to appeal to a Davidsonian solution. However, with a desire for both realism and autonomous mental causation, and with a desire to confront Kim on his own terms, we will assume that this move is off the table.

### **6.3 - Explanatory Exclusion, Events and Descriptions**

In Section 5.1 we saw Kim solve the problem of explanatory exclusion by endorsing an extensional model of explanatory individuation. According to this move, mental and physical descriptions both refer to the same objective relation between events, so they state the same explanation. Now that we have a deeper understanding of Kim's view of events, I want to consider whether this solution is viable or not.

#### ***6.3.a – Problems with Extensional Individuation***

A number of criticisms have been leveled against the extensional model of explanatory individuation. These criticisms largely revolve around the strange consequences that arise from two conceptually distinct descriptions stating the same explanation. Ausonio

Marras, for example, argues that explanations are supposed to provide some sort of epistemic relief – a principle which Kim agrees with (Kim, 1988, p. 225). He then asks us to consider the following two causal claims: ‘The earthquake caused the collapse of the building’ and ‘the event that caused the collapse of the building caused the collapse of the building’ (Marras, 1998, p. 443). While the first explanation helps us understand what happened, the second explanation is epistemically vacuous. However, on Kim’s model, to have the second explanation is to have the first explanation. The second description refers to the objective causal relation between the mentioned events, so it is the same explanation as the first one. Marras recoils: “anyone in possession of the one explanation is thereby in possession of the other. And this, I believe, is counter-intuitive” (Marras, 1998, p. 443; See also Campbell, 2008, p. 84-85; Campbell and Moore, 2009). Anyone who has the second explanation does not necessarily know that the earthquake caused the collapse of the building, so we have reason to believe that these two explanations remain distinct.

Andre Fuhrmann joins Marras in rejecting Kim’s model. He considers the case of a baby who is crying because (i) the water in the tub is too hot, and/or because (ii) the mean kinetic energy of the H<sub>2</sub>O molecules in the tub has passed a certain threshold value (Fuhrmann, 2002, p. 185). He points out that the objective relation between events is the same, but this does not mean that the two explanations are “interchangeable” (Fuhrmann, 2002, p. 185). Rather, “for those who do not know of the relation between heat and molecular movement, that is to say, for most people, (ii) would not even pass as an explanation at all” (Fuhrmann, 2002, p. 185). On Kim’s model, these descriptions both state the same explanation, regardless of the fact that one is more scientifically advanced, and the other provides more epistemic gain to ordinary folk.

Finally, Sophie Gibb provides a further paradox resulting from the suggestion that two intensionally distinct explanations are the same. She says that explanations are a relationship between statements, a supposition that Kim endorses at times (Kim, 1989, p.92-94; Kim, 1988, p. 225-226; Kim, 1994, p. 52-54). But the statement ‘the earthquake

occurred' is not the same as 'the event reported on p. 7 of the Times occurred', for they are not synonymous. However, according to Kim's model, 'the earthquake caused the building to collapse' is the same explanation as 'the event reported on p. 7 of the Times caused the building to collapse'. We have a case where the statements are different, but the explanations involving the statements are the same. Gibb protests: "surely for the two explanations to be identical, it is necessary that [the statements are identical]" (Gibb, 2009, p. 12).

There is a common thread that winds through all of these objections. While we have reason to believe two different descriptions do not state the same explanation, the extensional model insists that these two different descriptions do state the same explanation. According to Marras, Kim's approach implies that an epistemically fruitful description is the same explanation as an epistemically vacuous description. According to Fuhrmann, it implies that a scientifically precise description is the same as a folk description. And Gibb claims that Kim's view entails that descriptions involving different statements state the same explanation.

### ***6.3.b – The Constitutive Property Reply***

At this point, Kim is faced with a choice: either abandon the extensional model of explanatory individuation or find a way to overcome these deficits from within the extensional framework. All of the critics mentioned above suggest the former route. However, this approach introduces an unpalatable dilemma for Kim. Kim either has to exclude psychological explanations or reject the principle of explanatory exclusion. Recall that Kim's acceptance of the extensional model allows two conceptually different descriptions to state the same explanation, thereby dodging exclusion pressures. If Kim abandons the extensional model, psychological explanations will remain distinct from neurological explanations, and they will have to be excluded. In order to avoid this consequence, Kim has to significantly alter or abandon the principle of explanatory exclusion. Marras and Gibb suggest that he alter the principle of explanatory exclusion

(Marras, 1998, p. 449; Gibb, 2009, p. 16), while Campbell and Fuhrmann suggest that it can be abandoned (Campbell, 2008, p 89; Fuhrmann, 2002, p. 184). However, Kim has shown no indications of modifying or abandoning this principle. Rather, he continues to talk about the same principle as recently as his latest book (Kim, 2005, p. 17).

Perhaps there is a solution from within an extensional framework that Kim has in mind? Kim's property exemplification model of events presents one such solution. Consider a slightly revised version of our previous example: the hurricane's being windy caused the building to collapse. Now consider these two descriptions for the first event: 'the hurricane's windiness (at  $t$ )' and 'the event that caused the building to collapse'. According to the property exemplification model, 'windiness' is the constitutive property, for it is a property of the object and it is intrinsic to the nature of the event. On the contrary, 'causing buildings to collapse' cannot be the constitutive property, for it is not a property of the hurricane, nor is it intrinsic (the hurricane's windiness could have still occurred if no building collapsed, and it mentions the external building). Kim can therefore respond by saying that 'being the event that caused the building to collapse' does refer to the same event as 'the hurricane's windiness (at  $t$ )', but 'being the event that caused the building to collapse' does not refer to the constitutive property of the event, whereas 'the hurricane's windiness (at  $t$ )' does. This being the case, it is not true that both descriptions refer to exactly the same thing, and since explanations are individuated by what they refer to, these can no longer be considered the same explanation. And, if these two different descriptions are not the same explanation anymore, then the aforementioned problems with two different descriptions stating the same explanation vanish. At least one critic (Campbell, 2010) thinks this response may have a limited amount of success.

### ***6.3.c – Problems with the Constitutive Property Reply***

If Kim's extensional model faces the problem that two different descriptions stated the same explanation, Kim can resolve this problem by arguing that these two descriptions

refer to different properties of the event, so they do not state the same explanation after all. But now we have to worry about whether this solution is too successful. For, psychological descriptions are different from neurological descriptions, so isn't it also the case that they both refer to different properties of the event as well? If so, then the result is that these descriptions are different once again, and the psychological description faces renewed exclusion pressures. Kim clearly does not want this, so he is going to have to say that the constitutive property of an event can be re-described.

Indeed, Kim does leave some room within his theory of events for a property to be re-described:

It is not part of the account in question that the use of different predicates – nonsynonymous, logically inequivalent predicates – invariably leads to a multiplicity of properties. ‘Is blue’ and ‘has the color of the sky’ pick out the same property, namely the color blue (Kim, 1976, p. 43).

The description ‘is blue’ refers to blue, while ‘has the color of the sky’ also refers to blue. So long as two descriptions refer to the same property, that one property can be instantiated as the constitutive property of an object, and hence we will be able to describe the constitutive property of an event in more than one way. Thus, we can still say that the psychological and neurological descriptions both refer to the constitutive property, so they are the same, and no exclusion pressure arises.

There are a number of reasons to think this solution will not work. First of all, a number of critics think the property exemplification model implies that any re-description will result in reference to a new event (Hedman, 1972; Katz, 1976; Bennett, 1988, p. 73; Evnine, 1991, p. 31). To see why, consider Jaegwon Kim’s discussion about events with Donald Davidson. According to Davidson’s model, one event can have numerous descriptions. To use an example found in the literature, Brutus’ stabbing Caesar is the same event as Brutus’ killing Caesar (Davidson, 1969, p. 170). Even though they are conceptually inequivalent descriptions, they both refer to the same event. Kim objects to this Davidsonian conclusion. According to Kim, two events are identical if they have the

same constitutive object, the same constitutive property and occur at the same constitutive time (Kim, 1969, p. 35). Since the property of stabbing Caesar is distinct from the property of killing Caesar, there are two events here, not one (Kim, 1966, p. 232; Kim, 1976, p. 42). In this way, any alteration to a description of an event threatens to pick out a different constitutive property, so any re-description may result in a different event.

In his response to these critics, Jaegwon Kim is somewhat sympathetic to the conclusions that they draw:

It is true that if an event description is altered so that a different generic event is picked out, then the resulting description, on my view, would pick out a different event. That much is clear enough. And the same applies to the names and descriptions of the constitutive objects and times of events (Kim, 1976, p. 43).

In general, any re-description will pick out a different constitutive event, and as a result there will be a different event. Kim introduces a caveat to this rule, which I have cited above. There he says that we may be able to use different predicates to describe the same property, thereby allowing for some limited re-description. However, he closes by saying that we can re-describe an event, but “what cannot be done is to re-describe [an event] by tampering with their constitutive properties” (Kim, 1976, p. 43). It is not clear what constitutes tampering with the constitutive property. Does this mean that we can never re-describe it, or does it simply mean that we can re-describe it so long as the re-description does not pick out a new constitutive property?

Additional elements of the property exemplification model provide five reasons to think we should not allow for any re-description. First of all, in “Events and their Descriptions” Kim wonders whether the conceptually inequivalent descriptions of ‘Socrates died’ and ‘Xantippe’s husband died’ can describe the same event (Kim, 1969, p. 202). He concludes that “serious difficulties arise in constructing a coherent general account of event-description in this direction” (Kim, 1969, p. 205). Two relevant problems with substituting nonequivalent statements for the same event are the “obvious

explanatory, and perhaps also causal, asymmetry” (Kim, 1969, p. 211; See also Kim, 1976, p.45) that ensues.

Let us consider the explanatory difficulties first. Consider the following explanation: ‘Whenever the blue screen appears, the student panics, the blue screen appeared, so the student panicked’. The two events described in this way offer a rich explanatory harvest. If we can re-describe one event, however, then we can also concoct the following explanation: ‘Whenever the blue screen appears, the student panics, the screen turned the colour of the sky, so the student panicked’. Here we have no immediate explanation for why the student panicked. We would have to know that blue is the colour of the sky, but we may not know this, or it might be night out at the moment. Since “we are interested in events primarily insofar as they are objects of explanation” (Kim, 1969, p. 213), we conclude that any description that does not provide explanatory insight should not be applied to the constitutive property of the event.<sup>36</sup>

In “Causation, Nomic Subsumption, and the Concept of an Event”, the aforementioned causal benefits of limiting re-descriptions of a given event begin to surface. Kim thinks that his model can account for the requirement that causation involves constant conjunction (Kim, 1973, p. 12). How does this work? According to

---

<sup>36</sup> Kim closes this paper as follows: “We are interested in events primarily insofar as they are objects of explanation and relata of causal relation, and it is by no means false or absurd to say that to explain why Brutus stabbed Casear is not the same as explaining why Brutus assassinated Caesar; and a specification of the causal conditions for the event given by ‘Brutus stabbed Caesar’ need not be the same as one for the event given by ‘Brutus assassinated Caesar’” (Kim, 1969, p. 213. See also Kim, 1976, p. 36; Kim, 1976, p. 45). We are interested in studying events for their explanatory fruitfulness, but there is a great deal of explanatory disparity between the descriptions ‘Brutus stabbed Caesar’ and ‘Brutus assassinated Caesar’. The first event does not necessarily explain the preceding event of the inauguration of a new emperor, as the cut may have been small, or through a toe, or not unto death. The latter fully explains the event of the inauguration of the new emperor, for it leaves no doubt that the old emperor is dead. The first event does not necessarily explain the preceding event of Brutus’ plotting, whereas the latter event does explain this prior event (assuming that all assassinations involve plotting), but stabbings do not necessarily do so.

his model, the constitutive property is intrinsic to the nature of the event, so it will be generic and necessarily repeated in similar circumstances. There are, therefore, constant conjunctions between pairs of generic events. As Kim explains:

Under our account, then, if Socrates' drinking hemlock was the cause of his dying, the two generic events, drinking hemlock and dying, must fulfill the requirement of lawlike constant conjunction (Kim, 1976, p. 12).

To return to our example, blue screens are constantly conjoined with students panicking. We can therefore say that the cause of the event ‘the student panicked’ was the event of ‘the screen turned blue’. It is more difficult to say that the cause of ‘the student panicked’ was ‘the screen turned the colour of the sky’, for there is no constant conjunction between panicking and screens that turn the colour of the sky. As mentioned, it might be night out, or more radically, we could be in a different world where the sky is purple or green. These possibilities remaining, the second description should not be directly applied to the constitutive property instantiated in the event.

Recall that Kim distinguishes between constitutive properties of an event and other properties of an event by claiming that a constitutive property is intrinsic, while the other properties are extrinsic. We need to ask whether ‘has the colour of the sky’ is extrinsic or intrinsic in a typical explanation involving blueness. Here we see that this description involves the external factors of the sky, and the colour of the sky. Moreover, we have seen that the sky may not be blue in some cases, so it is a coincidental feature of the event, not intrinsic to the event. Kim says as much himself when discussing the potential identity of ‘blue is the colour of the sky’ in another paper. There he says “at least one of the terms of the identity refers to a property via some particular(s) that stands in a certain definite relation to it” (Kim, 1966, p. 233). ‘The colour of the sky’ refers to particulars that stand in a relation with the actual property, so it is clearly extrinsic. But if ‘having the colour of the sky’ is an extrinsic description, it will be an extrinsic property of any event, and so it should not be applied to constitutive properties.

Of course, these points are relevant to more than the colour of the sky, for Kim attempts to provide a mental and a physical description of one constitutive property as well. But here we see the same problems. With respect to explanation, to say: ‘Whenever a loved one dies, sadness occurs. A loved one died, so a certain neural turbulence occurred’ does not provide much insight. We still do not know why this neural turbulence occurs, so we should not re-describe the constitutive properties of these events in this way. Similarly, with respect to causation, mental causes are multiply realizable, so we will not likely find a strict causal law between sadness and these particular neurons. We will, however, find a strict law between these neurons and certain specific neurons that fired immediately prior to it. We would lose this constant conjunction as soon as we give a mental re-description to these constitutive properties. Finally, Kim’s functionalism leads him to conclude that mental causes are extrinsic. Functional properties rely on their relations to their environment, whereas physical properties are intrinsic (Kim, 1998, p. 103). This being the case, functional descriptions will refer to extrinsic properties of the event, not the intrinsic constitutive physical property.

Not only should we refrain from re-describing the constitutive property for those three reasons, but we have also seen that constitutive properties are picked out by a theory. Presumably, the relevant constitutive property is picked out by some physiological theory. Psychological theory, however, is different from physiological theory. If a constitutive property is picked out by a physiological theory which is different from psychological theory, then it is not picked out by the psychological theory. But, if the psychological theory does not pick out the constitutive property, and constitutive properties must be picked out relative to a theory, it follows that the psychological description will not refer to the constitutive property of the event.

Finally, Jaegwon Kim argues that the property exemplification model provides something of a “structural similarity” or “structural isomorphism” (Kim, 1973, p. 8) between events and sentences. As Kim says, “as property designators we may use

ordinary tensed predicative expressions ...  $[(a,b,c,t),^2$  stands between <sup>1</sup> and <sup>3</sup>]  
corresponds, by the existence condition, to the sentence ‘*b* stands between *a* and *c* at *t*’”  
(Kim, 1973, p. 9). Here we see that the predicative expressions involving ‘*a*’, ‘*b*’, ‘*c*’ and  
‘*t*’ designate and correspond to *a*, *b*, *c* and *t*, respectively. In the same way, Kim uses ‘*x*  
has *p* at *t*’ to describe the event of object *x* having *p* at *t*. There is an isomorphism  
between the statement ‘*x* has *p* at *t*’ and the event of *x* having *p* at *t*. If we used an  
alternate description, we would no longer have this isomorphism. For example, there is  
no isomorphism when we use ‘*H* has *N* at *x*’ to describe *x* having *p* at *t*, so we may not  
permit it.<sup>37</sup>

---

<sup>37</sup> A number of critics find this required correspondence between causes and explanations in Kim’s explanatory realism as well (McIntyre, 2002, p. 95. See also Stueber, 2005, p. 253-254;). Indeed, Kim often mirrors explanations and causes in his writings. For example, he says “*C* is an explanans for *E* in virtue of the fact that *c* bears to *e* some determinate objective relation *R*” (Kim, 1988, p. 226; see also Kim, 1989, p. 105-106). Kim does not say *H* is the explanans for *W* in virtue of the fact that *c* bears to *e* some determinate objective relation. He also insists that “correct explanations are correct *because* they reflect the objectively existent determinative relationship among events ... what makes an explanation *correct* as an explanation is the fact that it mirrors what’s *out there* independently of our cognition” (Kim, 1981, p. 307; emphasis in original. See also Kim, 1988, p. 226; Kim, 1989, p. 94-96). So deep is this assumption that descriptions must accurately mirror the objective causal relation that Kim sometimes fails to mark a distinction between the lower case cause *c* and the upper case explanation *C*: “a causal explanation of *E* in terms of *C* is a ‘correct explanation’ only if *C* is in reality a cause of *E*” (Kim, 1989, p. 94). Not only does this mean that explanations are causes, but it also means that the explanation *C* perfectly mirrors the cause *C*. Similar patterns emerge elsewhere as well. For example, Kim crosses his categories when he says “let us suppose that our explanandum is some object, *x*, having *M* at *t*” (Kim, 2005, p. 111. See also Kim, 1988, p. 233). Kim simply assumes that the explanation will so closely resemble the actual event that he blurs the distinction at times. To this end, Kim even goes so far as to blend the principle of causal exclusion with the principle of explanatory exclusion at times, calling them/it by the title of causal/explanatory exclusion (Kim, 1993, p. 281; Kim, 1993, p. 291; Kim, 1989, p. 44; Kim, 1996, p. 148). Moreover, some have noted the remarkable similarities between the two arguments (Fuhrmann, 2002, p. 182; Gibb, 2009), as though the causal principle is the same as the explanatory principle. These similarities have not been

This structural isomorphism between statements and events has led Nicholas Unwin to argue that Kim requires a strict “one-to-one correspondence between events and facts” (Unwin, 1996, p. 317). Similarly, Jonathan Bennett criticizes Kim’s view of events because it implies that semantics must echo reality too closely (Bennett, 1988, Pg. 73; Bennett, 1991, Pg. 658). In his response, Kim does not deny Bennett’s interpretation, but rather confirms the link between metaphysics and semantics:

A bridge from the metaphysics of events to the semantics of event names is available: an event name of the form ‘the exemplification by *S* of *P* at *T*’ names the event which is the exemplification of the property ‘*P*’ names, by the substance ‘*S*’ names at the time ‘*T*’ names. Why is this the case? Consider the following uninformative but inescapable truism: ‘the exemplification of *P* by *S* at *t*’ names the exemplification of *P* by *S* at *T*. This is an unremarkable principle which can be thought of as a special case of this schema: “...” names ...’ exemplified by such statements as “‘Socrates’ names Socrates”. It is clear that [this] taken together with the property exemplification account of events, implies the semantical claim (Kim, 1991, Pg. 643-644. See also Kim, 1976, Pg. 37-39).

Here we see that descriptions “mirror” (Kim, 1981, p. 307) or “track” (Kim, 1994, p. 68) events. But if one description mirrors the event, and a second description is different from the first description, then this second description will seemingly not mirror the event. This being the case, the property exemplification model seems to privilege the first description.

We have, then, five reasons to suppose that the property exemplification model does not allow for a re-description of the constitutive property. But we have also seen that Kim hints at the possibility of allowing for such a re-description. At the very least,

---

lost on Kim’s critics, as a number of them have assumed that the two arguments are the same (Pereboom, 2002; Block, 2003).

Kim needs to clarify whether his property exemplification model allows for re-description of the constitutive property or not. If it does not, then he loses his solution to the exclusion problem. If it does, then he loses certain explanatory, causal and correspondence benefits. He also needs to revise his conclusion that constitutive properties cannot be extrinsic and that they must be picked out relative to a theory.

Moreover, if the constitutive property can be re-described, he also opens himself up to several further responses from his critics. First of all, now that we are allowed to re-describe the constitutive property, what prevents us from saying that a certain description such as ‘the cause of the collapse of the building’ is in fact a re-description of the constitutive property? Kim cannot argue that this description is extrinsic, or that it is not picked out relative to a particular theory, for these criteria also block psychological descriptions. But if these descriptions do refer to the constitutive property, then all of the original problems return. Namely, we have two different descriptions which state the same explanation once again.

Secondly, perhaps Kim can successfully use the constitutive property reply to retain the distinctness of certain descriptions that do not obviously refer to the constitutive property of the event. But we can easily re-create the problem on a deeper level. Namely, it still seems to be the case that ‘the constitutive property of the event that caused the building to collapse caused the building to collapse’ refers to the same objective causal relation between events as ‘the windiness of the hurricane caused the building to collapse’. Here we cannot say that the first description does not refer to the constitutive property of the event, for it clearly does. So now we have the same problem all over again. Namely, these two descriptions state the same explanation despite the epistemic vacuity of the one, and the intensional differences between the two. We must conclude that Kim’s solution to the problem of explanatory exclusion fails since it relies on a contestable extensional model of explanatory individuation. For this reason, along with the reason discussed in Section 5.4, we will need to seek a new solution to the problem of explanatory exclusion.

In this chapter I outlined Kim's property exemplification model of events. I then considered Davidson's resolution to the problem of causal exclusion by endorsing a coarse-grained model of events. While Davidson's solution may be successful, its nominalism does not provide us with as much realism as we may want, nor does it provide us with autonomous mental causation or deal with Kim on his own terms. For these reasons, we will not be able to solve the problem of causal exclusion by endorsing the Davidsonian model of events. I then spent some time demonstrating how Kim's model of events cannot prevent Kim's extensional solution to the problem of explanatory exclusion from being deficient. This leaves us in need of a new solution to the problem of explanatory exclusion; which will be offered in Section 10.5.

## **Chapter 7 - Causal Exclusion and Causation**

In chapter two we were introduced to a potentially inconsistent tetrad which makes up the problem of causal exclusion. In chapter three we rejected various attempts at resolving this problem. In chapters four and five we rejected the plausibility of abandoning the principles of causal exclusion and irreducibility, whereas a rejection of causal completeness and mental causation were ruled out early in chapter two. In chapter six I suggested that there are two possible metaphysical solutions. Chapter six dealt with the metaphysical solution involving events, but it was not favoured in the end. In this chapter I deal with a metaphysical solution revolving around Jaegwon Kim's model of causation.

Jaegwon Kim's conception of causation in general, and of sufficient causation more specifically, bears directly on the origination of the problem of causal exclusion. After all, the principle of causal exclusion is a principle about causation, and more specifically, is a principle about sufficient causation. Therefore, in this chapter I clarify Kim's background assumptions on causation, and sufficient causation, and consider whether they hold up under scrutiny.

This chapter is divided into four parts. In Section 7.1, I articulate Kim's model of generative causation, which construes causation as a law-like objective relation between events. In Section 7.2, I consider and ultimately reject the move made by some to overcome the causal exclusion problem by endorsing a counterfactual model of causation instead of the generative model. In Section 7.3, I demonstrate that Kim considers an event a sufficient cause despite the fact that certain background conditions necessarily play a role as well. This discovery will then be tied in with results found in Section 4.3 where I argued that a mental state can play a necessary, determinative role while the physical cause is still sufficient. Finally, in Section 7.4, I argue that the distinctness implicit in the nonreductive position implies that mental states and microphysical events may play a role in different ways.

## 7.1 – Kim’s Principles of Causation and the Generative Model

Over the course of many works, Kim discusses six principles of causation, of which he finds some more plausible than others. First of all, Kim mentions and dismisses the Humean view of causation. Hume thinks that our knowledge of causation is merely a matter of constant conjunction. We observe that event  $c$  normally happens immediately before another event  $e$ , so it is probable that the next time  $c$  occurs,  $e$  will follow it. Hume does not say that  $e$  must necessarily follow  $c$  because necessary conjunction “exists only in the mind, not in objects” (Hume, 1888, p. 165). Necessity, however, remains an integral part of causation for Hume (Kim, 1998, p. 229). Since necessity is integral for causation, and necessity does not objectively exist, causation cannot be objective (Kim, 1988, p. 229). As we saw in Section 2.4, and will see again below, Kim thinks that causation is objective, so Hume’s model is dismissed (Kim, 1988, p. 226ff).

Kim also mentions the Hempelian deductive-nomological account of causal explanations. Hempel gives a causal explanation of a given effect by constructing a deductive argument with a law-like generalization as the first premise, and a proposition describing the effect as the second premise (Hempel, 1965). The conclusion of this argument is the causal explanation of the given effect. These causal explanations focus on deduction, derivability and intelligibility. According to Kim, these conceptual characteristics largely ignore objective components: “Hempel’s primary focus in analyzing the structure of explanation is on the logical and conceptual characteristics of statements making up an explanation, not on the events or other entities these statements describe and their interrelations” (Kim, 1988, p. 236). Again, since Kim thinks causality has an objective component, he dismisses this model of causal explanation as well.

Third, in the above discussion involving Hume, we see that causation involves necessary laws. In order for one event to be the cause of another event, we must be able to form a law that relates the two. Kim, at times, endorses this view (Kim, 2009, p. 35 of 49; Kim, 1988, p. 229; Kim, 1973, p. 572). This nomological account of causation has been defended by many, including Donald Davidson (Davidson, 1970). It has, however,

undergone a significant amount of criticism as well (McLaughlin, 1993, p. 40; Pinedo, 2006). For Kim, an immediate question about endorsing this principle arises: Hume argued that laws are necessary, but the objective world does not necessarily do anything, so causality is not part of the objective world. Here Kim agrees that laws are necessary, but he also thinks causality is objective, so he seems forced to conclude that the objective world is law-like. Davidson does not endorse this additional conclusion. Kim, however, in so far as he holds onto the necessity of objective causation, is committed to some such view.

Fourth, Kim spends quite some time discussing the possibility that causation involves counterfactual dependence. The counterfactual account of causation counts any potential cause that can pass a simple test as an actual cause for an effect (Lewis, 1986). The test: if potential cause  $c$  does not occur, would effect  $e$  still occur? If the answer is no, then  $c$  is indeed a cause of  $e$ . Kim's relationship with counterfactual dependence is complex. At times he relies on a counterfactual analysis of causation (see Section 2.1.d). At other times, he rejects the counterfactual analysis of causation completely (Kim, 1998, p.40ff; Kim, 2007, p. 233ff). The bulk of his work demonstrates a rejection of the counterfactual model, for reasons we will see in Section 7.2, so we can conclude that this model is ruled out.

Fifth, Kim thinks in terms of event causation. We have seen that the causal exclusion argument is framed in terms of events. It is a physical event which causes another physical event, and it is a mental event which is in danger of exclusion on account of this fact. As we saw in the last chapter, an event is an occurrence or a happening, so event causation implies some sort of change (Kim, 2006, p. 7). However, as we also saw, the stable object in which the change occurs is also integral in the analysis of the event. This means that an object undergoes some alteration – some property is instantiated in it – and this event in turn causes some future event to occur.

Finally, and most importantly, Kim endorses a model of objective causation. Kim argues that “causality itself [is] an objective feature of reality” (Kim, 1988, Pg. 229). The

world bounces about, and would continue to do so even if no human existed. Thus it is fair to suggest that causality is out in the world, and is happening independently of human conceptualization. Kim calls this position Causal Realism, and contrasts it with causal irrealism, which supposes that causation is a construction of the mind or of some conceptual apparatus (Kim, 1988, p. 229-230). Causal realism supposes that there is a particular cause ( $c$ ) and effect ( $e$ ) which occurs in the world, as well as some objective relation ( $R$ ) which links these two events together (Kim, 1988, p. 226). As already hinted at, Kim's conviction that causality is objective rules out the Humean and Hempelian accounts of causation on account of their irrealism. Of the four remaining principles, let us provisionally remove the counterfactual principle, since we will see the doubt that Kim expresses about it. We can combine the three remaining principles that he seems to fully endorse (objective causation, event causation, nomological causation) and construct a rough view of Kim's theory of causation: causation is a law-like objective relation between events, where these laws are presumably capturable by scientific vocabulary.

A number of people have complained that Kim does not sufficiently clarify, or make explicit, this exact view on causation (Menzies and List, 2010; Walter, 2008, p. 681). In recent years, Kim has articulated his view of causation in greater detail. The resultant picture is called the generative model of causation, or the productive model of causation. As this model is fleshed out, it confirms the basic picture presented above. There are, however, two developments. Specifically, the principles of necessity and counterfactual dependency are to some degree marginalized.

Kim first makes reference to the generative model of causation in *Mind in a Physical World*. There he contrasts “genuine generative and productive causal processes” with “non-causal regularities” (Kim, 1998, p.45). As an example of the difference between these two he considers a moving car and its shadow. A moving car is substantive and has genuine causal powers, while the shadow, though ever-present, does not generate anything. Of course, Kim makes this point in the context of the exclusion argument: physical events are generative in the sense that they pack enough punch to

make future events happen. Supervening mental events, on the contrary, always accompany these physical events as shadows, but they do not produce any future effects themselves – or so he argues.

In two separate discussions with Barry Loewer, Kim fleshes out the generative account of causation further.<sup>38</sup> In these articles, he first of all confirms that he has a generative model of causation in mind (Kim, 2002, p. 675; Kim, 2007, p. 235. See also Kim, 2005, p. 18; Kim, 2005, p. 38; Kim, 2005, p. 47; Kim, 2009, p. 33 of 49). Further, Kim explains that the generative model requires some sort of contact between cause and effect (Kim, 2007, p. 228; Kim, 2009, p. 6 of 49). This contact involves spatio-temporal location (Kim, 2007, p. 235; Kim, 2009, p. 6 of 49). This fits nicely with the view that causation is an objective relation between events.

This does not fit as well with the view that causation is nomological, however. Notice in the above passage, Kim argues that genuine causation is generative, while “only apparent” and “parasitic” causation still figures in “regularities” that are “law-like” (Kim, 1998, p. 45). For example, as a matter of law, whenever the car shadow appears on section one of the road, the shadow will soon appear on section two of the road. Kim does not think that the appearance of the shadow on section one caused the appearance of the shadow on section two however. There is a law connecting the appearance of the

---

<sup>38</sup> Barry Loewer delivers a number of objections to the generative account of causation. Perhaps the most trenchant objection: Barry Loewer responds to Kim by arguing that contemporary microphysics is largely devoid of such causal and generative vocabulary: “But causation as production fits ill with contemporary physics. Russell famously said that causation so understood ‘is a relic of a bygone age, surviving, like the monarchy, only because it is erroneously supposed to do no harm’. What Russell had in mind is that the fundamental laws and facts of physics do not mention causality.” (Loewer, 2002, p. 661). According to Loewer, microphysics does not talk in terms of productive causality, so it can be rejected. Kim concedes that contemporary quantum mechanics does not frequently speak directly of causality (Kim, 2002, p. 675). However, Kim argues that such themes as interference and force seem to imply productive causation (Kim, 2002, p. 676; Kim, 2009, p. 34 of 49).

shadows, but this is not causation because the shadow did not generate the effect itself. More importantly, of course, Kim makes the same point about mental causation. Mental properties, due to their dependence on physical properties, may regularly appear in a certain order, but this still does not make them causes of the relevant effect. Kim explains:

Non-causal nomological regularities can be generated by underlying causal processes, in the way regularities governing medical symptoms are generated by the progress of an underlying pathology. The crucial question, therefore, is whether these mind-body regularities represent not just laws, but causal laws (Kim, 2009, p. 45).

Symptoms nomologically necessitate the development of some pathology, but these symptoms are still not causes of the effect in question.

There is a way to reconcile this negative stance on nomological regularities with Kim's aforementioned endorsement of law-like causal connections between events. Namely, we can think of causation as a sub-set of nomological regularities, where causation involves nomological regularities, but there are more regularities than causal relations. Thus, the relation between the physical cause and the effect will be law-like and causal, whereas the relation between the mental cause and the effect will be law-like, but not causal. This emerges in Kim's analysis of the counterfactual account of causation as well; a fact which we now turn to.

## 7.2 – Counterfactual Causation and Causal Exclusion

A number of nonreductive physicalists suggest that we can avoid the exclusion problem by simply endorsing a counterfactual model of causation (LePore and Loewer, 1987; Horgan, 1989; Baker, 1993, p. 93; Burge, 1993, p. 115; Loewer, 2007, p. 255-256; Walter 2006; Walter, 2008). How does this work? On the counterfactual model we ask whether the effect would still occur if the presumed cause does not occur. If not, then we conclude that this presumed cause is in fact a cause because of this demonstrable

dependency relation. If so, then the occurrence of this event is irrelevant to whether the effect occurs or not, so it is not a cause.

In the case of mental causation, then, we need to ask whether this effect would still occur if the prospective mental cause does not occur. As we saw in Section 3.1, nonreductive physicalism suggests that mental events are determined by their subvening physical bases. In Section 4.2 we saw that this dependency relation is so strict that wherever this physical base occurs, the supervening mental event will occur as well. Therefore, if the supervenient mental cause does not occur, then the subvening physical cause does not occur either, and hence the effect does not occur either. So the counterfactual ‘had the mental cause not occurred, then the effect would not have occurred’ is true.<sup>39</sup> Given the truth of the counterfactual, we can conclude that the mental cause is in fact a cause of the effect. Moreover, the counterfactual ‘had the physical cause not occurred, then the effect would not have occurred’ is also true. This means that the physical cause is a cause as well, so causal completeness can remain true. On the counterfactual model of causation, both the physical cause and the mental cause are actual causes of the effect, so we need not worry about abandoning causal completeness or mental causation.

---

<sup>39</sup> There are certain problems here. For example, the subvening physical cause may be replaced by a physical cause that lacks a supervenient mental cause, and this physical cause does cause the effect, so the effect can still occur without the mental cause. Alternatively, the subvening physical cause may be replaced by a different physical cause that subvenes the same mental cause (due to multiple realizability), and this new physical cause does cause the effect, so the effect is counterfactually dependent on the mental cause, but not on the original physical cause, implying the original physical event is not a cause. For reasons similar to this Bennett (2003) has proposed a ban on replacing physical causes of the same effect on the grounds that the nearest possible world is not a world where the effect still occurs due to a replaced physical cause. If anything, these difficulties provide further reason to reject the counterfactual model of causation.

Is this a viable solution to the problem of causal exclusion? Notice that on Jaegwon Kim's generative account of causation, this will not work. Kim has already marked a distinction between shadows and supervening events that occur with law-like regularity and genuine productive causes. According to Kim, it may be the case that if the shadow had not occurred just before the green light then the car would not have driven through the intersection. This does not mean, however, that the shadow is a cause. The shadow does nothing to generate the movement of the car through the green light, so the shadow is not a cause. Barry Loewer summarizes Jaegwon Kim's use of the generative model of causation (where  $N$  means a neural cause, and  $M$  is a mental cause, and  $E$  is a given effect):

Kim is thinking of causation as a relation in which the cause generates or produces the effect. If both  $N$  and  $M$  are producing  $E$  then they do so independently. So if  $N$  produces  $E$  then  $M$  is dispensable as a producer of  $E$  ... So Kim's argument against overdetermination, as he makes clear, depends on his conceiving of causation on the model of one event producing another (Loewer, 2002, p. 658-659. See also Loewer, 2007, p. 253; Walter, 2008).

On Kim's generative model of causation, the mental cause is excluded because the physical cause can produce all of the punch needed for the effect. On the nonreductive physicalist counterfactual model of causation, the mental cause is included because the relevant counterfactual linking the mental cause to the effect holds true. A number of nonreductive physicalists have suggested that Kim is begging the question since he relies upon a model of causation where the problem occurs rather than the nonreductive physicalist model of causation where the problem does not occur (Loewer, 2002, p. 659; Bontly, 2005; Walter, 2008, p. 682). True enough, the nonreductive physicalist need not accept the generative view of causation, and if they do not, then they can have unreduced mental causation via a counterfactual account.

In his response, Kim acknowledges Loewer's concern that the exclusion principle only holds if we accept the generative view of causation, but he does not directly address

the question begging nature of this line of reasoning. Sven Walter, as a typical nonreductive physicalist, has not been impressed by this (Walter, 2008, p. 682, Walter, 2006, p. 77ff). Rather, Kim spends his time demonstrating why a strictly counterfactual account of causation is futile (Kim, 1998, p. 40ff; Kim, 2007, p. 233ff; Kim, 1973).

Kim's central argument against the counterfactual analysis is that, like in the case of the nomological relation, causation is a proper sub-set of counterfactual dependence (Kim, 1973). There are, for example, logically dependent counterfactuals that do not indicate a causal relation. For example, "If yesterday had not been Monday, today would not be Tuesday" (Kim, 1973, p. 570). This counterfactual is true, but presumably the fact that yesterday was Monday did not cause today to be Tuesday. Moreover, counterfactual dependency includes relations of determination, and some of these are not causal. To use Kim's example once again, "If my sister had not given birth at  $t$ , I would not have become an uncle at  $t'$ " (Kim, 1973, p. 571). There is a dependency relation between these two events, but they are spread out so far across space that it is difficult to imagine how one event that occurs in Florida (i.e., the birth) could cause another event that occurs in Vancouver (i.e., becoming an uncle). As we saw in Section 3.1, the supervenient relation between physical and mental events is another dependency relation that renders the counterfactual true, but is considered determinative, not causal. In this case, if the physical event does not occur, then the mental event truly will not occur, even though the physical event does not cause the mental event to occur. Kim generalizes these results by simply concluding that, "counterfactual dependency is too broad to pin down causal dependency" (Kim, 1973, p. 571).<sup>40</sup> Some counterfactual dependencies are not causal relations, though most causal relations may be discernible through a counterfactual analysis (cases of independent overdetermination being an exception).

---

<sup>40</sup> In later works Kim elucidates several other counterfactual dependency relations that do not seem to be causal, such as the relation that effect 1 has with effect 2 of a common cause, and omissions. See Kim, 2008, p. 236-238.

For Kim, the most egregious counterfactual dependency that holds despite the absence of a causal relation is the one we hinted at above. Namely, a counterfactual dependency still holds between an epiphenomenal mental event and the effect (Kim, 1998, p. 71; Kim, 2007, p. 234). If the counterfactual dependency still holds despite the admittedly non-causal nature of this relation, then how can a counterfactual dependency reveal causality? Not surprisingly, Kim concludes:

Causation as generation, or effective production and determination, is in many ways a stronger relation than mere counterfactual dependence, and it is causation in this sense that is fundamentally involved in the problem of mental causation (Kim, 2005, p. 18).

For their own part, a number of other philosophers agree that the counterfactual account does not separate epiphenomenal events from causal events, so it is not an ideal model of causation (Crane, 1995; McLaughlin, 2005). With these considerations in mind, we can now conclusively remove counterfactual considerations from Kim's view of causation. We are left with the generative model of causation which states that causation involves a law-like objective relation between events.

What should we conclude about endorsing the counterfactual model to overcome the causal exclusion problem? As with the Davidsonian model of events, we cannot convince Kim that the exclusion problem is resolved by simply appealing to a counterfactual model of causation, for Kim has reason to reject this model of causation. On the other hand, nonreductive physicalists will have to respond to these concerns before they can resolve the problem of causal exclusion by endorsing the counterfactual analysis.

Barry Loewer offers a reply to some of these worries (Loewer, 2007, p. 256-257). Without the space to consider all of Loewer's replies, I will simply focus on the pressing issue of how epiphenomenal mental events can be considered causal on the counterfactual model. Loewer responds to this concern by noting that many epiphenomenalists adhere to the nomological necessity of the mental event whenever the

subvening physical event occurs (Loewer, 2007, p. 257). This being the case we can find a world with different laws where the physical event causes the effect without an accompanying mental event. In this world the counterfactual ‘if the mental event had not occurred the effect would not have occurred’ is false, so the mental event is demonstrably non-causal.

To refute Loewer’s suggestion, we simply need to find one epiphenomenalist who holds that the physical cause metaphysically necessitates the mental event, and we can no longer find an instance where the aforementioned counterfactual is false. Since the counterfactual ‘if the mental event had not occurred the effect would not have occurred’ is true, we once again have a case where an epiphenomenalist would have to admit to the mental event being causal on the counterfactual model. Without venturing any further into the debate, we can at least begin to see why a resolution to the causal exclusion problem by endorsing a counterfactual model of causation is not very appealing.

### **7.3 – Causal Sufficiency and Necessarily Determinative Background Conditions**

In this section I offer additional considerations for one of the conclusions reached in Section 4.3. Namely, in Section 4.3 I suggested that mental states may be able to play a necessarily determinative role while the physical event still retains its causal sufficiency. As it turns out, on Kim’s model of causal sufficiency a background condition can play a necessary role without it being included in the sufficient cause. If this is the case, then mental states can operate like a background condition, where they play a necessary determinative role, but we can still consider the physical event a sufficient cause.

Let us begin with the principle of causal sufficiency. Causal sufficiency is a key ingredient in the causal exclusion problem: it is only the presence of a sufficient cause that generates pressure to exclude another cause. If some cause is not sufficient to bring about an effect on its own, then exclusion pressures do not arise at all. Rather, we conclude that this cause is only a partial cause, and it requires additional help to generate the effect.

Kim introduces and then contrasts these two types of situations. First he considers the example of a man who climbs a ladder. In this case the physiological cause is sufficient to generate this effect (Kim, 1989, p. 82). He then considers the example of an automobile accident. In this case the icy road is not sufficient to generate this effect (Kim, 1989, p. 90); rather, the faulty breaks are necessary as well. How do we know when a cause is sufficient or not? Kim turns to a counterfactual analysis of causation to discern the difference (Kim, 1989, p. 91-92). Let us recall the basic idea: would the effect have happened if only the cause under consideration had occurred? If the answer is yes, then this cause is sufficient. If the answer is no, then this cause is partial. In the case of the man climbing the ladder, only the physiological cause is needed to make the effect happen. In the case of the car accident, the effect would not have occurred if there were only icy roads, but no faulty breaks.

Of course, we have seen that Kim also rejects the counterfactual model of causation; an inconsistency which causes the following problem. On the counterfactual model, Kim's example of the physiological cause of the climb is actually not a sufficient cause. Sure enough, if the physiological cause had not occurred, then the man would not have climbed the ladder. But in order to secure sufficiency, we need to be able to say that if only the physiological cause had occurred, then the effect would have still occurred. However, if no one had ever invented ladders, the man would not climb the ladder, no matter what his physiological state. If the asteroid had not narrowly missed the earth one million years ago, thus enabling life to continue to thrive, then the man would not climb the ladder either, for there would be no man to climb a ladder. If the universe operated under a different set of basic laws, then the man would float up to the roof, and so would not climb the ladder. For these, and infinitely many more reasons, it can truly be said that if they had not occurred, then the climbing would not have occurred either. The implication of the counterfactual analysis is that a countless number of background conditions are necessary for the effect to occur, so it is not the case that the physiological cause alone is sufficient.

Of course, Kim has rejected the counterfactual model as well, so it comes as no surprise to see Kim rejecting this analysis. His concern is that on this model we would literally have to include every background influencing factor in the universe. In another paper he expresses the worry as follows:

The realist scheme also yields a more global sense of ‘complete explanation’, one in which a complete explanation of an event specifies its entire causal history in every detail. This is an idealized sense of completeness, and no explanation can be complete in that sense (Kim, 1998, p. 234).

The entire causal background of an event certainly would be a complete and sufficient cause, but Kim argues that this is too broad. For Kim, these dependencies are not causes; rather, they are necessary background conditions which can be bracketed out. They set the stage for the cause to occur, but they are not themselves counted as causes (see Armstrong, 1978, p. 155; Broadbent, 2008; Longworth, 2006, p. 81). His strategy runs as follows: grant that certain background conditions are necessary for the effect to occur, assume that these background conditions remain in place, then find the event that is sufficient to generate the effect. Given the sufficiency of this event, when in these background conditions, we can exclude any other event from making a causal contribution.<sup>41</sup>

---

<sup>41</sup> This move is somewhat counterintuitive. Causal sufficiency, or completeness, has an obvious air of largesse to it. In order to distinguish it from a regular cause, or a partial cause, it must, by very definition, touch on all of the relevant factors. If this event is sufficient, then nothing else should be necessary, but these background conditions are necessary. Kim is right, this does lead to an exorbitantly complex cause. Donald Davidson provides a glimpse of the enormity of this sufficient cause: “If we consider an event that is a ‘full, sufficient’ cause of another event, it must, as Mill pointed out long ago, include everything in the universe preceding the effect that has a causal bearing on it” (Davidson, 1993, p. 15. See also Loewer, 2007, p. 252; Putnam, 1984, p. 8-9). Anything less than this is, by very definition, only a partial cause. Let us, however, leave this problem aside and assume for the sake of the argument that background conditions, though necessary, are not part of the sufficient cause.

There are several important issues to pull out of this discussion. First of all, we clearly need to answer the following question: if certain background conditions are necessary, how can the event alone be sufficient? Kim's answer seems to be that the cause alone is sufficient only after presupposing the necessary background conditions as well. This is an important discovery, for it confirms the general pattern outlined in Section 4.3. In Section 4.3 I argued that the physical cause is not sufficient if the mental event is necessary as well. One way to secure both the sufficiency of the physical cause and the necessity of the mental event was to argue that the mental event is only necessarily present, but is not causal. If the mental event is only necessarily present, but has no causal power, then the physical cause must do all of the causal work, so the physical cause must be sufficient. I rejected this option because it led to epiphenomenalism, but then I suggested that we can deploy a nuanced version of this strategy. Namely, if the mental state plays a necessary, non-causal but determinative role, then the mental state can play a determining role, while the physical cause will still be causally sufficient, since it is forced to do all of the causal work. Here we see a similar pattern arising. Namely, this physical event is sufficient in itself, even though these background conditions play a necessary role as well.

There are at least three instances where the background conditions are necessary, but they are still excluded from the sufficient cause. As mentioned above, the laws of nature and the ancient past are necessary conditions, though they are not parts of the sufficient cause as Kim defines it. Finally, certain stable background conditions are necessary at the moment when the sufficient cause occurs, even though they do not qualify as part of the sufficient cause according to Kim's model. To use a textbook example, a match is struck against a rock and it ignites. What causes the match to ignite? The event of the match striking is the cause. But the presence of oxygen in the room is a necessary stable condition as well. On this example, the oxygen is not only necessarily present, but in all such worlds which operate under the laws of nature witnessed here, the oxygen plays a determinative role as well. The oxygen, despite being necessarily present,

and playing a role in the ignition of the match, is still not part of the sufficient cause, nor does it impede the striking of the match from being the sufficient cause. Or again, the roughness of the rock that the match was struck against is a necessary condition and appears to play a determinative role as well. Presumably, however, the striking of the match is still the sufficient cause.<sup>42</sup> If these background conditions play a necessary role, though they do not detract from the sufficiency of the cause, then it is plausible that mental states can also play a necessary role without detracting from the sufficiency of the physical cause.

#### 7.4 – Causal Heterogeneity

In order to construe mental states as playing a necessary, non-causal determinative role we will need to insist that mental states have an influence in a different way than physical events do. Kim, however, thinks mental properties are identical to physical properties, so it follows that mental properties cause certain effects in the same way that physical properties do. This means that for Kim causation is homogeneous. Wherever causation occurs, it looks the same. This assumption rears its head on a number of occasions. First, and most obviously, the causal exclusion problem is framed in terms of a physical property which is instantiated as a physical event in competition with a mental property

---

<sup>42</sup> Some other stable background conditions which certainly seem to be determinative in some sense include: the presence of the moon, which determines tides to change; the presence of gravity, which determines stones to fall and desks to exert pressure on the floor; magnetic and electric fields, which exert constant force; the density of water, which determines materials to float or sink; the structure of the pipes in one's house, which determines that water will flow into the city sewer system rather than leak out into the living room. We can ask whether these relations of determination are causal. As we have seen, some say they are not causal, but they are merely background conditions. If these are counted as causes, and they are necessary, then we can no longer say that a certain other event itself is causally sufficient for the occurrence of the effect. The reason for this, again, is that if another cause is necessary, the first cause cannot be sufficient alone and by itself. For this reason, Kim needs to argue that these background conditions are not causal, even though they are necessarily present and seem to play a determining role.

which is instantiated as a mental event. In both cases we have a property, an instantiation or an event that occurs at a time, and in an object. Kim conceives of the type of causal contribution that they would both make as the same.

Another example of this assumption occurs in Kim's description of the difficulty of mental causation. He says:

According to property dualism, however, mental phenomena are distinct from neural phenomena, and it becomes a *prima facie* mystery by what mechanisms, or through what intervening links, these supposedly nonphysical phenomena can cause the muscles to contract...if my desire is to cause my legs to move, it must somehow make use of, or ride piggyback on, the causal chain from my motor cortex to the leg muscles (Kim, 2005, p. 153; Kim, 2007, p. 239)

Notice that Kim assumes mental causation involves such things as links and chains. He goes on to talk about mental causation in terms of connections and causal paths as well. This vocabulary harkens back to generative causation, where causation occurs through a chain of production or movement. After making this point, Kim argues that this mental pushing, or "telekenesis" (Kim, 2005, p. 153), is impossible to imagine.

Finally, the assumption of causal homogeneity comes out in Kim's discussion on the nature of mental causation as well. He says:

We care about mental causation because we care about human agency, and agency requires the productive/generative conception of causation ... an agent is someone who, on account of her beliefs, desires, emotions, intentions, and the like, has the capacity to perform actions in the physical world: that is, to cause her limbs and other bodily parts to move in appropriate ways ... these causal processes all involve real connectedness between cause and effect, and the connection is constituted by phenomena such as energy flow and momentum transfer (Kim, 2007, p. 236; See also Kim, 2002, p. 675; Kim, 2009, p. 33 of 49).

When we make a decision, it seems as though we are producing certain bodily movements to occur, in much the same way as physical processes generate certain

effects. As Loewer suspects, Kim thinks of both physical causation and mental causation in terms of “one event producing another” (Loewer, 2007, p. 659).

Causal homogeneity makes sense for a reductive physicalist who thinks that mental properties are physical properties. Of course, if there is an identity between mental and physical properties then the mental cannot play a role in a different way than the physical does. A nonreductive physicalist, however, adheres to the distinctness of the mental and the physical. Not surprisingly, numerous nonreductive physicalists argue that the manner in which the mental plays a role will be distinct from the manner in which the physical makes its contribution. Tim Crane, for example, points to a number of nonreductive physicalists who argue that “if there are mental causes they are not causes ‘in the same way’ that physical causes are” (Crane, 1995, p. 232). He cites Fred Dretske’s use of structuring causes, and Jackson and Pettit’s use of program explanations as examples of nonreductive approaches which assume that mental causation occurs in a distinct way than physical causation does. Similarly, Tyler Burge expresses reservations about “thinking of mental causes on a physical model – as providing an extra ‘bump’ on the effect” as well (Burge, 1993, p. 115. See also Marcus, 2005, p. 22).<sup>43</sup> The nonreductive physicalist, therefore, has every reason to suppose that the mental states can play a different sort of role than physical causes. Kim’s argument to the contrary makes sense if properties are identified, but the nonreductive physicalist holds on to the distinctness of mental and physical properties.

Causal heterogeneity allows us to conceive of the role of the mental differently than the role of the physical. To this end, notice that I have been careful to discriminate between mental states and physical causes. This shift in vocabulary represents the

---

<sup>43</sup> The nonreductive physicalist who shares this view that mental distinctness plausibly implies the distinctness in the way in which mental states will play a role have a straightforward response to Kim’s charge of telekenesis. Namely, the problem with mental pushing is not rooted in the difficulty with endorsing *mental* pushing, but rather in the difficulty with endorsing *mental pushing*.

distinction being drawn here between mental determination and physical causation. On this model, physical causation involves the Kimian doctrine of two events related by a law-like causal relation. Mental states, however, will be construed as enduring structures of objects in which these events occur, and in so being they can be represented as being states, rather than events. I will, of course, flesh this model out more fully in the following chapters.

For now it suffices to briefly summarize the line of reasoning. Causal heterogeneity allows mental states to play a role in a different way than physical events do. This different role allows us to construe mental states as determinative rather than causal. Determinative relations, we have seen, can be necessary, and can play a role, without impinging on the sufficiency of the physical cause. If this is all possible, then we may have the beginnings of a resolution to the problem of causal exclusion. Namely, mental states are necessary and determinative, so we have an important role for the mental. However, they are not causal, so this does not prevent physical causes from remaining the only sufficient cause of a given effect, thereby preserving causal completeness. Or, in other words, recall that Kim's causal exclusion principle says that "no single event can have more than one sufficient *cause occurring* at any given time" (Kim, 2005, p. 42, emphasis added). If, as I have been saying, mental states are not events or occurrences, then we can accept the causal exclusion principle while simultaneously endorsing the influence of these background mental states.

In summary, Jaegwon Kim endorses a generative model of causation which construes causation as a law-like objective relation between events. A number of philosophers have attempted to dodge the exclusion problem by endorsing a counterfactual model of causation. This move will not convince Kim, for he brings forth a number of weighty considerations against the counterfactual model. Rather, these considerations may turn us towards something like a generative model of causation. I then pointed out that Kim's model of sufficiency implies that certain background conditions can be necessarily present and play a determinative role without encroaching

on the causal sufficiency of the physical event. This suggests the following resolution: if mental states are construed as something like a background condition then we can achieve the necessity and the determinative role of the mental without violating causal completeness or the principle of causal exclusion.

## Chapter 8 - Kim on Mereology

At this point we have ruled out various methods of resolving the causal exclusion problem such as supervenience, emergentism and the constitution view. We then ruled out the plausibility of rejecting one of our four fixed assumptions. We have also ruled out certain metaphysical shifts that would help us avoid the causal exclusion problem, such as altering our view on events and causation. On the positive side, I have also suggested that if we can find a necessary non-causal determinative role for mental states, then the causal efficacy can be left in the hands of the micro physical event while the mental states can simultaneously play an important role. I will now spend the next two chapters arguing that if we conceive of mental states as mereological structures of their parts we can successfully find this non-causal determinative role for the mental.

To make this argument I will need to develop a mereological model, and to do this I will analyze the strengths and weaknesses of Kim's mereological system. As an entry point into Kim's mereology, I will consider the generalization problem, which is an argument commonly raised against the principle of causal exclusion. This chapter is divided into four sections. In Section 8.1, I outline the generalization problem and Kim's initial "higher level" attempt at preventing the exclusion argument from leaking out into all of the special science domains. In Section 8.2, I demonstrate why Kim's higher level response is inadequate. In Section 8.3, I track a gradual shift into a second distinct response to the generalization problem. Rather than retaining the layered view of the higher level solution, Kim now collapses the levels into one by saying that "higher level" properties are configurations of micro properties, thereby ensuring the causal efficacy of these "higher level" properties. This leaves us with the question of whether the configuration of micro properties is identical to all of the individual micro properties. In Section 8.4, I provide a number of arguments for why this identification cannot be made. However, the many micro properties are causally sufficient themselves, so this leaves the configuration of micro properties a target for causal exclusion. This leaves Kim in a

difficult bind; a bind which can be alleviated by taking the steps suggested in chapter nine.

### **8.1 – The Generalization Problem and the Higher level Solution**

The generalization problem is one of many possible replies to the causal exclusion argument. It contends that special science properties supervene on underlying physical processes in the same way that mental properties do, so they are in the same position as mental properties (Burge, 1993, p. 102; Baker, 1993, p. 77; Van Gulick, 1992, p. 325). If mental properties are excluded because they supervene on neural properties, then neural properties are excluded because they in turn supervene on atomic properties, and so on and so forth. Typically this response is accompanied by a *reductio ad absurdum* which takes it as obvious that special science properties are causally efficacious, and since mental properties share the same fate as special science properties, mental properties are efficacious as well.

Jaegwon Kim responds to the generalization problem in the aptly titled article “Does the Problem of Mental Causation Generalize?” In it he argues that if mental properties are similar to special science properties, then the generalization worry would not save mental properties, but would rather endanger special science properties as well. There is no need to fear, however, for Kim marks a clear distinction between mental causation and special science causation, so the presumed analogy cannot be maintained. This distinction revolves around the difference between a higher order property and a higher level property. A higher order property, or second-order property, is a property that supervenes on a lower order property, but this lower order property is still on the same level as the higher order property. For example, the highest shelf in a bookcase on the sixth floor of an apartment is higher than the lowest shelf of the same bookcase. However, both shelves are on the same floor, or level, of the apartment. Importantly, the reason higher order properties are on the same level as lower order properties is that they are both properties of the same object: “supervenient mental properties and their

subvenient base properties are properties of *the same entities*. A human is in pain and is in an appropriate neural state" (Kim, 1997, p. 290). Higher order mental properties and lower order neural properties are properties of the same object, namely the human brain.

Conversely, a higher level property is a property of a different object on the mereological scale than a lower level property. Returning to the apartment example, the sixth floor is higher than the fifth floor because they are different levels in the apartment. For Kim (at this point), the world is mereologically organized, so various micro particles form a whole molecule. These whole molecules then function as parts which make up a whole cell, which in turn make up a whole biological organism (Kim, 1997, Pg. 291). On this scheme it is not the case that higher level properties are properties of the same object as lower level properties. To use Kim's example, the higher level property of ductility is a property of certain objects, namely certain molecules. However, lower level properties such as charm and spin are not properties of these molecules, rather they are properties of their individual micro particles.

One such higher level property is known as a micro-based property. A micro-based property is a property of a whole that is decomposable into the properties and relations of the parts of that whole. Kim offers two examples of a micro-based property – I will focus on the second. ‘Being a water molecule’ is a micro-based property (Kim, 1997, p. 292). This means that being a water molecule is having two hydrogen atoms and one oxygen atom in a specific bonding relationship. ‘Being a water molecule’ is a higher level property because it is impossible for any of the lower level parts to have this property by itself. Hydrogen does not have the property of being made up of two parts hydrogen and one part water in a certain bonding relationship, nor does oxygen.

Since higher level properties, such as micro-based properties, belong to different objects than lower level properties do, Kim says there is no reason to suppose the supervenience argument would apply. Exclusion pressures do not arise until it is determined that there are two properties of the same object fighting for causal efficacy. To use a crude example, the causal properties of a mouse do not compete with the causal

properties of the dog down the street because these properties belong to different objects. Similarly, a micro-based property would not compete with certain micro properties because these properties belong to different objects. However, mental properties do belong to the same brain as neural properties, and so the competition between them remains a concern.

Before moving on, notice that Kim does not identify micro-based properties with the specific properties and relations that make up the micro-based property. Rather, he says that lower level properties construct and constitute the micro-based property (Kim, 1997, p. 296). Perhaps Kim's most vivid statement to this end is found in an earlier exchange with Trenton Merricks. He says:

Macro properties of a thing, therefore, depend on, and are determined by, its micro properties; they are not identical with, or replaceable by, the latter... properties of a whole supervene on the properties and relations characterizing its parts ... This only means that the properties of a whole are fixed when the properties and relations of its parts are fixed; it does not mean that they are identical with, or eliminable in favour of, the latter properties (Kim, 1995, p. 163-164).

Kim acknowledges that micro-based properties depend and supervene upon lower level properties, but will not admit an identity.

This identity cannot be posited for a very important reason: Kim's refutation of the generalization problem is predicated on a distinction between levels. This requisite distinction between higher level macro objects and lower level micro properties cannot be sustained if he simultaneously posits an identity between the two. Kim makes much the same point himself: "if  $x$  is 'macro' in relation to  $y$ ,  $x$  and  $y$  cannot be identical. If  $P$  is a macro property relative to  $Q$ , it logically cannot be that  $P=Q$ " (Kim, 1995, p. 162. See also Kim, 1984, p. 261). Since Kim relies on this distinction between levels, it is safe to conclude that a reductive identity between macro and micro levels is not possible for Kim at this time.

## 8.2 – Problems with the Higher level Solution

The generalization problem supposes that all macro properties that supervene on a subvenient base face the threat of epiphenomenalization. Kim's response is that macro properties of the special sciences are properties of different objects than their micro constituents, and properties of different objects are not causal competitors. Numerous critics have rejected Kim's answer (Noordhof, 1999; Marras, 2000; Gillett et al., 2001; Schroder, 2002; Bontly 2002; Block 2003). I want to look briefly at three of these critical responses, as they all share a common concern.

Paul Noordhof suggests that higher level properties that are constituted from lower level properties still supervene on lower level properties, despite the fact that they are properties of different objects.<sup>44</sup> The reason for this is that as soon as a dependence relation exists between a higher level and some base, then by definition, the base can carry all of the causal weight itself no matter what object the higher level property belongs to. To this end, Noordhof argues that micro-based properties have a micro base which determines the causal power of the micro-based property (Noordhof, 1999, p. 111). In the same way that mental properties face exclusion pressure when confronted with a sufficient base property, so micro-based properties face exclusion pressure when confronted with the fact that the micro constituents determine the micro-based properties.

Ned Block makes much the same point. Block, however, is able to capitalize on Kim's later admission that micro-based properties "supervene on specific mereological configurations involving these micro properties" (Kim, 1998, p. 117). Now that Kim acknowledges the supervenience relation between micro-based properties and specific micro constituents, it becomes even more probable that the supervenience problem can be applied to higher level micro-based properties. This is exactly what Block does (Block,

---

<sup>44</sup> Noordhof (1999, p. 109). He cites Kim (1997, p. 293-296) which demonstrates that higher level properties are constituted from micro-constituents. He then cites Kim (1984, p. 259ff) to demonstrate that constitution amounts to supervenience.

2003, p. 144). Specifically, he says it would be a mistake to assume that a micro-based property causes a later micro-based property to occur, because the micro-based property supervenes on a specific mereological configuration which fully causes the later micro-based property to occur.

Both Noordhof and Block argue that the generalization issue remains a problem because micro-based properties turn out to supervene on their micro-constituents and thus the problem remains. Thomas Bontly, however, takes a different route to the same result. He grants, for the sake of the argument, that micro-based properties are properties of a macro object, and thus there is no competition between these properties and the micro properties which are properties of other micro objects. But imagine, Bontly urges us, that Descartes were here today, and Descartes said mental properties belong to a different object (namely, the soul) than neurological properties, which belong to the body (Bontly, 2002, p. 84). Would Descartes avoid the exclusion problem by simply arguing that mental properties are properties of a soul, not a body? If so, nonreductive physicalists could flock to Cartesianism in order to avoid the exclusion problem. Bontly suggests Descartes cannot escape from the problem and this is because “what really generates the exclusion problem is the fact that one and the same event seems to have both a mental and a physical cause, and it matters not at all whether the two causes are exemplified by the same object” (Bontly, 2002, p. 84). Kim cannot escape the exclusion problem by positing two separate objects any more than Descartes can dodge exclusion worries by positing a soul and a body.

The common theme present in all of these criticisms is that exclusion applies whenever there are two competing causes for one event. It does not matter if these causes are of different orders or on different levels. All that matters is whether there are two causes for one effect. For example, if the mouse fully causes the cat to jump, then we can exclude the simultaneous occurrence of the dog’s barking from playing a causal role in making the cat jump, even though the dog is not the same object as the cat. On Kim’s distinction between levels, there are still two purported causes for one effect, so

exclusion still applies. This being the case, Kim cannot avoid the problem by introducing a distinction between levels and orders.

### 8.3 – The Generalization Problem and the Identity Solution

In the first section Kim attempted to block the generalization argument by positing a distinction between higher order and higher level properties. As it turned out, however, it was possible to reconstruct the generalization argument for any macro property which supervenes on, or in some other way is determined by base properties, regardless of the fact that these base properties may be properties of different objects. In this section I track a discernible shift in Kim's tactics. Specifically, Kim begins to articulate an identity solution to the generalization problem.

In a reply to Noordhof's concerns raised above, Kim announces a need to revise his higher level response to the generalization problem:

For various reasons I now believe what I said in the target paper about micro-based properties and causal powers is incomplete in certain respects and is in need of further elaboration ... [Noordhof's] discussion, however, brings to light some points of unclarity and incompleteness in the target paper, and for this I am grateful - the issues discussed in this note are more fully treated in Kim, 1998 (Kim, 1999, p. 117).

Kim announces here that an addition, or at least a clarification, is forthcoming. But what does Kim have in mind? He seems to be referring to an identity claim of some kind.

The first place where this crucial identity appears is in the sentence before the passage just quoted. He says “To be sure, [micro-based property]  $P$  is determined by having parts  $x_1, \dots, x_n$  such that  $P_1(x_1)$ , etc. But this is trivial and unsurprising, for  $P$  is this property!” (Kim, 1999, p. 117). Here it seems as though Kim is now willing to identify the micro-based property  $P$  with the lower level micro properties that determine

it.<sup>45</sup> In the same way Kim chooses the identity option over the supervenience option in the mental causation debate, so he endorses identity in the generalization debate.

The next place where Kim appeals to identity to resolve the generalization problem is in *Mind in a Physical World*. Recall that Kim says this text contains the key additions. His discussion in *Mind in a Physical World* follows his earlier writing quite closely - even verbatim at times. At the end of his discussion, however, an addition is discernible. Consider the following fragment from the article written in 1997:

Difficulties of this sort do not arise for micro-based properties in relation to their constituent properties, because the former do not supervene on the latter taken individually or as a group. It is an interesting question ... (Kim, 1997, p. 296).

Compare this to the reworked version from 1998:

Difficulties of this sort do not arise for micro-based properties in relation to their constituent properties because the former do not supervene on the latter taken individually or as a group. Rather, they supervene on *specific mereological configurations* involving these microproperties—for a rather obvious and uninteresting reason: they *are* identical with these micro-configurations. Now it is an interesting question ... (Kim, 1998, p. 118-119).

Kim adds that a micro-based property is identical to the specific mereological configuration of the micro properties. There is an important difference here, however. In the first passage, Kim identifies the micro-based property with having certain micro parts.

---

<sup>45</sup> Gillett and Rives argue that Kim and Noordhof are talking past each other on this point, since Kim seems to have identity in mind, while Noordhof assumes Kim does not (Gillett et al., 2001, p. 91). This interpretation, however, does not give proper weight to the fact that the identity solution does not appear in the earlier text, a fact Gillett and Rives acknowledge (Gillett et al., 2001, p. 91). This interpretation does not account for Kim's acknowledgement of a forthcoming addition either. If Kim had identity in mind all along, he would not need to clarify or add to his earlier works. Finally, this interpretation does not take into account the aforementioned considerations that specifically barred Kim from appealing to this identity

In the second passage, Kim identifies the micro-based property with a “configuration” involving these micro properties. This difference will become important below.

In *Physicalism or Something Near Enough*, the identity appears along with an explanation of how this identity solves the generalization problem. Kim’s analysis is found in his response to Ned Block (Block, 2003). Block wonders whether there may be no bottom level of elementary particles, hence exclusion may drain all causal powers into a bottomless pit, leaving no causation anywhere. In responding to this possibility, Kim considers the example of water once again. The micro-based property ‘being H<sub>2</sub>O’ is the property of being made up of two hydrogen atoms and one oxygen atom with a specific structure (Kim, 2005, p. 68). This much had been said in his earlier comments on water, but now he adds the identity: “Being water is having this kind of microstructure. Having this microstructure is the microstructural essence of water, and being water just is having that structure” (Kim, 2005, p. 68). ‘Being H<sub>2</sub>O’, as a higher level micro-based property, is here identified with the lower level microstructure. Kim then points out that hydrogen and oxygen are macro objects as well, which have a certain microstructure also. Again, however, he identifies the macro object hydrogen with the microstructure of the parts. He continues this process all the way down the mereological ladder (and here L-1 stands for a lower level and M stands for a micro-based property):

These identities in turn imply the following series of identities: M<sub>L</sub> = M<sub>L-1</sub> = M<sub>L-2</sub> = M<sub>L-3</sub>... Voila! These are the identities we need to stop the drainage...the primary response to the drainage argument is the point that for downward causal drainage to occur, the reduction option must be ruled out for purely physical levels, including microphysical levels, and it is far from obvious that this can be done. In fact, the drainage problem provides us with one more reason to perform a reduction against the irreducibility premise of the supervenience/exclusion argument (Kim, 2005, p. 69).

By reductively identifying the micro-based property with the configuration of lower level properties, Kim stops the drainage. How does this happen? In the same way that he

identifies the mental with the physical to provide mental causation, so he now identifies the macro property with the causally potent lower level to bestow causal status to the macro level. This solution is as effective in preventing higher level epiphenomenalism as his solution to the mental causation problem is.

Before moving on, let us take notice of one important point. In the previous passage Kim argued that a nonreductive approach, which posits higher levels as distinct from lower levels, will not work for preventing drainage or solving the causal exclusion problem. In fact, recall that the higher level solution was embedded in a layered model of the world (Kim, 1997, p. 289-291), which prevented any identity from going forward. Perhaps it is not surprising to see that Kim begins to express deep reservations about this layered model of the world. In his 2002 article entitled “The Layered Model”, Kim tentatively concludes it is a mistake to suppose the world itself is mereologically ordered:

But this talk of higher and lower can only be provisional: the question of reducing [higher level properties] to ‘lower level’ properties is open, and if the reduction goes through, there is no higher or lower. Under a successful reduction, the properties initially regarded as ‘higher’ have been reductively absorbed into lower level properties, and the hierarchy of properties with which we began collapses (Kim, 2002, p. 19. See also Kim, 1999, p. 33).

Kim not only doubts the nonreductive higher level solution, but the viability of the layered model which sustained it as well. In its place, Kim posits both a reductive identity of the levels and the flattened world that ensues. The layered model and the flat model are incompatible – identity implies there is one level, while higher/lower talk implies there are at least two (Kim, 1984, p. 261; Kim, 1995, p. 16; Kim, 2005, p. 54).

This being the case, Kim ought to make a clear break from the higher level response to the generalization problem, along with its layered ontology, to the identity response with its flat ontology. Kim, however, continues to mention the possibility of levels in his later discussions on the generalization problem (2005, Pg. 57) – and this unnecessarily clouds the issue. Perhaps for this reason, some critics continue to argue

that Kim's response to the generalization problem involves levels and orders (Baker, 2009). Not only has the levels response been refuted, but the levels response is no longer available to Kim, given his flattened ontology. What remains is the view that special science properties retain their causal efficacy because they are identical with a configuration of causally potent microphysical properties.

#### **8.4 – Problems with the Identity Solution**

In the previous section we saw that Kim replaced his ill-fated higher level solution with an identity solution. Can the identity solution hold up under scrutiny? In order to answer this question, we need to discern exactly which entities are targeted for identification. As it turns out, this is not a simple matter - indeed, this will be part of the subject matter of this closing section. It is clear that one of the entities is the micro-based property. Beyond this, several authors suppose the second entity is the lower level properties and the lower level individual relations between these properties which make up the micro-based property (Bontly, 2002, p. 87-88; Gillett et al., p. 91). As we have seen, this interpretation seems to be in line with Kim's inaugural discussion of the identity solution (Kim, 1999, p. 117). Moreover, it fits well with Kim's insistence that a micro-based property is decomposable into the many properties and individual relations between the parts. I will consider another possibility below, but for now let us assume the identity is between the micro-based property and the micro properties and individual relations between these micro properties.

There are a number of problems with identifying a micro-based property with the micro properties and individual relations between these micro properties. The most popular criticism of this identity is that the prospect of multiple composition makes such an identification impossible (Bontly, 2002, p. 87; Shroeder, 2002, p. 324; Block, 2003, p. 145; Walter, 2008, p. 691). Ned Block clarifies the problem by using the example of jade. Jade can be composed of calcium magnesium silicate (nephrite) or sodium aluminum silicate (jadeite) (Block, 2003, p. 145). The micro-based property of being

jade can therefore be composed of different micro properties in different relations. This is problematic because we now cannot identify jade with the specific micro configuration known as nephrite, for this would rule out jade being jadeite as well. Nor can we identify jade with the disjunction of ‘being nephrite or jadeite’, for Kim does not think such disjunctive properties exist (Kim, 1992; Kim, 1998, p. 103-110; Block, 2003, p. 147).

Of course, this argument from multiple composition closely resembles the argument against psychophysical reduction from multiple realizability. Not surprisingly, Kim responds to the problem of multiple composition in the same way that he responds to the problem of multiple realizability. Specifically, he focuses on identifying individual instances of jade, rather than jade as a type: this jade is nephrite, that jade is jadeite, though all jade is not nephrite (Kim, 2005, p. 57). This gives an instance of jade all the causal power that an instance of nephrite has on a particular occasion.

Sven Walter has recently responded to Kim’s analysis (Walter, 2008, p. 690-692). According to Walter, Kim’s response is similar to the nonreductive physicalist position on mental causation. Namely, many nonreductive physicalists identify token mental events with token physical events, but refrain from identifying mental and physical properties. Kim rejects the nonreductive physicalist move because his property exemplification model of events cannot allow token events to be identical without their properties being identical as well (Kim, 2005, p. 42; Marras and Ylu-Vakkuri, 2008). We have already seen why the properties themselves cannot be identical, so the property instantiations cannot be identical either.

Kim offers up a second possible reply to the argument from multiple composition. He says that we can perhaps identify jade with the disjunctive kind ‘being jadeite or nephrite’ (Kim, 2005, p. 58). This solution would give jade all the causal power that jadeite or nephrite has, while retaining consistency with Kim’s property exemplification model of events. There are two problems here. First, as previously noted, Kim once argued that disjunctive kinds do not really exist. This move would force Kim to modify his stance on the existence of disjunctive kinds. More importantly for our purposes here,

however, Sven Walter has noted a second problem with this response (Walter, 2008, p. 691). Presumably, being jadeite is a property in its own right. If this is so, then we need to ask whether the property of being jadeite has sufficient causal powers in its own right to determine what an instance of jadeite can do. Seemingly, the answer is yes, for if we said no then we would be suggesting that an individual realizer is not causally sufficient. If the answer is yes, then what work is left over for the disjunctive property of ‘being jadeite or nephrite’ to do? There is none. The individual micro property excludes the disjunctive property. This pattern will repeat itself in an important way below.

Another reason to refrain from identifying a micro-based property with certain micro properties and their many relations is that this implies that a singularity is a plurality. A micro-based property, such as being a water molecule, is singular. However, there are numerous micro properties that constitute it. For example, there are two property tokens of being a hydrogen atom, and another property token of being an oxygen atom (which can each be decomposed in turn). There are also a number of bonding relationships. For example, one hydrogen atom stands 95.84 picometres away from the oxygen atom, the other hydrogen atom stands 95.84 picometres away from the oxygen atom, and the two hydrogen atoms stand some distance away from each other as well, while the angle formed by these two hydrogen atoms is 104.45°. Can a singularity be successfully identified with this plurality? Cynthia MacDonald answers much the same question in the negative:

A substance is a single thing. Thus it is not identical with its properties, taken collectively, since a single thing cannot be identical with many things, and these properties are many things (MacDonald, 2005, p. 86).

The problem is that ‘being a water molecule’ has the property of being singular, while ‘being hydrogen’ and ‘being oxygen’ and ‘being bonded’ are jointly a plurality. Since

the water molecule has a property that the conglomeration does not have (and vice versa), they cannot be identical.<sup>46</sup>

Kim seems to have two options at this point, neither of which ends well. First of all, Kim can defend the identity of the whole with its parts. This doctrine, known as composition as identity, is endorsed by David Armstrong (Armstrong, 1997, p. 12), but is most vigorously defended by Donald Baxter. As an example of how it works, Baxter suggests that six plots of land located beside each other do not jointly make up a seventh whole plot of land which can be sold separately. Rather, “the six parts are the so-called seventh thing” (Baxter, 1988, p. 197).

In making this move, Baxter rejects Leibniz’s principle that *identicals* are indiscernible. In its place, he argues that one identical object is discernibly different from itself at times:

So [to] say that *identicals* are discernible, [means] that there exists some *a* such that *a* has and lacks a property. For *a* insofar as it is one way has the property and *a* insofar as it is another way lacks it (Baxter, 1988, p. 205-206).

If the criticism is that a whole cannot be identical to various parts, Baxter responds by saying that an object can be singular in one way, yet not singular in another way, while still remaining self-identical.

---

<sup>46</sup> This point is similar to the argument that Thomas Bontly makes. Bontly argues that the indiscernibility of *identicals* prevents us from identifying properties if they are not instantiated by the same objects at the same time (Bontly, 2002, p. 88). If there is only one property, then surely wherever it is, it will also be. However, according to Kim’s model, the micro-based property is a property of a different object than the micro properties are. This being the case, the micro-based property cannot be identical to the micro properties. Kim, for his own part, argues that micro-based properties cannot stand in a determination relation with the micro properties, for the micro properties are properties of the parts of the object, not the object itself (Kim, 1999, p. 117). If micro-based properties cannot be determined by micro properties since they are properties of different objects, then for the same reason micro-based properties cannot be identical to micro properties either.

Baxter's argument has met some resistance (Van Inwagen, 1994; Yi, 1999; McDaniel, 2008). Amongst the complaints is the palpable confusion over how an object can be discernibly different from itself (Van Inwagen, 1994, p. 214; Harte, 2002, p. 24). Indeed, Baxter's suggestion brings fresh and literal meaning to such statements as 'I am not myself today'.<sup>47</sup> Numerous philosophers who are sympathetic with the composition as identity view refrain from endorsing an outright identity on account of this perplexity. For example, David Lewis suggests that composition is merely analogous to identity because: "what is true of the many is not exactly what's true of the one. After all, they are many while it is one" (Lewis, 1991, p. 87; see also Sider, 2007). If we follow Lewis in rejecting the outright identity, then we are not able to sustain the identity between one micro-based property and the properties of its many constituent parts.

Not only is there a discernible difference between the singularity and plurality of the properties being considered for identification, but there are further differences as well. A macro property and the micro properties that compose it, for example, do not share the same modal properties. The textbook example of this principle suggests that the piece of artwork called 'David' has the property of being a statue, and the properties of being certain marble pieces. Plausibly, being a statue is not identical to being these certain pieces of marble because these marble pieces can survive disfiguration (for example,

---

<sup>47</sup> Importantly, there is reason to suppose that Kim would not endorse Baxter's model. Baxter argues that we can count something as many, or we can count it as one. The one-many identity, therefore, is simply a matter of choosing to count something in one way or another. The way we choose to count something is relative to human interests (Baxter, 1988, p. 210). Kim, on the contrary, thinks that the relevant micro-based properties for the generalization debate are genuine natural kinds. Nature itself cuts things up into pluralities or singularities, and it is our job to count these properly. Baxter rarely uses natural kinds in his examples, preferring instead to talk about such things as boxes of juice and the legal parceling of land (Baxter, 1988, p. 199; p. 200; p. 210). In these examples, there is no natural kind that is really a singularity or a plurality, which needs to be counted as such. Rather, the way these objects are counted depends on human interest. Beyond this, Baxter is clear that within one way of counting, we still cannot identify a singularity with a plurality (Baxter, 1988, p. 193, p. 201, p. 209).

being broken in pieces), whereas the statue cannot. (Baker, 1997; Johnston, 1992; Shoemaker, 1999). Or to use a slightly different example, the animal named ‘Bouncy’ has the macro property of ‘being a rabbit’ and also is composed of numerous molecules. The macro property of ‘being a rabbit’ is possibly not identical to having these many molecules because Bouncy can continue to be a rabbit while she slowly loses all of these molecules and gains new ones. Or again, if the molecules that make up the rabbit decompose and are spatially separated from one another, the molecules still exist, whereas the rabbit does not. Since the macro entity does not have the same persistence conditions as the micro parts, we can conclude that they are not the exact same thing.

Not only do we possibly have diverging numerical and modal properties between these two entities, but it is also plausible that the composition as identity thesis entails mereological essentialism (Merricks, 1999, p. 193; Van Cleave, 1985, p. 98). According to mereological essentialism, if a whole is identical to certain parts at a time, and over time these parts are replaced by other parts, then the whole is replaced by another whole as well. To use Van Cleave’s example, imagine that object *O* is identical to parts *FGH*. Now part *H* is replaced by part *K*, so we now have parts *FGK*. Van Cleave points out that “*FGH* is simply not identical to *FGK*” (Van Cleave, 1985, p. 98). Since object *O* is *FGH*, and *FGH* is different from *FGK*, object *O* is different from whatever object *FGK* is. Object *O* is replaced, it does not persist through time or change. However, we typically think that macro-objects can survive modifications to its parts. Partly for this reason, Van Cleave says we should not identify an object with the bundle of properties. However, he argues that if we were to downplay the ontological status of the object, and focus instead on our ability to translate statements about the object into statements about the bundle of properties, we may be able to salvage this view (Van Cleave, 1985, p. 103).

In a discussion of Van Cleave’s work, Kim agrees that a danger of identifying objects with bundles of properties is that “it makes every property that an object has an essential property of it” (Kim, 1993, p. 129). However, Kim suggests that Van Cleave’s grammatical solution to this problem amounts to “fictionalism” (Kim, 1993, p. 128)

about individual objects. Rather than resorting to fictionalism, Kim suggests that we can use a relationship of supervenience to avoid the identity but preserve realism. This is done by seeing the supervenient domain as the object, and the base domain as the bundle of properties. Then, as Kim explains:

No member of the supervenient domain is identified with a member of the base domain ... nevertheless we may say ... that all truths about the supervening domain are fixed once the truths about the base domain are fixed... Thus, uncoordinated multiple-domain supervenience seems to give us a perspicuous general scheme that captures the metaphysical determinative relationship underlying the new bundle theory without the latter's implicit commitments to complex issues about meaning and translation. Freedom from such commitments is precisely what makes the metaphysical version more appealing than the translation version of the theory (Kim, 1993, p. 128-129).

In this passage Kim refrains from identifying the supervening object with the micro-properties, while still endorsing supervenience to improve the ontological status of objects. Of course, at this point, the supervenience option was alive for Kim, so it seemed a natural solution. However, as Kim moves away from the supervenience thesis, the problem returns. If an object is certain specific parts, and those parts change, then we do not seem to have the same object anymore.

For all of these reasons, it seems as though Kim ought to reject the identity between a micro-based property and certain micro properties and the individual relations between these micro properties.<sup>48</sup> Perhaps, for the sake of the argument, we can reject

<sup>48</sup> Indeed, Kim at times makes statements suggesting that a self-identical object cannot have any discernible differences from itself. He argues: "Distinct properties are just distinct, and we can't pretend they are the same. I don't think it's good philosophy to say, as some materialists used to say, 'But why can't we just say that they are one and the same? Give me good reasons why we shouldn't say that!' I think that we must try to provide positive reasons for saying that things that appear to be distinct are in fact one and the same" (Kim, 1998, p. 98). Beyond this, we have seen Kim's hesitancy to identify macro properties with micro properties, possibly on account of the fact that they simply seem

the indiscernibility of identicals, and accept that a singularity is many, and that two objects can be the same despite their having differing modal properties, or despite the fact that they have different properties at different times. There is still reason to believe that this would not help in our discussion of the generalization problem. Recall that Kim solves the generalization problem by identifying the higher level with the causally potent lower level. But this solution presupposes that the indiscernibility of identicals is true. Consider Ned Block's summary of Kim's argument:

As I understand Kim, the answer is that the key feature of micro-based properties is that they are identical with micro-configurations. If water = H<sub>2</sub>O ... there is no worry about the causal efficacy of water draining down to H<sub>2</sub>O – if H<sub>2</sub>O is efficacious, so is water, since identicals are indiscernible (Block, 2003, p. 145).

H<sub>2</sub>O has causal properties, and water is H<sub>2</sub>O, and since the indiscernibility of identicals says that an identical substance has all identical properties, water must have the causal properties that H<sub>2</sub>O does. If Kim were to reject the indiscernibility of identicals, a critic could lodge the following complaint against his position: H<sub>2</sub>O has causal powers, and water is H<sub>2</sub>O, but water is discernibly different from H<sub>2</sub>O in one important respect.

Namely, H<sub>2</sub>O has causal properties that water does not have, so water does not inherit the causal powers of H<sub>2</sub>O on this occasion, regardless of the identity. Of course, such a criticism sounds patently absurd. And it is. But this is only because identicals cannot be discernibly different in this, or any other way.

There was, however, a second route that Kim could take to secure the required identity. Rather than identifying a micro-based property with various micro properties and individual relations between these micro properties, he can identify the micro-based property with a specific configuration of micro properties, which is also singular. This move would avoid the difficulties with identifying the macro property with numerous

---

discernibly different. Kim also doubts that we can identify first-order properties with second order properties at times (Kim, 1998, p. 103).

micro properties and relations by identifying the macro property with one configuration of the micro properties. Patrick McGivern thinks this is Kim's move:

It appears that *one* higher level property is being identified with *many* lower level properties. I take it that Kim's talk of "specific mereological configurations" is meant to suggest that "configurations" of microproperties are to be understood as individuals, hence avoiding any worries about identifying one macro level property with many (McGivern, 2008, p. 59).

McGivern's analysis gains support from Kim's second articulation of the identity solution, which we saw in *Mind in a Physical World*. There Kim marks a distinction between the many micro properties that constitute the micro-based property taken individually, or even all of these micro properties taken as a group, and the mereological configuration of the micro properties. Recall that he says: "difficulties of this sort do not arise for micro-based properties in relation to their constituent properties because the former do not supervene on the latter taken individually or as a group. Rather, they supervene on specific mereological configurations involving these microproperties" (Kim, 1998, p. 117). Kim distinguishes between constituent properties taken individually or even in a group and a specific mereological configuration of micro properties. He then identifies the micro-based property with the specific mereological configuration of micro properties (although he does use the plural 'configurations'), which perhaps implies that Kim does not want to identify the micro-based property with the many micro properties and their relations, but rather that he wants to identify the micro-based property with the mereological configuration, or the mereological relation, of the micro properties.

This proposal also gains support from Kim's view of the relationship between a macro object and its parts. The relationship that a macro object has to its parts is similar to the relationship that a macro property has with its micro properties. It comes as no surprise, therefore, to see that Kim assumes much the same relationship exists between a macro object and its parts. Kim illustrates his view with the help of a baseball:

The shattering of the glass was caused by the baseball and certainly not by the individual particles composing it. True, the baseball is a composite object made up of its constituent molecules, atoms, particles, or what have you, that this complex structure consisting of microparticles broke the window. But there is no mystery here: the baseball = this composite structure of microparticles (Kim, 2005, p. 56).

Here a macro object is identified with a composite structure of microparticles. Elsewhere Kim argues that a table or a biological organism can be reduced by “being identified with a structure (‘swarm’) made up of its parts” (Kim, 1995, p. 162). In a similar way that a macro object is the structure of its parts, so a micro-based property now seems to be the mereological configuration, or mereological relation, of specific micro properties.

As noted, this proposal helps to alleviate the difficulty of identifying a whole with many parts. It also resolves the problems associated with multiple composition, and with one object having differing modal and temporal properties than itself. However, it leads to other problems. For now we can ask whether this individual configuration is itself identical with the many micro properties and their individual relations that make it up, or whether it is merely constituted by them. If the configuration is identical with the many micro properties and their relations then the previous problems arise anew. Specifically, how can one configuration have multiple compositions; how can an individual configuration be identical to numerous micro properties and relations; etc...

Suppose we avoid the repetition of these problems by choosing to say the individual configuration is merely constituted by its micro properties and individual relations. This move, however, threatens to bring the danger of epiphenomenalism back in full force. Recall that Kim identified micro-based properties with some underlying entity in order to secure the causal potency of micro-based properties. If it turns out that micro-based properties are identical with a configuration of micro properties and their relations, but this configuration is not identical with these micro properties and relations, then it turns out that the configuration (and, by virtue of the identity, the micro-based

property) is causally excluded since the micro properties and individual relations can do all of the causal work themselves.

Consider, for example, Kim's example of the ten kilogram table. The table has a mass of ten kilograms. The table also has a micro-based property of being decomposable into a six kilogram top and a four kilogram pedestal in an on-top-of relation. Suppose the micro-based property is identical to this configuration, but is not identical to the micro properties of the top being six kilograms and the pedestal being four kilograms. These two micro properties, however, together add up to the total ten kilograms of the table, and there is no more weight to the table. The micro properties contribute all of the weight. So, if the configuration is not identical to these micro properties, then the exclusion principle suggests that the configuration would be excluded.

This argument is similar to the one raised earlier in the discussion of multiple composition. We saw that the property of 'being jadeite' did all of the causal work, so it excluded the disjunctive property of 'being jadeite or nephrite' from making a causal contribution. In the same way, the micro properties of 'being six kilograms' and 'being four kilograms' amount to all of the ten kilogram weight, so there is nothing left for a configuration of these properties to contribute. For this reason it seems like Kim has to identify the configuration with the properties and their relations. But this brings back the difficulties associated with making this sort of identity.

Perhaps there is a way out of this dilemma. Perhaps Kim could say that a micro-based property is identical to a configuration of micro properties and relations, thereby dodging the problems of identity. However, he could also avoid the causal problems by saying that the configuration does provide powers that the micro properties and relations of the configuration do not provide in themselves. Let us flesh out this possibility a little. Recall that Kim claims micro-based properties are akin to what David Armstrong calls structural properties (Kim, 1998, p. 84). David Armstrong, however, marks a clear distinction between what he calls relationally structural properties and non-relationally structural properties (Armstrong, 1978, p. 71). Non-relationally structural properties are

structural properties where the relations between the parts do not influence whether or not a whole has a certain macro property. Armstrong uses the example of an object being a mass of one kilogram. It doesn't matter how the parts are arranged, the macro object will still have the property of weighing one kilogram if the parts together have that weight.

On the contrary, relationally structural properties are structural properties where the relations between the parts do influence whether a whole has a certain macro property or not. Consider the isomers fructose and glucose. Fructose and glucose are both composed of  $C_6H_{12}O_6$ . They have the same parts, which have the same intrinsic properties. They are, however, structured differently. Because of the structural differences, fructose has certain macro properties that glucose does not have, and vice versa. For example, fructose is sweeter than glucose, and causes less insulin secretion in humans than glucose. This tells us something important: at times, structure may make a contribution itself.<sup>49</sup> If structure makes a contribution, then perhaps we can identify the micro-based property with the structure, or configuration, thereby securing some efficacy for this micro-based property.

With this possibility in mind, let us return to the previous objection: if the micro properties and their individual relations can do all of the causal work themselves, then what work is there left for the configuration to do? Rather than falling prey to the exclusion argument here, we can now say that the configuration of the parts does make a unique contribution. Clearly the specific manner in which the molecules composing glucose are arranged makes a difference, since the same molecules configured differently yield different results. But then, does this mean that the micro properties and their individual relations cannot do all of the causal work themselves?

---

<sup>49</sup> Although Armstrong marks out this division, Kim does not. Rather, Kim simply places both of these types of structural properties under the more generic category of being a micro-based property - a fact that has troubled at least one critic (Shroeder, 2002, p. 323-324).

There are a number of available answers that Kim has here. First of all: yes, it means the micro properties and their individual relations cannot do all of the causal work themselves. If the configuration makes a contribution, then the micro properties and relations must leave some room for this additional role. This move seems to reject causal completeness. Causal completeness states that every event has a sufficient microphysical cause, but we seem to have just suggested that the micro properties and their relations do not do all of the causal work themselves.

A second available reply is to insist that the micro properties and their relations can do all of the causal work themselves, which keeps causal completeness intact. Kim, of course, makes this reply. However, this reply returns us to our original problem. Namely, since the micro properties and their individual relations do all of the causal work, we can exclude the configuration from making a causal contribution. We may attempt to get around this dilemma by returning to the identity solution, whereby we identify the configuration with these micro properties and relations and thereby achieve a causal role for the configuration. However, the identity solution has already been ruled out. There is another option, which I will explore in the next chapter.

For now, we can conclude that Jaegwon Kim's supervenience argument was accused of generalizing to all of the macro properties of the special sciences. Kim's original response was to posit mereological levels for the special sciences in contrast to higher orders in the case of mental causation, but this did not work. Kim then blocked off the supervenience argument by identifying "higher level" properties with causally potent "lower level" properties and relations, but this does not work either. For a number of reasons we could not identify the higher level property with micro properties and their individual relations. If we attempt to identify the higher level property with a configuration of micro properties and relations we will face additional problems. Namely, this configuration could not be identical to the micro properties and their relations. However, if it is not, then it seems like any role for the configuration is excluded, since the micro properties and relations are sufficient to cause a given effect.

## **Chapter 9 – Specific Mereological Structure and Non-Causal Mereological Determination**

At this point I have rejected a number of replies to the causal exclusion argument. I have also rejected the viability of abandoning one of the four principles in our seemingly inconsistent tetrad. I then argued that shifting our metaphysical stance on events or causation will not work either. I proposed that a resolution might be found within a mereological system which construes mental states as non-causal determinative structures of parts. However, to this point we have only seen that configurations of micro properties and relations continue to face causal exclusion pressures since the individual micro properties and relations are causally sufficient for the effects in question. In this chapter I will suggest that the configuration of micro properties does play a role beyond that of the individual micro properties and relations, but this role is a non-causal determinative role.

This chapter is divided into four sections. In Section 9.1, I clarify what I mean when I talk about a structure of parts. I also demonstrate that structure makes a contribution, and how structure makes a contribution. In Section 9.2, I discuss how we can legitimately conceive of the contribution of the structure of parts as determinative rather than causal. In Section 9.3 I consider an objection to macro to micro determination called the problem of co-dependence.

### **9.1 – Mereological Determination**

I want to clarify some terminology that was used in the last chapter, and will be important in this chapter as well. First of all, we have seen that Kim's notion of a micro-based property is best construed as a specific mereological configuration of certain micro properties. We have seen that these micro-based properties are akin to what Armstrong calls structural properties. Armstrong then draws a distinction between relationally structural properties and non-relationally structural properties, where the specific

structure makes a difference with relationally structural properties. Let us follow Armstrong's vocabulary, since it emphasizes the role that structure can have.

These structural properties are also what Armstrong calls particularizing properties (Armstrong, 1978, p. 116-117). For my purposes, particularizing properties characterize the whole object as the type of object that it is. For example, 'being a rabbit' is a particularizing property of Bouncy the rabbit. Or, to use some of the examples we have seen from Kim, a baseball has the particularizing property of 'being a baseball', and water has the particularizing property of 'being water'. These particularizing properties are structural properties in the following sense: 'being a rabbit' is instantiated as a specific structure of Bouncy's parts, and 'being water' is instantiated as the specific structure of its two hydrogen parts and its oxygen part. I will also refer to these structural properties as the macro property, as it characterizes what the macro object is, and it is that type of macro object in virtue of having the specified structural property.

Having established what a structural property is, let us move on to a demonstration that it makes a contribution to macro level effects. We have already seen how structure makes a contribution in the case of glucose and fructose. These two molecules have the same parts, but are structured differently. As a result of their differing structure they have different causal powers. This same pattern is discernible when we consider the aforementioned case of the rabbit. Many parts arranged randomly will not have the property of being able to chew, or to jump. However, these same parts arranged as a rabbit do have the ability to chew and jump. Or again, two hydrogen atoms and one oxygen atom arranged at a distance of one kilometer apart will not have the same ability to bond with water molecules, though these parts arranged as a water molecule do. Virtually any composite object will lose certain powers when its parts are radically rearranged. Perhaps this is why David Armstrong considers it necessary to separate out a category of relationally structural properties, for there are so many cases where structure makes a difference.

One area where the role of structure is ably analyzed is in the literature on emergentism. Before considering this analysis, I want to make a cautionary note about emergentism. In Section 3.4, I ruled emergentism out on account of the fact that emergent properties supervened on causally sufficient configurations of micro properties. It is true that if emergent properties are supervenient properties, they will be susceptible to the causal exclusion problem. And, it is also true that a number of emergentists insist that emergent properties are supervenient properties. However, emergentism comes in a number of varieties, not all of which construe emergent properties as supervenient properties.<sup>50</sup> Rather, some consider emergent properties to be structural properties (Gillett, 2006; El-Hani & Emmeche, 2000; El-Hani 2002; Murphy, 2006. See also Campbell and Bickhard, 2001, p. 19; Van Gulick, 1993, p. 250ff). Within the British emergentist tradition, Carl Gillett argues that Samuel Alexander was a structural emergentist (Gillett, 2006, p. 266-267). Alexander, in turn, traces the position back to the “ancient distinction of form and matter” (Alexander, 1920, p. 47). This structural emergentism is similar to the position I will be taking.

Structural emergentism can be contrasted with the supervenient emergentism that we saw earlier in a number of ways. First of all, supervenient emergentism is clear that emergent properties are not structural properties, or relations between parts:

The notion of an emergent property can then be understood in part by way of contrast with structural properties. An emergent property is a property of a composite system that is wholly nonstructural...Emergent properties are

---

<sup>50</sup> There is the familiar distinction between epistemological and ontological emergentism. The epistemological emergentist will countenance conceptual layers (Bedau, 1997; Nagel, 1961, p. 369), while the ontological emergentist will endorse a hierarchical model of reality itself (O'Connor and Jacobs, 2003; Silberstein and McGeever, 1999). There are those who endorse synchronic emergentism, who claim that emergent properties have causal power at the same time as underlying micro properties do. On the contrary, there are those who endorse diachronic emergentism, who argue that emergent properties display their causal efficacy through time. See Humphreys, 2008; Stephan, 2002; Rueger, 2000 for a discussion of diachronic and synchronic emergentism.

nonstructural properties of composite individuals (O'Connor and Wong, 2005, p. 666. See also, Markic, 2004, p. 75-76; O'Connor, 1994, p. 97-98).

Although supervenient emergentism recognizes the importance of structural properties, they do not identify emergent properties with structural properties.

Structural emergentism, on the contrary, identifies an emergent property with the configuration, or mereological relation, or structure of the parts. Consider, for example, Nancey Murphy's analysis:

The downward efficacy of supervenient mental properties or events can only be seen when we understand supervenience in terms of the relation of the base properties to a broader causal system - the supervenient property is the base property's relation to a broader system, which entails its enmeshment in a more complex set of causal processes (Murphy, 2006, p. 238).

Here we see that Murphy identifies the supposed supervenient, or emergent, property with the relations between the base properties of a system. Likewise, Robert Van Gulick, though not specifically endorsing structural emergentism, takes a similar position. He argues that "patterns of organization ... are picked out by the predicates of the special sciences" (Van Gulick, 1993, p. 250). According to Van Gulick, the special sciences do not refer to properties over and above the composite object, but rather special science predicates refer to the patterns of organization of a composite object. As a final example, Samuel Alexander argues that we should not consider an emergent quality as something merely correlated with "a certain configuration of its basis," but rather, we should "identify the quality with its peculiar form of body" (Alexander, 1920, p. 47). Here again, we see a philosopher refraining from identifying an emergent property with a supervening property of the whole, but rather identifying an emergent property with the configuration of the parts of the whole.

Not surprisingly, the structural emergentist considers the contribution that emergent properties make quite differently as well. On supervenient emergentism, there were parts, then there was an arrangement which provided for certain new powers, then

there was an emergent property which supervened on the arrangement of the parts. The problem for them was that the parts together with the arrangement of the parts could do all of the work themselves, so there was no work for the supervening emergent property to do. On structural emergentism, there are parts and there is an arrangement which produces some new powers. It is this arrangement which is the emergent property, so the emergent property has all of the power that the structure of the parts has.

In order to understand how structure plays a role for the structural emergentist I will first mark a distinction between intrinsic properties and conditional properties. Intrinsic properties are properties that an object has at all times. Examples of intrinsic properties include mass, shape and possibly charge. Intrinsic properties can be contrasted with conditional properties, or, in other words, “latent” properties (Shoemaker, 2002, p. 54), or perhaps, “potential” properties (McLaughlin, 1992). Conditional properties are properties that an object can possibly have, but does not necessarily have. A person, for example, can have the property of being a professor, but does not have to have this property. A cardboard box can be a home, but does not have to be. Presumably, though this will not concern us, there are also certain properties that an object can never have. The person, for example, can never be the heat of the sun – nor can the cardboard box for that matter.

According to a variety of structural emergentists, structure plays a role in generating certain effects by determining whether its parts will instantiate certain conditional properties or not. This determinative relation can be inhibitory; in the sense that the structure determines that certain conditional properties will not be instantiated by its parts. Or, this determinative relation can be excitatory, in the sense that the structure triggers the parts to instantiate certain conditional properties. Let us consider examples of both, beginning with the latter, and using an example from Sperry (1969). Imagine that we have various parts such as a flat piece of rubber, a flat piece of metal, and certain thin rectangular metal rods. These pieces all have certain intrinsic properties, such as mass and hardness. These pieces can also have certain conditional properties, such as the

rod's potential to be a weapon. These pieces, once arranged in the shape of a wheel, instantiate some of these conditional properties. For example, the rods instantiate the properties of being spokes, the metal instantiates being round and being a rim, and the rubber instantiates being round.

These freshly instantiated conditional properties bring causal power to the parts which the parts did not have before. For example, the wheel is placed on a smooth hill and it starts to move. What causes the wheel to move? The roundness of the rim and rubber cause the wheel to move downhill. Roundness, however, is a conditional property that the rim and the rubber only instantiate when structured as a wheel. If these parts had not been structured as a wheel, they would not have moved downhill. We can say that the structure of the wheel plays a role by determining that its rim and rubber will instantiate the causally efficacious conditional property of being round.

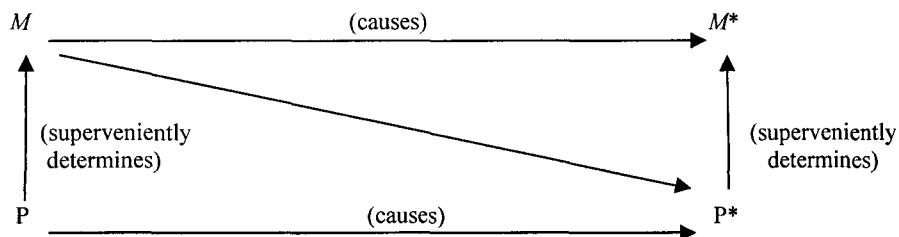
Thus, as Carl Gillett summarizes, structural or emergent properties are “efficacious by determining the contributions of powers by other property instances and not solely by itself contributing powers” (Gillett, 2006, p. 274). Or, as Nancey Murphy affirms, these structural emergent properties “can affect which causal powers of their constituents are activated or likely to be activated” (Murphy, 2006, p. 229). Similarly, Robert Van Gulick argues that “the pattern of which it is a part may affect which of its causal powers get activated” (Van Gulick, 1993, p. 251. See also Sperry, 1986, p. 266; Shoemaker, 2002; Francescotti, 2007, p. 57ff).

As hinted at above, not only do structural properties determine that certain conditional properties will be instantiated, but structural properties also constrain certain conditional properties from being instantiated. El-Hani and Emmeche explain that structural emergentism entails “the modification of a complex system’s parts by the system itself as a constraint implied by being part of a pattern” (El-Hani and Emmeche, 2000, p. 263; See also Shroder, 1998; Kistler, 2009, p. 602). Thus, to use their example, a certain group of molecules arranged randomly in a test tube may interact in any number of ways. However, once we bind these molecules into a cell, these molecules are now

limited in how they can interact with one another. The molecules being structured as a cell now “have a much more ordered distribution in space-time than they would have in its absence” (El-Hani and Emmeche, 2000, p. 262). Not only can structure limit which conditional properties its parts will instantiate in this way, but we have seen that structure also triggers certain conditional properties of its parts to be instantiated. Clearly then, the structure that certain parts are placed within plays an important role by determining which conditional properties will and will not be instantiated.

## 9.2 – Non-Causal Mereological Determination

We have seen how the structure of the parts can make a contribution, but the remaining question is how can we conceive of this contribution as being non-causal? Hopefully the turn to determination rather than causation is not simply an *ad hoc* move designed to dodge the causal exclusion problem by a technicality. To be sure, there is ample reason to construe mereological structure as determinative, but non-causal. Let us begin by returning to Figure 3 in Section 3.2.



**Figure 3: The Problem of Causal Exclusion**

This diagram outlines the problem of causal exclusion for those who endorse supervenience. According to this diagram, the vertical arrows from the subvening physical properties to the supervening mental properties are not causal, rather they are arrows of determination (Kim, 2005, p. 36; Kim, 2005, p. 42-45). For this reason, it is not likely that mental properties are caused by their subvening bases, but rather they are determined by them.

Why does Kim mark this contrast between a horizontal causal relation and a vertical determinative relation? Kim explains:

What do I mean by “vertical determination”? Consider an object, say this lump of bronze. At any given time it has a variety of intrinsic properties, like color, shape, texture, density, hardness, electrical conductivity, and so on. Most of us would accept the proposition that the bronze has these properties at this time in virtue of the fact that it has, at this time, a certain microstructure – that is, it is composed of molecules of certain kinds in a certain specific structural configuration. I describe this situation by saying that the macroproperties of the bronze are vertically determined by its synchronous microstructure. The term “vertical” is meant to reflect the usual practice of picturing micro-macro levels in a vertical array, with the micro underpinning the macro. In contrast, we usually represent diachronic causal relations on a horizontal line (Kim, 2005, p. 36).

According to this passage, vertical lines are determinative because they represent a mereological relation between the lower level physical property and the higher level mental property. Mereological relations, therefore, are determinative, but not causal. Kim makes the same point on a number of different occasions as well (Kim, 1999, p. 32; Kim, 1998, p. 47-50). He even goes so far as to suggest that John Searle is one of the few contemporary philosophers who “gives ... highly idiosyncratic causal readings” (Kim, 1998, p. 47) to the relation between the physical cause and the supervening mental cause.

What reason is there to think that mereological relations are not causal? There are a number of reasons, each of which are rooted in Hume’s argument that causes and effects are separate and independent from one another (Hume, 1888, p. 78-80; See also Bechtel and Craver, 2007; Lewis, 2000, p. 78).<sup>51</sup> As we have seen, causation probably involves a bump or connection between two different things. This bump from one event to another is not possible if the two events are somehow amalgamated or combined

---

<sup>51</sup> Hume’s argument has had such an intuitive appeal that it led a number of philosophers to use the logical connection argument to demonstrate that mental events are not causal (Ryle, 1949; Melden 1961). The argument states that causes and effects are independent, but reasons are inseparably linked, so reasons cannot be causes.

(Bechtel and Craver, 2007, p. 552). This means that a cause must be temporally independent from, and more specifically, often prior to an effect. Kim hints at this in the aforementioned passage, where he says that causal relations are diachronic relations that occur across time. This also means that a cause must be spatially separate from an effect (Kistler, 2009, p. 604). Finally, and most basically, it means that a cause must be independent from the effect in the sense that the event that is a cause of an effect in some circumstances is not always a cause of the effect. Of course, there is a law-like relation between causes and effects, but this does not mean that some third event could not intervene in between the first event and the effect to prevent the effect from occurring, and ensuring that the first event is not a cause. For example, in normal conditions power outages may cause computers to shut off, but this causal relation is not guaranteed because a third event, such as the commencement of battery power may intervene and prevent the effect from occurring.

Mereological relations, contrary to causal relations, are intimately connected and dependent. As we saw in Section 4.2, nonreductive physicalists and Kim alike conceive of the relation between the subvening physical base and a higher level mental event as metaphysically, or at least nomologically, necessary and dependent. In the same way, the parts of a whole determine that there will be a whole, and we cannot have a whole without its parts. Not surprisingly, therefore, macro properties are determined synchronically by micro properties (Kistler, 2009, p. 598). Macro properties overlap in the same space as the micro properties do as well. ‘Being a rabbit’ is instantiated in the same space as the micro properties that compose the rabbit. And finally, macro properties are necessitated by micro properties. It is at least nomologically impossible to have the subvening physical event without the higher level mental event occurring as well. For these reasons we cannot consider the determinative relation from micro properties to macro properties as causal.

If micro properties non-causally determine the instantiation of macro properties because of the mereological relation between the two, then it is possible that macro

properties would non-causally determine the instantiation of micro properties because of this mereological relation as well. Clearly the relation from the macro property to its constitutive micro properties occur at the same time, in the same space and are necessarily connected as well (El Hani and Emmeche, 2000, p. 260; Gillett, 2006, p. 285). This being the case, we can begin to doubt whether a purported relation from  $M$  to  $P$ , or from  $M$  to  $P^*$ , or from  $M^*$  to  $P^*$  would be causal after all. According to the above diagram, causal relations are horizontal, and none of the relations from a mental state to a physical event is horizontal. Rather, the relations from mental states to physical events are either vertical or diagonal.

These considerations suggest a fairly simple solution to the problem of causal exclusion. Namely, the vertical arrow from the physical property  $P^*$  to supervening mental property  $M^*$  is not a causal arrow, so there is no reason to suppose there is causal competition between the causal arrow from mental property  $M$  to the later mental property  $M^*$  and the non-causal vertical arrow that states that the physical property  $P^*$  determines the occurrence of  $M^*$ . Furthermore, the line from the mental properties  $M$  and  $M^*$  to the physical properties  $P$  and  $P^*$  are not causal, but determinative, so there is no problem with  $M$  and  $M^*$  having an influence on  $P$  and  $P^*$  without violating causal completeness or causal exclusion (Sabates, 1996, p. 105; Jacob, 2002, p. 651; Marras, 2007, p. 309-310; Gibbons, 2006, p. 89).

Kim, of course, has no problem agreeing that in the case of upward determination, where the subvening properties ( $P$ ) determine the supervening properties ( $M$ ), we have a determinative relation rather than a causal relation. His concern rises when we attempt to suggest the determinative relationship is symmetrical in the sense that macro properties ( $M$ ) can determine the instantiation of micro properties ( $P$ ) as well. One of his concerns would be that downward mental determination violates causal exclusion since it seems to provide an additional cause on top of the sufficient physical cause. But, as noted, if the determinative relationship from the mental ( $M$ ) to the physical ( $P$ ) is not causal, there is no reason to suspect the principle of causal exclusion has been violated. The mental state

( $M$ ) plays a non-causal role in determining that the physical cause ( $P$ ) will occur, and this physical cause ( $P$ ) is then the sole sufficient cause of the effect ( $P^*$ ). Kim raises another concern against the sort of downward determination that I suggest here as well, called the principle of determinative exclusion, but I will save an analysis of this concern until section 10.2.

### **9.3 – Mereological Determination and the Co-Dependence Objection**

I have proposed that if we construe mental states as playing a non-causal determinative role we can secure a role for the mental while simultaneously endorsing causal exclusion and causal completeness. This solution requires macro to micro determination, which Kim rejects. I have dismissed two of Kim's objections to this form of mereological determination, but he also raises the co-dependence objection to mereological determination.

In his discussion with the emergentists, Kim labels the macro-to-micro form of mereological determination by the name of "reflexivity". Reflexivity, according to Kim, is when "a whole causes one of its micro-constituents to change in a certain way" (Kim, 1999, p. 29). Given our recently completed discussion on mereological determination, we should first of all amend this definition to read as follows: a whole *determines* one of its micro-constituents to change in a certain way. Kim levels two challenges at this doctrine of reflexivity. First of all, he argues that it is "circular and incoherent" (Kim, 1999, p. 28) to think that the parts determine the whole (from supervenience) while simultaneously believing that the whole partially determines the parts (from reflexivity). Kim articulates the difficulty of endorsing this model in a criticism of Roger Sperry's views about an eddy of water:

When each and every molecule in a puddle of water begins to move in an appropriate way – and only then – will there be an eddy of water. But in spite of this, Sperry says, it remains true that the eddy is moving the molecules around

“whether they like it or not”. Thus, reflexive downward causation is combined with upward determination (Kim, 1999, p. 27).

In the case of an eddy of water, the movement of the water determines that the eddy occurs, so the eddy cannot reciprocally determine that the water will move around. Or, to return to our wheel example, the arrangement of the parts determines that the wheel will be round, so we cannot also say that the wheel determines that the parts of the wheel will be round. Call this the Co-Dependence Objection.

Kim argues that we may be able to get around this problem by endorsing a mutual causal interdependence, or by supposing that reflexive downward causation occurs over time. A number of authors have taken this second route and used a diachronic model to respond to Kim’s concern here (El Hani and Quiroz, 2005, p. 169; Symons, 2002). However, given the fact that mereological relations are non-causally determinative in part because they occur at the same time, we cannot avoid the co-dependence objection by turning to a diachronic model of downward causation now.

Perhaps we can endorse the mutual dependence response. As mentioned, Kim finds this move incoherent. Part of his objection stems from what he calls the causal power actuality principle, which states that “for properties to exercise their causal/determinative powers they must actually be possessed by objects at the time; it cannot be that the objects are in the process of acquiring them at that time” (Kim, 1999, p. 29). In other words, for the spokes of a wheel to spin they need to have the property of ‘spinning’ at that time. However, if the spinning of the spokes is determined by the macro property of being a wheel, which is in turn determined to occur by the structure of the parts of the wheel, then it seems like the spokes did not already have the property of spinning when the spinning occurred.

In response, it is not clear that mutual dependence is incoherent. Mutual dependence is readily observable in nature. The heart depends on the brain, while the brain depends on the heart. Both depend on the lungs, though the lungs depend on them as well. All three depend on the nose, but the nose would not function without the brain,

heart and lungs. Any symbiotic relationship, which so often characterizes the constituents of life, relies on a certain degree of mutual co-dependence.<sup>52</sup> What we need, then, is not a demonstration that co-dependence does not occur, but an account of how to make sense of it.

There are two conceptual hurdles that we need to get over before we can have a working model of mereological co-dependence. First of all, Kim at times assumes that the mutual dependence is causal (Kim, 1999, p. 28). Causal co-dependence may be difficult to conceive of because of the necessary distinction between cause and effect that we have already seen. For example, a cause is usually prior to its effect, so there seems to be a one way relationship between the two, and it is therefore hard to imagine how the effect can have an influence on the cause. The dependency involved in the mereological relation is not causal, however, and this fact makes it easier to imagine the mutual dependence in two ways. First of all, the co-dependency is now conceived of as one of mereological determination where there is a fundamental tie between the parts and the whole. The intimacy of the mereological relation, where part and whole are neither spatially nor temporally distinct, makes it easier to imagine the mutual dependence.

Secondly, due to the mereological relation there can be no time lapse between the determinative relations, so Kim cannot argue that the parts need to have their property from the beginning of the event so they cannot be in the process of acquiring this property due to the influence of the macro property. To be sure, the parts need to have the property they instantiate at the beginning of the event, but due to the temporal overlap

---

<sup>52</sup> Of course, as we will see in a moment, Kim does not think these are examples of macro to micro co-dependence. However, since I define the macro in terms of the structure between the parts, this is exactly what mereological co-dependence amounts to. One part (i.e., the heart) depends for its existence on the structure it is placed within. We cannot imagine a heart continuing to function properly if it is taken out of the body and placed on a table. At the same time, however, the body depends for its existence on the heart working properly. If we took the heart out of the body, the body would no longer work. The part is dependent on its place in the structure, while the structure of the parts is dependent on the parts.

implicit in the mereological relation, the macro property must have determined that the micro part will instantiate this property from the beginning of the instantiation onward.

With these hurdles out of the way, we now need to present a plausible model of co-dependence. From Kim's example of the eddy cited above it is clear that Kim is still operating under the assumption that mental states are supervenient properties rather than structural properties. To be sure, Kim's discussion of the eddy occurs in a paper where he is criticizing the supervenient model of emergent properties. We have already rejected this supervenient emergentism in favour of structuralism, so it will come as little surprise to learn that Kim's objection is slightly off the mark.

According to structuralism, it is not a property that arises out of a configuration of the water molecules that determines that the water molecules will be swift and swirling. Obviously the eddy-like configuration of water molecules is sufficient to determine that the water molecules will be swiftly swirling, and so no higher level property is needed to determine that this will occur. Rather, on structuralism the structural property is the eddy-like configuration of water molecules. Of this we can truly say that the water molecules had the potential to be swiftly moving in a swirling manner, but they did not do so until they were in an eddy-like structure, so the structural property of being an eddy does determine that the water molecules will be swift and swirling. This is how a whole determines the properties of the parts while the parts simultaneously make up the whole.

Let us take stock. We entered this chapter faced with the problem that configurations of micro properties and relations continue to face causal exclusion pressure since the individual micro properties and relations are causally sufficient. I resolved this problem by demonstrating that configurations of micro properties and relations do play a role beyond that of the individual micro properties and relations. This role is determinative, however, not causal. That configurations play a non-causal determinative role arises naturally from the conclusion that configurations are mereological structures, which lend themselves to determinative influence, not causal influence.

## **Chapter 10 – Resolving Causal/Explanatory Exclusion**

In this final chapter I attempt to pull together the various parts of this dissertation and configure them into a response to the parallel problems of causal/explanatory exclusion.

In Section 10.1, I motivate and outline my doctrine of structuralism, and then demonstrate the model of mental causation that arises from it. I argue that structuralism gives us autonomous mental determination (i.e., a necessary, distinct role for the mental), though it does not strictly speaking give us mental causation. In Section 10.2, I consider how structuralism provides a viable response to the problem of causal exclusion. I argue that certain metaphysical considerations that we have looked at allow us to conceive of mental states as structures of parts which play a necessary and determinative role in the occurrence of certain microphysical causes. This determinative role is not causal, so the principle of causal exclusion is not violated, nor is the principle of determinative exclusion.

In order to fully resolve the problem of causal exclusion, we will need to consider how structuralism endorses and renders consistent the three other potentially inconsistent principles we first set out with. To this end, in Section 10.3, I argue that structuralism gives us causal completeness since mental states, and other macro properties, play a non-causal determinative role. Structuralism also gives us determinative completeness when the physical is construed broadly. Section 10.4 discusses structuralism and irreducibility. I argue that structuralism is nonreductive in nature, though it is perhaps not as robustly nonreductive as certain supervenient models. As evidenced in all of these sections, in order to achieve a reconciliation of these four principles via structuralism, each principle has to be slightly weakened. Nevertheless, I argue that structuralism still presents the most promising solution to the problem of causal exclusion.

After presenting structuralism as a response to the problem of causal exclusion, I turn back to the problem of explanatory exclusion. In Section 10.5, I present a nuanced form of the dual explananda reply to the problem of explanatory exclusion. This nuanced

model is rooted in structuralism, and it evades a critical objection to traditional dual explananda responses.

### **10.1 – Structuralism and Mental Causation**

I want to begin by summarizing the doctrine of structuralism. In chapter four I argued that it is possible for mental states to be necessary and determinative for a given effect without being causes for the effect. In chapter seven we began to see how this may be possible. Namely, we can say that an effect has background conditions that are necessarily present and even partly determinative while still insisting that they are not part of the sufficient cause of the effect. With this end in mind, I first of all argued in Section 7.4 that irreducibility implies a possible distinctness between the role that mental states and physical events play. Thus it is plausible that mental states can be determinative while physical events are causal. In chapters eight and nine this determinative role was fleshed out. Namely, mereological structure plays a non-causal determinative role. If mental states are mereological structures of their parts then we may have a non-causal determinative role for mental states.

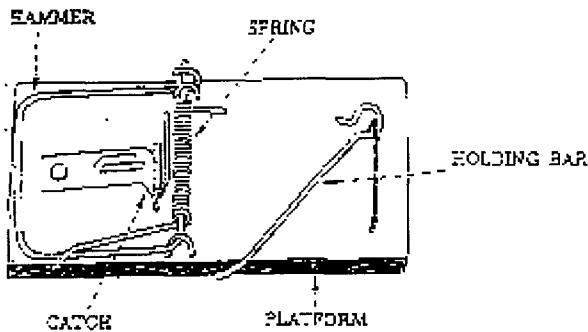
This non-causal determinative role is reached as follows. Mereological structure triggers or inhibits the instantiation of the conditional properties of its parts. These conditional properties of the parts in turn add new causal powers. There are a number of questions that are worthy of attention here. First of all, what does mental causation look like on the structuralist model? And secondly, and most importantly, does structuralism satisfactorily deliver mental causation for us?

#### ***10.1.a – The Structuralist Model of Mental Causation***

I want to first of all outline what mental causation looks like on the structuralist model, and then decide on whether we can be satisfied with the account of mental causation it offers. There are a number of different ways that structuralism can be applied to mental causation (El Hani and Emmeche, 2000, p. 267; Murphy, 2006, p. 232). First, and most

obviously, as ‘being a rabbit’ or ‘being a wheel’ is the structure of the parts of the rabbit and wheel respectively, so ‘being a mind’ will be the structure of the parts of the mind. Or, to shrink the model down, ‘being the belief that cake contains chocolate’ will be the structure of the parts of the belief that cake contains chocolate.

In order to refine this analysis, let us consider a simple macro object in detail, and then draw out the parallels to the case of mental causation. Let us use a mousetrap, as illustrated in Figure 6, as our macro object.



**Figure 4: Mouse Trap**

The property of ‘being a mouse trap’ is instantiated as the structure of a particular mouse trap. This mouse trap has a number of structural or functional parts. For example, there is a hammer, which plays the role of slamming down on the mouse. There is bait, which plays the role of luring the mouse. There is the trigger, which plays the role of unleashing the hammer. There is the holding bar, which plays the role of holding the hammer in suspended preparedness to strike. There are further functional properties of the mousetrap as well, but this will suffice for our purposes. Since these properties are strictly defined in terms of their roles within the system of the mousetrap, without regard for their physical realizers, it makes sense to consider these functional properties to be properties of the macro system, rather than properties of a particular realizer.

There are additional reasons to conclude that these functional properties are properties of the macro system, rather than properties of the realizing parts at this point. For example, a particular realizer, before and after being placed within the system, will not have the functional role property. For example, a piece of cheese, before being in a

mousetrap will not be bait, nor will it lure any mice. Rather, it will only be yummy food. This helps us to see that the functional role property is not necessarily a property of the particular realizer, but rather it is a property of the macro system. In addition, the functional role property is only loosely attached to a particular realizer, but it is essentially attached to the system. I can, for example, remove the cheese and replace it with peanut butter. If bait is a property of the cheese, the bait would presumably be gone once the cheese is gone. This is not the case, however, for the peanut butter can slide into this role. On the contrary, I cannot remove the hammer and the holding bar of the mousetrap and still have bait. If there is no hammer, the mouse will enjoy the cheese without reprisal, so the cheese is not bait. Since bait is integrally associated with the system, and only loosely associated with a particular realizer, we can see that this functional property is a property of a macro system. Moreover, functional properties are defined by the causal relations they bear to their inputs, outputs, and other parts of the system. Since functional properties are defined in terms of their interactions with other parts of the system, it makes sense to construe these functional properties as properties of this macro system, or as functional parts of this macro system. Finally, the multiple realizability of functional properties works well with the assumption that functional properties are properties of a macro system, where we can replace the micro objects and yet the macro object and its properties persist. For all of these reasons, it is plausible to conceive of functional properties as properties of a higher level system or object.

Although functional properties are properties of a macro system, there is also a sense in which these functional properties belong to whichever realizer is carrying out the specified task. This is evident from the above example of the cheese and the peanut butter. Although ‘being bait’ is a functional property of the mousetrap system, whenever cheese occupies this role, it seems right to conclude that the cheese has the property of being bait. In other words, the macro system “lends” its functional properties to whichever realizer is carrying out the specified task.

This “lending” can be clarified by returning to the distinction between intrinsic properties and conditional properties. Recall that intrinsic properties are properties that an object has at all times. Examples of intrinsic properties include mass and shape. Intrinsic properties can be contrasted with “conditional” properties (Shoemaker, 1980), or, in other words, “latent” properties (Shoemaker, 2002, p. 54), or perhaps, “potential” properties (McLaughlin, 1992). Conditional properties are properties that an object can possibly have, but does not necessarily have. As we saw, a person can have the property of being a professor, but does not have to have this property. A cardboard box can be a home, but does not have to be. Or, to return to our example, a piece of cheese can be bait, but does not have to be.

Since ‘being bait’ is a conditional property of the cheese, the cheese will only be bait at certain times. When will the cheese be bait? Plausibly, the cheese is bait when a mousetrap “lends” the functional property of being bait to the cheese. Or, in other words, the mousetrap determines that the cheese instantiates one of its conditional properties.

Whenever the cheese is properly arranged within the system of the mousetrap, it instantiates the conditional functional property of being bait in the mousetrap. Outside of its proper placement in a mousetrap it is not bait, and when it is properly placed in a mousetrap it is bait, so it stands to reason that the proper placement in a mousetrap determines that the cheese will instantiate the conditional property of being bait. This is how the macro system lends the functional property to the realizer.

This suggestion that a system, or a macro organism, can play a determinative role by triggering the instantiation of conditional properties of micro objects is not new. We have seen how Robert Van Gulick, for example, argues that “the pattern of which it is a part may affect which of its causal powers get activated” (Van Gulick, 1993, p. 251). Or, as Nancey Murphy affirmed, what I have been calling a system “can affect which causal powers of their constituents are activated or likely to be activated” (Murphy, 2006, p. 229. See also Sperry, 1986, p. 266; Shoemaker, 2002; Francescotti, 2007, p. 57ff; Gillett, 2006, p. 274). To return to a previous example, the macro system of a wheel determines

that the spokes and rubber will spin downhill. These spokes have the potential to move down a hill, but until they are placed within a wheel-like system, they will simply rest on the ground. Or, a wing has the potential to flap, but until it is suitably attached to a bird, it will not flap. The macro bird determines that the wing will be able to flap.

We are now in a position to evaluate the causal processes of the mousetrap.

Imagine that a mouse breaks his neck in the mousetrap. What causes the mouse to break his neck? The slamming of a metal piece onto the neck of the mouse causes the mouse's neck to break. But is the metal piece's slamming downward an intrinsic property, or a conditional property of the metal piece? The metal piece has many intrinsic properties such as its hardness and its shape, but it is not always slamming down. A metal piece placed randomly on the floor does not have the property of slamming downward, and since it does not have this property, the metal piece will not kill the mouse. Rather, the mouse may simply play with the metal piece, or push it around with its nose.

So, what determines that the metal piece instantiates the property of slamming down? Presumably, when the metal piece is appropriately placed in the macro system of the mousetrap, it will slam down. If the metal piece is not playing the role as a hammer in a mousetrap, and is not appropriately connected to a spring and a trigger, the metal piece will not slam down. To generalize this point, if a realizer does not play a certain role (i.e., does not instantiate a conditional functional property), then this realizer will not cause the effect in question. Functional properties of a macro system, therefore, play a role by determining that the realizer will play the role that it plays.

Let us try to translate this discussion on the mouse trap into the mental causation debate. On this model, mental states are functional structures of a system, or, mind. Mental states are defined in terms of their functional roles within the mind, and in terms of their inputs and outputs. Since these mental states belong to the macro mind, they can be realized by various different neural configurations. These neural realizers can do all of the causal work in bringing about a certain effect, but they will not do this work if they are not functioning as a certain role in the system. Thus, once they instantiate the role

property, or in other words, once they are appropriately placed within the system of the mind, the realizers are sufficient causes of the effect in question.

For example, Frank has a belief that broccoli is somewhat tasty, another belief that chocolate cake is delicious, another belief that he is on a diet, a final belief that broccoli is healthier than cake, a percept that there is chocolate cake and broccoli on the table, and a hunger for food. Frank's hunger for food has a certain neural realizer. This neural realizer has certain intrinsic properties, such as its mass and shape. It also has the conditional property of potentially functioning as Frank's hunger for food. This property is conditional because if it were removed from the system, if it were placed by itself on a scientist's table, it would not have this functional property anymore.<sup>53</sup> Once functioning as Frank's hunger for food, that is, once it is appropriately connected with Frank's belief that broccoli is healthier than chocolate, and Frank's percept that there is chocolate and broccoli on the table, this neural realizer is sufficient to cause Frank to eat broccoli.

Does this leave Frank's hunger for food with no role to play in generating Frank's eating of the broccoli? Seemingly not, for if this neural configuration is not functioning as this hunger, if the realizer does not have the role property because it is not properly arranged within the macro system of Frank's mind, then the realizer will not cause Frank to eat broccoli. The mental state of 'Frank's hunger for food' is a structural, or functional, property of Frank's mind which determines that the neural configuration will play the role of being Frank's hunger for food, which in turn determines that the neural

---

<sup>53</sup> Some have objected to functionalism on the grounds that there are law-like generalizations from the physical realizer to the effect, so the functional role property is not necessary (Malcolm 1968; Kim 1989). On higher level functionalism, however, there is a simple response to this objection. Namely, there is no law-like generalization from the neural configuration and Frank eating broccoli. As we have seen, if this neural configuration is placed on a scientist's table and prodded to fire, this would not result in Frank eating broccoli. Rather, this neural configuration, only after it is appropriately placed within Frank, bares a law-like relation with the effect of Frank eating broccoli. But, this added caveat states that this neural configuration, once instantiating the role property, stands in a law-like relation with the effect, so the role property is necessary in order to secure the nomological relation.

configuration will cause Frank to eat broccoli. The neural configuration has the potential to cause Frank to eat broccoli, but until it is appropriately placed in Frank it will not do so. We can once again say that the structure of the macro object (i.e., Frank) determines or triggers the realizer (i.e., neural configuration  $A$ ) to instantiate a conditional property (i.e., being Frank's hunger for food), which will then cause Frank to eat broccoli.

### ***10.1.b – Structuralism and Functionalism***

Structuralism bears some resemblance to contemporary role functionalism. Role functionalism, which is the original, and still the most popular form of functionalism, is different from Kim's model of functional reduction in an important way. Kim defines mental states in terms of their causal roles in a system, which is similar to what the role functionalist does. However, Kim then reductively identifies this role property with the realizer of this role. This move leaves Kim with a reductive model of realizer functionalism. Role functionalists, on the contrary, do not identify the role property with the realizing property. Rather, they leave the role property unreduced to the realizer property, even though they acknowledge that the realizer property is the property that plays the causal role specified by the role property.

As we have seen, structuralism supposes that there are role properties in macro systems, which are in turn realized by, but not reduced to, certain micro parts. Perhaps it is not surprising to see the similarity with contemporary role functionalism. After all, Hilary Putnam originally modeled functionalism on a similar distinction being drawn here. As he says “what we [functionalists] are really interested in, as Aristotle saw, is form and not matter” (Putnam, 1979, p. 302. See also Putnam, 1979, p. 279).

Functionalism investigates structure, which is exactly what structuralism does, so we can expect certain similarities.

This similarity between structuralism and role functionalism is important for two reasons: (1) role functionalism faces an objection that structuralism may face as well; (2) it sheds light on our question about whether structuralism delivers an acceptable form of

mental causation for us. I want to consider both of these points, beginning with the first. Jaegwon Kim is one of a number of people who raise concerns about the nonreductive version of role functionalism. The complaint is that the realizers can do all of the causal work, so the role property can be excluded (Kim, 1998, p. 51). Ned Block offers several examples of this problem. He wonders whether a bull gets angry because the cape is provocative, or because the cape is red (Block, 1980, p. 184)? According to Block, the role property of ‘being provocative’ can be excluded because the bull actually gets angry in virtue of the redness of the cape. Later on, Block points out that it is not the dormitivity of a sleeping pill which causes sleep, but the actual chemical in the pill (Block, 1990, p. 155). Similarly, Frank Jackson notes that it seems to be the molecular makeup of the glass, not the fragility of the glass, which causes the glass to shatter when it hits the ground (Jackson, 1996, p. 393). In all of these examples, the realizing property seems to carry all of the causal weight, so the role property can be excluded. For their own parts, a number of other commentators have noticed this same problem with role functionalism as well (Witmer, 2003, p. 200; Bennett, 2007, p. 330).<sup>54</sup>

<sup>54</sup> Kim’s solution, as we have seen, is to identify the role property with the realizer (Kim, 2006, p. 169). One of the reasons against this identity given in Section 5.2.a was that a role property cannot be identical to a realizer property. We can now see an additional reason as to why this identity is so difficult to sustain. Namely, the role property is not always a property of the same object as the realizer property. Let us say that the property of ‘being bait’ is identical to the property of ‘being tasty’. The mouse trap has a structural component of having a bait role, but it may not have the realizer property of ‘being tasty’. For two potentially identical properties to actually be identical, they must be instantiated by all the same objects at all the same times, and this is not the case here. Furthermore, cheese whiz can continue to instantiate ‘being tasty’ when it is removed from the trap, but is cannot still ‘be bait’ at that time, so again, we see that there is not only one property instantiated here, because they did not both persist. Finally, let us say that the property of ‘being bait’ is instantiated for ten hours in a row. For six hours the trap has cheese whiz on it, but it is placed outside in the rain and it slowly washes away. At the six hour mark some yellow plastic is placed in the trap as bait. By the eighth hour, all of the cheese whiz is washed away, but the yellow plastic still remains. ‘Being tasty’ was not instantiated for ten hours, but ‘being bait’ was, so it is not likely that this instance of bait is identical to the instance of tastiness.

Structuralism offers a solution to this problem. On structuralism, a role property is a property of a macro system or object, where this macro system determines that the micro part will play the role specified by the role property. Outside of the macro system, it is not necessarily the case that the realizer property would do the work it does, so the role property has a determining influence. Thus, on the traditional account of role functionalism, we posit that Frank's desire for food is the second order role property, while the neural configuration is the realizer. We then ask whether the neural configuration is sufficient for causing Frank to eat broccoli. The answer is yes, so the functional role property makes no causal contribution. On structuralism, however, Frank's desire for food is the functional property of the higher-level system of Frank's mind. Frank's desire for food will have a neural realizer as well. We can once again ask whether the neural configuration is sufficient for causing Frank to eat broccoli. This time the answer is still yes, but only if the neural configuration is appropriately connected to the other beliefs and desires that Frank has. The neural configuration, on its own, is not sufficient to cause Frank to eat broccoli. It is only after the neural configuration is properly arranged, or in other words, plays the role of hunger in Frank, that the neural configuration is sufficient to cause Frank to eat broccoli.

### ***10.1.c – Structuralism and Mental Determination***

Having demonstrated that role functionalism can overcome the causal exclusion problem if we adopt a structuralist conception of role properties, I now turn to the second reason for considering role functionalism in our context. Namely, there is a need to be certain that structuralism offers a satisfactory model of mental causation. At first glance, the answer seems to be no. Why? As argued above, there is a marked distinction between causation and determination on this model. Causation occurs at the micro level, while mereological structure merely plays a determining role. Since I have construed mental states as structures, this implies that the mental may play a determinative role, but not a

causal role. It may be disappointing to learn that structuralism only achieves mental determination, not mental causation.

Note, however, that non-causal determinative relations still make a significant objective difference. This was evidenced in Section 7.4, where background conditions such as the presence of oxygen were necessary and determinative even though they were not counted as part of the sufficient cause for the striking of the match.<sup>55</sup> This was also displayed in Section 9.2 when I noted the “upward” determinative relation that many agree occurs between physical events and their supervening events. In this case, these supervening events exist and make up a significant portion of our mental lives, so we can conclude that the determinative relation makes a significant objective difference here. Finally, and more relevant to the case at hand, it is apparent that the structure of the parts plays a significant role. If the leg of the bird were not properly attached to the rest of the bird, the muscles would not contract and thereby grab the worm – and this makes a significant difference to the worm at least. The same parts arranged in radically different ways are capable of doing radically different things. Clearly then, the move from mental causation to mental determination is not a significant demotion, for a significant influence is still evident.

There are several other tests that structuralism needs to pass before concluding that it satisfies the principle of mental causation. For example, I argued in Section 4.1 that mental states must be necessary factors in the occurrence of a given event. Many of the solutions we considered in chapter three could not offer a necessary role for mental states. The problem was that the physical cause was sufficient, which, by very definition,

<sup>55</sup> It is possible that the oxygen as a background condition is arbitrary and interest sensitive. For example, if we are filling the room with various gases to see whether matches ignite or not, we may consider the presence of oxygen as the sufficient cause, while the ignition of the match is the background condition. I take background conditions to be stable states that persist throughout the duration of a law-like causal relation between two generic events. Since generic events are nomologically related, we can pick these out as the causes, while background conditions persist throughout the occurrence of these events.

meant that the mental cause was not necessary. Certain positions attempted to dodge this problem by insisting that the mental event is necessarily present. Of course, the mental event can be necessarily present, but not play any role whatsoever. But, as soon as this mental event necessarily played a causal role, the physical cause was, by definition, not sufficient. As it turns out, structuralism dodges this problem. Structural properties are necessary because only a specific configuration of the parts guarantees that the effect in question will occur. Does this make the physical cause insufficient, since the mental state is necessary? No, since the mental state is non-causal, the physical cause still has to do all of the causal work.

In Section 2.1 reason was also given to prefer that mental states play a distinct role, or that autonomous mental causation should be preferred over reduced mental causation. Several of the solutions I considered in chapter three, as well as Kim's solution considered in chapter five, did not offer a distinct role for the mental. Rather, the mental event had the role of the physical event. On structuralism, mental states do make a distinct contribution, since the mental state does not reduce to its parts, nor does the determinative power of the mental state reduce to only being the power that its parts already have. Rather, the configuration is distinct from its parts, and though its parts do all of the causal work, the mental state plays a distinct role by triggering (or inhibiting) which micro properties will be instantiated.

I also concluded that the type of model of mental causation that we end up with must accord with the arguments that ground mental causation in the first place. In other words, the argument for mental causation from ethical grounding, epistemological grounding and common sense only support a version of mental causation that demonstrably secures ethical grounding and epistemological grounding, while simultaneously adhering to our common sense intuitions of what mental causation is. While a discussion on how structuralism delivers ethical and epistemological grounding takes us too far afield, it is important to address whether or not structuralism offers us an intuitive model of mental causation.

At first glance, structuralism may seem to offer a counterintuitive model of mental causation. After all, isn't it the case that mentality seems to "pop up"? For example, sudden decisions, or occurrent feelings of desire, or the sudden consciousness of beliefs seem to fit more properly as events which come and then go. Perhaps we run the risk of violating the principle of common sense by construing the mental as structures, or more generally stable states.

My response is much like Jaegwon Kim's response. On the issue of mental causation, Kim is not specifically interested in the qualitative or conscious aspects of mentality, such as an occurrent belief or desire. Rather, Kim is concerned with the causal, or functional role of these mental states. Functionalism, which is arguably the dominant model of mental causation available today, does this as well. Structuralism clearly follows suit here, and in so doing it is no more counter-intuitive than a currently dominant model, and to the extent that Kim endorses a form of functionalism, it is no more counter-intuitive than his position.

Placing structuralism within this common camp does not guarantee that the position is intuitive, however. It is possible that all of these models are all counterintuitive. With this in mind, it is important to point out how structuralism can be reconciled with a common sense view of the mental. It is not unusual to consider beliefs, desires, and even decisions as stable background determinants. For example, Lisa's unflinching belief that chairs are sturdy has disposed her since childhood to sit down whenever she is near a chair. Susan's longstanding desire to be a millionaire influences her to play the lotto every week, or to work at a less than desirable job. Larry's erstwhile decision to finish his bachelor's degree influences him to stay up late at night to study. To be sure, these beliefs and desires have qualitative properties as well, but an analysis of these qualitative dimensions is beyond the subject matter at hand. For my purposes here, it suffices to demonstrate that construing the mental in terms of mental states, rather than mental events, is compatible with common sense intuitions. I conclude, once again, that structuralism does not offer us the most satisfying version of mental causation. Sacrifices

have to be made in order to hold on to all four intuitions. Rather, it offers us necessary distinct mental determination, which seems to be substantial enough.

### 10.2 – Structuralism and Causal Exclusion

With the structuralist model in place, I want to now consider how it fares when confronted with the problem of causal exclusion. Perhaps the response that structuralism makes to the problem of causal exclusion can be clarified by mapping the structuralist solution onto Kim's original diagram of the problem of mental causation. I should note, however, that structuralism does not neatly fit within Kim's articulation of the problem of mental causation, so this method of plotting the structuralist solution onto Kim's diagram will fail in time. That being said, let us return to the original figure, but this time add the proposed structuralist solution in bold.<sup>56</sup>



**Figure 5: Causal Exclusion Solution**

According to Kim, causal exclusion pressures will arise when there are two arrows pointing at the same event. He thinks that we cannot allow an arrow from a mental event  $M$  to a later physical cause  $P^*$  because  $P^*$  already has a sufficient physical cause  $P$ . According to structuralism, there is no causal arrow from  $M$  to  $P^*$ , so the mental will not face exclusion pressures in this way. Rather, there is now an arrow from  $M$  to  $P$ , which attempts to express the view that  $P$ 's placement in the structure  $M$  determines that  $P$  will cause  $P^*$ . Or, in other words, when  $P$  has the functional role property  $M$ ,  $P$  will cause  $P^*$ . Or, again, when  $P$  is a part of macro system  $M$ ,  $P$  will cause  $P^*$ . It may seem like this leaves  $M$  with no role to play, since the causal arrow goes from  $P$  to  $P^*$ . It is true

<sup>56</sup> It is better to conceive of mental determination as involving one large downward arrow that persists for the duration of the causal process from  $P$  to  $P^*$ , or better yet, to conceive of determination as a surrounding background arrow.

that the physical event  $P$  causes  $P^*$  to occur. However, if  $P$  were not properly structured within the system of  $M$ ,  $P$  would not cause  $P^*$  to occur, so we cannot get rid of the mental state from the diagram.<sup>57</sup>

Presumably, however, there will be some previous physical cause  $P^*$  which is a sufficient cause for  $P$ . In this case there are two arrows pointing at the physical cause  $P$ . There is the prior physical cause  $P^*$  and the downward arrow from the mental state  $M$ . Exclusion pressures are supposed to arise whenever two arrows converge on the same event, so there may now be a problem. I have shown, however, that the arrow from mental state  $M$  to  $P$  will not be causal, so there is no need to worry about  $P$  having more than one sufficient cause. The previous physical cause  $P^*$  is the sufficient cause, while  $M$  plays a non-causal determinative role.

This solution will leave Kim unsatisfied, however, since he thinks that tension arises whenever two arrows converge on the same event, whether they are causal arrows

---

<sup>57</sup> There is also a question about how the relation from  $M$  to  $M^*$  will work on this model. Imagine that  $M_1$  is the mental state of desiring to not be thirsty,  $M_2$  is the mental state of believing water prevents thirst, and  $M^*$  is the mental state of desiring water. Presumably,  $M_1$  is realized as a structure of neural parts  $P_1$ ,  $M_2$  is realized as another structure of neural parts  $P_2$ , and  $M^*$  is realized as the specific structure of neural parts  $P_1$  and  $P_2$ . On the structuralist model, the neural structure  $M_1$  of parts  $P_1$ , together with  $M_2$  of parts  $P_2$  causes neural structure  $M^*$  of parts  $P^*$  to occur. Are the parts  $P_1$  and  $P_2$  sufficient to cause  $P^*$  to occur? Yes, but only if  $P_1$  is structured  $M_1$ -wise and  $P_2$  is structured  $M_2$ -wise and  $P^*$  is structured  $M^*$ -wise, where  $M^*$ -wise means that there is a specific structure between neural activity  $P_1$  and  $P_2$ . In other words, if the parts  $P_1$  are arranged randomly, then  $P^*$  will not occur, if the parts  $P_2$  are arranged randomly then  $P^*$  will not occur, and if the interrelation of the parts  $P_1$  and  $P_2$  is significantly different, then  $P^*$  will not occur. Thus, although the parts are sufficient to cause  $P^*$ , this is only true under the assumption that the structures  $M_1$ ,  $M_2$  and  $M^*$  are in place.

or not (Kim, 2005, p. 36).<sup>58</sup> Since there is a horizontal causal arrow pointing to  $P$ , we may have to exclude any downward arrow from  $M$ , whether it be causal or determinative. Unfortunately, here is where structuralism can no longer be diagrammed within Kim's picture. As noted, structuralism supposes that mental structures persist throughout the causal process from  $P$  to  $P^*$ , so it is not accurate to suppose that a determinative arrow from  $M$  to  $P$  will compete with a causal arrow from  $P^*$  to  $P$ , for the determinative arrow does not exactly go to  $P$ . If anything, mental state  $M$  should be diagrammed as one wide

---

<sup>58</sup> Kim has most recently made this point by appealing to Edward's Dictum, which states that "there is a tension between 'vertical' determination and 'horizontal' causation. In fact, vertical determination excludes horizontal causation" (Kim, 2005, p. 36). I have not framed this rebuttal through a use of Edward's Dictum for a number of reasons. First of all, it draws inspiration from Edward's view that the universe is literally re-created every millisecond. I take it that most people in the mental causation debate do not accept this view. Secondly, notice that Edward's Dictum suggests that determination trumps causation, so it would not be the mental state that faces exclusion pressure on my model of downward determination, but rather the horizontal causal arrow. For example, since the structure of the rabbit is sufficient to determine that a lump of flesh is a functioning heart at a given time, we could exclude a prior cause of the flesh's being a functioning heart at this time. This is not to suggest that this is my proposed solution. Rather, as becomes clear, I do not accept the fundamental tension between determination and causation that Kim attempts to motivate. I take it that determination and causation are different relations, so a sufficient cause does not rule out a necessary determinative contribution as well. For example, ten strawberries may be the sufficient amount of strawberries for a pie, but this does not exclude the required two stalks of rhubarb. I take it that Kim would argue that causation is a determinative relationship (Kim, 2005, p. 36), so the issue is that ten strawberries is the sufficient amount of fruit for a pie, which does exclude the addition of rhubarb. However, if the recipe stipulates that rhubarb is necessary, then clearly the proper interpretation should be that ten strawberries is the sufficient amount of strawberries, not the sufficient amount of fruit. In any event, I pursue many of these same issues through an analysis of Kim's principle of determinative exclusion, which presents a stronger threat to my position.

arrow that stretches from  $P^*$  to  $P$ , or even better,  $M$  should be a circle which surrounds the entire causal process from  $P^*$  to  $P$  on all sides.

This suggests a slight revision in our concept of “downward” determination, or “downward” causation, where any sort of arrow from the mental state  $M$  to the physical event  $P$  is conceived of as being downward, and occurring at roughly the same time as the causal arrow from the physical event  $P^*$ . Recall that the notion of downward causation, or downward directionality in general, was first raised in the discussion of supervenient emergentism. Supervenience theorists, as well as supervenient emergentists, conceive of mental properties as higher level properties that supervene on structures of parts. On structuralism, however, the emergent property is no longer considered to be supervening on the structure of parts, so any influence that structuralism suggests will not have a “downward” directionality. Rather, a mental state is the configuration of the parts, and this indicates that the mental state is on the same level as the physical events, surrounding the physical events. If a specific name is needed for this position, call it surroundward determination, rather than downward determination. The intuition that the mental ought to be construed as a surrounding determinative influence rather than a downward determinative influence, is predominant among a number of structural emergentists (Sperry, 1991, p. 230; El Hani and Emmeche, 2000, p. 263ff; Pihlstrom and El-Hani, 2002; Moreno and Umerez, 2000), and is also noted by others who have similar inclinations (Van Gulick, 1993, p. 252; Shoemaker, 2002, p. 58; Silberstein, 2001, p. 86; Silberstein, 2006, p. 204). This appeal to Kim’s diagram, therefore, although helpful as an introduction to how structuralism avoids the problem of causal exclusion, fails as we look at the specifics of a surrounding, or background, mental state.

With this in mind, I want to now consider the causal exclusion principle in its written form, in order to get a clear understanding of how structuralism fares when confronted with the exclusion principle. The principle of causal exclusion states, “No single event can have more than one sufficient cause occurring at any given time” (Kim,

2005, p. 42). Does structuralism violate this principle in some way? As noted, it does not violate this principle because the mental state does not play a causal role, so there will be no more than one sufficient cause for a given effect. We have seen that Kim does not think this response is entirely satisfying. After all, doesn't the exclusion problem simply arise again? Rather than facing the problem of causal exclusion, we now face something like Kim's parallel principle of determinative exclusion.

Determinative exclusion is motivated by the same intuitions as causal exclusion, but is a more general principle which states that "if the occurrence of an event  $e$ , or an instantiation of a property  $P$ , is determined/generated by an event  $c$  – causally or otherwise – then  $e$ 's occurrence is not determined/generated by any event wholly distinct from or independent of  $c$ " (Kim, 2005, p. 17). Here Kim captures the intuition that any time two arrows converge on one effect, whether they be causal or not, we will face pressure to exclude one of them. Does structuralism violate the principle of determinative exclusion? Determinative exclusion states that if an event  $p$  determines the effect  $p^*$  then no other event determines  $p^*$ . In our case, this means that  $m^*$  cannot determine  $p^*$ . We need to ask ourselves, therefore, whether event  $p$  determines effect  $p^*$  all by itself. According to structuralism, the answer to this question is both a yes and a no. I want to consider the 'no' answer first.

If event  $p$  is capable of determining that effect  $p^*$  will occur all by itself, then we could remove the surrounding structure and event  $p$  will still determine that effect  $p^*$  occurs. There is reason to think that event  $p$  is not capable of determining that effect  $p^*$  will occur on this scenario. To see this, imagine that we can re-arrange any structure that may surround event  $p$ , and ask whether effect  $p^*$  still occurs after this structure has been re-arranged or removed. The answer, some of the time, will be that effect  $p^*$  still occurs. Event  $p$  is a sufficient cause no matter what shape the surrounding structure takes. To use Kim's example, the table will weigh ten kilograms if it is table-shaped, or all of the blocks of wood are still on the carpenter's table. Thus the dropping of the table on the scale will cause the scale to measure ten kilograms no matter how the table is structured.

As we have seen, this is an example of what Armstrong calls a non-relational structural property. The structure does not matter – because mass is an intrinsic property, so the parts will have the same mass at all times, no matter how they are shaped.

Recall that Armstrong argues that relational structural properties also exist. Here the relations between the micro parts do make a causal difference. Is Armstrong correct in making the assumption that sometimes structure makes a difference? We have seen a number of examples where the structural relations between the parts do play a role. Thus, let us imagine that a heap-shaped arrangement of molecules is arranged puppy-wise. Once these molecules are arranged puppy-wise, let us say that one particular molecule makes up part of an amino acid in the collagen of the lower jawbone of the puppy. This being the case, does this molecule gain additional causal capabilities that it did not have in its previous amalgamation? Seemingly, it has the same causal capabilities based on its intrinsic properties such as mass, and it also has new causal capabilities. For example, it now has the ability to stay bonded in place when confronted with a fairly strong force (being part of a bone now, not a blob). It also now has the ability to be part of a lick, bark or chew, which each bring certain capabilities that it previously did not have. It often moves through space (whenever the dog moves), though it did not do this very often, or as swiftly, while clumped on the ground. The list could go on. Suffice it to say that the list of conditional properties that are only instantiated when structured appropriately is long and not usually contested.<sup>59</sup> This being the case, it seems as though

---

<sup>59</sup> Perhaps this is why Kim, when discussing the causal powers of parts and wholes, does not solely talk about the sufficiency of the micro-properties, but rather he talks about the sufficiency of the micro properties and relations among those micro properties as being sufficient. For example, micro-based properties are considered identical to “specific mereological configurations involving ... microproperties” (Kim, 1998, p. 118), they are not considered identical to these parts taken individually or as an arbitrary group. It is the micro particles and the relations that have causal power. Or again, Kim argues that emergent properties are excluded because there is a sufficient cause in terms of the micro properties *and* relations: “A whole,  $W$ , has emergent property  $M$ , where  $M$  emerges from the following configuration of conditions:  $W$  has a complete decomposition into parts  $a_1, \dots, a_n$ ; each  $a_i$  has property  $P_i$ ; and relation  $R$  holds for the

our previous event  $p$  will sometimes only determine effect  $p^*$  when operating within the proper structural arrangement  $m^*$ .

On the other hand, once we have a certain background structure in place, there is reason to think that event  $p$  is sufficient to determine that effect  $p^*$  will occur. In Section 10.1, I demonstrated how this is possible. Following this model, it seems fair to suggest that once the pipes are in place in a house, a simple flush of the toilet is a sufficient cause to make the water descend into the septic tank. Or again, once the transistors and resistors are properly arranged in a computer chip, pressing the ‘p’ key is sufficient for causing the letter ‘p’ to appear on the screen. Or, once the wing is in place as a part of the bird, the wing is sufficient to cause wind to blow. The structure does not make a contribution at the time when the toilet is flushing, so all of the causal power rests with the flushing of the toilet. The structure does not make a contribution at the time when the ‘p’ is being pressed, so it would be incorrect to assume that there is competition between the pressing of the key and the structure of the circuit board.<sup>60</sup> In this way, there is an

---

sequence  $a_1, \dots, a_n \dots$  A whole has a certain emergent property,  $M$ , at a given time,  $t$ , and the fact that this property emerges at  $t$  is dependent on its having a certain micro-configuration at  $t'$  (Kim, 1999, p. 28. Kim 1993, 161-165; Kim 1993: 164. See also Kim 1997: 279). The subvenient base includes parts with properties *and* the relation  $R$ , it is not simply the parts with properties. The subvenient base, because it includes these parts in that configuration, is sufficient, so the emergent property can be excluded. The inclusion of the configuration in the subvening base makes an appearance in Kim’s later discussion of Sperry’s emergentism as well. Sperry argues that the flying capabilities of the bird caused the parts of the bird to end up high in the sky where they wouldn’t have otherwise. When arguing against Sperry’s example, Kim still includes the appropriate configuration of the parts of the bird: “think of the bird’s five constituent parts, its head, torso, two wings, and the tail. For the bird to move from point  $p_1$  to point  $p_2$  is for its five parts (together, undetached) to move from  $p_1$  to  $p_2$ ” (Kim, 1999, p. 30). Here Kim argues that the five parts in the configuration of ‘being together and undetached as a bird’ is causally sufficient for the parts being high in the sky. However, he does not suggest that the five parts in a random arrangement is sufficient to explain why the parts are high in the sky.

<sup>60</sup> Consider how El-Hani and Emmeche articulate a similar position as it pertains to the causal exclusion problem (where  $M_1$  is a mental cause,  $M_2$  is a mental effect,  $Sp_1$  is a set of relations at the neuronal level at an early time, and  $Sp_2$ ,

important sense in which structure is necessary, but at the same there is a sense in which the immediate cause is sufficient as well. Or, put in a slightly different way, the mental state is not an event that occurs at the same time as the physical event, so it is true that there is no more than the one sufficient physical event that determines the effect at a given time. The mental state, however, is still necessary, for the reasons outlined above.

One may object that this position essentially entails that mental states are necessary, but at the same time microphysical states are sufficient causes on their own. This seems to be contradictory, especially in light of the arguments made in Chapter Four. Fortunately, the results from the investigation of overdetermination in Chapter Four are helpful here. Recall that I argued that we cannot say that some cause *p* is sufficient while some other cause *m* is necessary. We cannot say that we need no other cause than *p* while we simultaneously say that we need another cause other than *p*. However, some versions of epiphenomenalism state that a mental event necessarily appears when a physical event causes a given effect, though it does not causally contribute to the occurrence of this effect. Since the mental event does no causal work, the physical cause must provide enough force on its own to make sure that the effect occurs. In this case, the mental event is necessary for the effect to occur, but the physical

---

is a set of relations at the neuronal level at a later time): “Briefly, we are not claiming that the supervenience base of the thought  $M_2$ ,  $Sp_2$ , may be causally overdetermined, arising out of two distinct sufficient causes,  $M_1$  and  $Sp_1$ .  $Sp_2$  is indeed caused, in the efficient mode, by events that take place, without exception, at the micro-level (within the pattern  $Sp_1$ ). Nevertheless,  $Sp_1$  plus the environment or context, constrain the efficient events that take place at the micro-level so that the resulting pattern is  $Sp_2$ , and not another one. This provides a distinct causal explanation, in the formal mode, of  $Sp_2$ ’s instantiation” (El-Hani and Emmeche, 2000, p. 267-268; see also Gillett, 2006, p. 281-282; Symons, 2002, p. 200; El Hani and Quiroz, 2005, p. 176). Here El-Hani and Emmeche suggest that the structural property plays a formal causal role in bringing about  $Sp_2$ , which does not interfere with the complete efficient causality of  $Sp_1$ . Or, in my own words, the structural property here plays a determining role that does not interfere with micro causation. Since it is not causation, nor does it occur at the same time as the microphysical cause, it does enter into a competition with the microphysical cause.

event is still a sufficient cause of the effect. In this case we can say that some cause  $p$  is sufficient while some other event is necessary. We can say that we need no other cause than  $p$  while we simultaneously say that we need event  $m$ , since event  $m$  is not causal, so there is no contradiction. We are not saying we need no other cause than  $p$  while also saying we need another cause other than  $p$  in this situation, for  $m$  is not causal, so what we need is not another cause.

The problem with this epiphenomenal model, of course, is that it leaves the mental with no role. The strength of this model, however, is that it demonstrates that a mental event can be necessary for an effect while the physical cause is still sufficient for the effect. The inconsistency only arises when we say that we need only cause  $p$  and we need another cause as well. The inconsistency does not arise when we say that we need only cause  $p$  and another non-causal event, or a determinative event as well. This is what structuralism suggests. According to structuralism, mental states are necessary as well, for if a proper structure does not surround the physical cause, the effect will not occur. However, for all of the reasons listed above, mental states are not causal, which means that the physical cause must provide enough force on its own to make sure that the effect occurs. In this way, the mental state plays a necessary determinative role, though the physical event is still the sufficient cause of the effect. In other words, the physical cause is the sufficient cause, but this sufficient physical cause is not entirely sufficient on its own to determine that effect  $e$  will occur. This is, of course, simply a more paradoxical and controversial way of saying that causes are sufficient *ceteris paribus*. A cause is sufficient, but only after assuming certain other necessary conditions.

This solution will not be entirely satisfactory to some. Structuralism is consistent with the principle of causal exclusion, and largely with the principle of determinative exclusion, but there remains a sense in which the sufficient physical cause also requires the mental to play a determinative role. As noted in the previous section, however, in order to hold on to all four of our seemingly inconsistent principles, there is a need to water-down each principle slightly.

### 10.3 – Structuralism and Causal Completeness

We are now faced with the question of whether structuralism preserves causal completeness. Not surprisingly, the simple answer to this question is that it does preserve causal completeness because mental states are not causal, so the physical properties of the parts do all of the causal work, as required. It is possible, however, to reframe completeness in terms of determination. In this case the principle of determinative completeness will state that all physical events have a sufficient physical determinant, whether this determinant is causal or otherwise. Seemingly, this articulation of completeness will be supported by the same types of arguments that supported causal completeness.

Framed in this way, structuralism may appear to violate determinative completeness on account of the fact that the physical properties of the parts may not be determinatively sufficient, as the mental structure is a necessary determinant as well. I want to argue that the principle of determinative completeness, so defined, is not violated by structuralism. To see why, first of all recall the dual meanings of the term ‘physical’ that were outlined in Section 2.2. ‘Physical’ can mean microphysical, which implies that only the objects and properties referred to by the vocabulary of physics are physical. Alternatively, ‘physical’ can mean broadly physical, where certain macro objects and properties picked out by the special sciences can be included as physical as well. These macro objects and properties would include chemical, biological and possibly psychological phenomena as physical as well.

I want to briefly look at what happens when we endorse both definitions of the physical, beginning with the broadly physical definition. In Section 2.2, I briefly noted that Kim and many others in the debate on mental causation use physical in the latter, broader, sense. Kim, for example, argues that “physicalism need not be, and should not be, identified with microphysicalism” (Kim, 1998, p. 117). This means that macro objects which are made up of microphysical parts are still considered physical:

Perhaps, the standard micro-macro hierarchical model abets the idea that ultimately the physical domain only comprises the lowest microphysical level—only microphysical particles and their properties. But this is wrong. The physical domain must also include aggregates of basic particles, aggregates of these aggregates, and so on, without end; atoms, molecules, cells, tables, organisms, mountains, planets, and all the rest belong, without question, in the physical domain (Kim, 1997, p. 293).

Kim's intuition is that if macro objects, such as molecules and rabbits, are entirely constituted by microphysical particles, then what reason is there to exclude these macro objects from the physical world? For their own part, a number of commentators agree with Kim on this point (Huttemann, 2004; Papineau, 2008, p. 127-129).

The consequence of endorsing a broad definition of the physical is that physical determinative completeness is not violated. We have seen that on the broad definition of the physical, the structural properties and objects referred to by the special sciences are still physical. Chemical properties such ‘being a water molecule’, and biological properties such as ‘being a rabbit’ are included within the broadly physical framework. Mental states, as we have seen, bear the same mereological relation to their parts as these other special science macro objects do. Thus, we can conceive of mental structures as fitting within the broadly physical net. If mental states are broadly physical, then the determinative contribution they make will still not imply that a non-physical contribution is being made. Thus, broadly physical determinative completeness is not violated.

Let us briefly consider what happens if we endorse the microphysical version of the physical. On this model only the properties and objects referred to by microphysics can be counted as physical. Microphysics mentions sub-atomic particles such as quarks and bosons, and maybe even strings, but there is no mention of special science objects such as rabbits. This being the case, these macro objects are not considered physical. Since we have construed mental states along the same lines as other special science objects, mental states will not be considered physical either. Structuralism, however,

supposes that mental states and other macro structures play a determinative role, which may mean that all of the determinative power does not rest with microphysical properties and objects. In this way, structuralism appears to violate microphysical determinative completeness.<sup>61</sup>

This leaves us with a stark choice. We can accept broadly physical determinative completeness, which safeguards the necessary determinative role of all special science properties and objects, including mental states. Or, we can accept microphysical determinative completeness and leave all special science properties and objects, including mental states, without a determinative or necessary role. As we have seen in the prior sections of this chapter, and we will see again in the next section, in order to achieve some semblance of all of our four founding principles, we will have to weaken each principle somewhat. In this case, in order to preserve macro determination in general, and most specifically our central intuition of mental determination, we will need to define the physical in broad terms.

This leaves us in a position somewhat akin to the view that Carl Gillett takes. Gillett, a structural emergentist who we briefly encountered earlier, concludes that microphysical determinative completeness is false, but broad physical determinative completeness is true (Gillett, 2006, p.278-279). To the extent that broad physical determinative completeness is true, and physical causal completeness is true in general, we have not violated the principle of causal completeness. This may not be ideal, but as warned about in Section 2.1, we may have to settle on a position that is “near enough” to fulfilling the requirement of physical causal completeness if we want to hold on to all four of our principles. As it turns out, we were able to hang onto physical causal completeness and a broadly physical determinative completeness.

---

<sup>61</sup> Some structural emergentists argue that the completeness of micro physics is not violated, but rather it is “supplemented” (El Hani and Emmeche, 2000, p. 264). For further discussion see El Hani and Emmeche, 2000; Sperry, 1991, p. 230; and Emmeche et al., 1997.

#### **10.4 – Structuralism and Irreducibility**

Having seen the scathed, but successful adoption of the first three principles, I now want to turn my attention to whether structuralism succeeds in delivering irreducibility. There are those who think structuralism fails here, since it looks like a form of reductionism. Timothy O'Connor, for example, argues that since this model locates the emergent property among and between the base properties, then there is “not anything ‘over and above’ the micro-properties” (O'Connor, 1994, p. 93). The emergent property, on the structural model, would neither differ in kind, nor be on a higher level than the base, and for these reasons it may not be a form of non-reductionism (Broad, 1925; O'Connor, 1994, p. 94). O'Connor is certainly right to conclude that structuralism does not endorse supervenient emergent properties in the traditional way where the emergent property sits on top of its base properties. Instead of talking of higher level properties, it may be more accurate to speak of macro properties, or better yet, mereological relations.

However, there is still reason to think structuralism is nonreductive in nature. As we have seen, the structural property is the specific configuration of the micro parts, and we cannot reduce it to these micro parts without losing the important configuration of the parts (El Hani and Emmeche, 2000, p. 269). Thus, ‘vegetable’ is composed of the letters ‘v’ ‘e’ ‘g’ ‘e’ ‘t’ ‘a’ ‘b’ ‘l’ ‘e’ and the specific configuration or ordering. If we attempt to reduce ‘vegetable’ to its micro parts (i.e., letters), then there is no guarantee that these letters will appear in the specific configuration that composes the word ‘vegetable’. Rather, we could get ‘telebagve’. In order to ensure that we get ‘vegetable’ we need to add the specific configuration of the letters. This is, of course, true for an actual vegetable as well. We cannot reduce a cucumber to its parts because we could still have all of the parts present, though arranged differently, and not have a cucumber. To guarantee that there will be a cucumber we need to have the parts of the cucumber and the fairly specific configuration of the parts as well.

Additionally, the structural property is a property of a macro object, while the micro properties and relations between the micro parts are not properties of the macro

object. Thus, ‘being a rabbit’ is a structural property of the macro object ‘Bouncy the rabbit’. The property of ‘being negatively charged’ or the relation of ‘being one nanometer west of molecule 241’ are properties of and relations between other micro parts. As we have seen (Chapter 8, Footnote 3), a reductive identity can only go through if the same objects always have the identical properties at the same time. Since these properties are not instantiated in the same object at the same time, they are not identical. Finally, for all of the reasons listed in Section 8.4, we cannot reductively identify the structure of the parts with the parts of the structure.

What about O’Connor’s concern that structural properties do not differ in kind from micro properties, so this is hardly a nonreductive position. The position developed here, however, posits a considerable distinction between structural mental states and microphysical events. Mental states, as structures, are stable for at least the duration of the event, while events are changes. Mental states play a determinative role, while microphysical events play a causal role.

One can also object that I have at times contrasted mental states with physical events, which seems to indicate that mental states are not physical, but at other times I include mental states within the physical domain. Indeed, I do contrast mental states with physical events, and by this I mean to say that mental states are not microphysical. In this case, mental states are distinct from, and not reducible to the microphysical. On the other hand, in the last section I also argued that when the physical is conceived broadly, it is the case that mental states are physical, as they operate within the spatio-temporal flux and are wholly composed of microphysical particles. In this sense mental states do reduce to the physical, as they are within the physical sphere. In sum, mental states are not microphysical, but they are broadly physical.

I must conclude along the same lines as our previous three sections have ended. Namely, structuralism does not offer us as much non-reductionism as we may like. Certainly, supervenient emergentism and supervenience based physicalism offers a little more of a distinction between the mental and the physical; while more extreme positions

such as Cartesian dualism offer us even more irreducibility still. However, in order to satisfy the other three principles, this seems like as much irreducibility as possible.

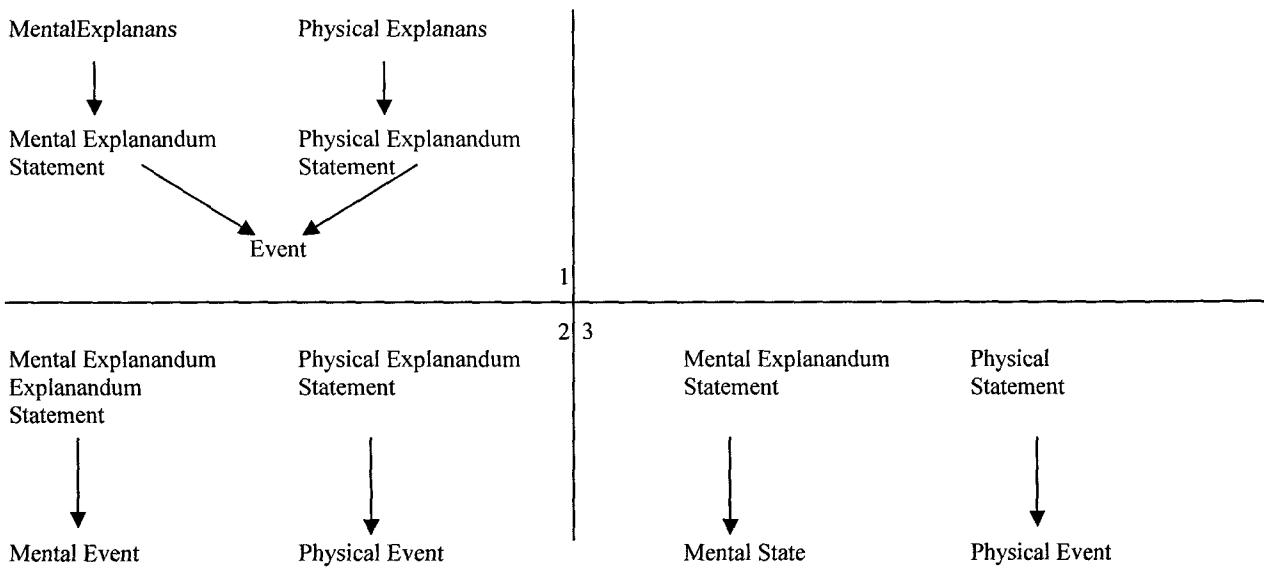
### **10.5 – Structuralism and the Problem of Explanatory Exclusion**

The causal exclusion problem was avoided by construing mental states as structures which play a determinative role in the causal power of the micro parts, but how is the explanatory exclusion problem solved? Recall that explanatory exclusion states that there can be no more than a single complete and independent explanation for any one event. Kim solves the problem generated out of explanatory exclusion by appealing to an extensional criterion of explanatory individuation. Mental descriptions and physical descriptions state the same explanation, and are about the same causal relation between events, so there is only one complete explanation for the given effect. This model failed because it supposes physical descriptions are complete descriptions of why the effect occurred, so we do not need mental descriptions of the same effect, so mental descriptions can be excluded (Section 5.4). Kim's model also fails because the extensional model of explanatory individuation is problematic (Section 6.3). For these reasons, we need to find another solution to the problem of explanatory exclusion.

An alternative popular response comes in the form of the Dual Explananda Reply. The dual explananda reply suggests that there are two different entities (i.e., events or explananda, depending on the version of the dual explananda reply one endorses) that need to be explained, a mental entity and a physical entity. These entities are distinct from each other, so they will each have a complete and independent explanation. Moreover, their distinctness eliminates any need for competition between them. Thus, there can be no more than one complete and independent explanation for any one thing, but the mental and physical are different things, so they can each have their own complete and independent explanations.

Perhaps the most popular version of the dual explananda reply can be termed the triangular dual explananda reply (Davidson, 1993, p. 16; Marras, 1998, p. 448;

Fuhrmann, 2002, p. 195; Campbell, 2008, p. 84; Liu, 2001, p. 16; McIntryre, 2002, p. 95). This version posits two distinct explanandum statements, one mental and the other physical, of the same event. These two explanandum statements are conceptually inequivalent, even though they both describe the same event. Once we have two different explananda, it is natural to suppose that there will be two different explanans statements as well, which explain their respective explananda statements. Thus we have two different explanations which refer to the same event. As Donald Davidson summarizes, “there is no reason why logically independent explanations cannot be given of the same event” (Davidson, 1993, p. 16). In Figure 7, Box 1, I illustrate this solution.



**Figure 6: Versions of the Dual Explananda Reply**

This triangular dual explananda reply can address the problem of explanatory exclusion in several ways. Marras, for example, argues that explanations are essentially intensional (See Section 6.3), so we need to rephrase the principle of explanatory exclusion in terms of explananda, not events (Marras, 1998, p. 449). Once we have framed the exclusion problem in terms of explananda, we can point out that there are in fact two distinct explananda that require an explanation. The mental explanation can involve an explanandum framed in mental terms, while the physical explanation can involve the explanandum framed in physical terms, and the apparent explanatory competition is dissolved.

Alternatively, we can point out that although the two explanations are conceptually inequivalent (and so are conceptually independent), they describe the same causal relation between events (and so are ontologically dependent). This being the case, we can argue that the two conceptually distinct explanations are still ontologically dependent, so there are not two independent explanations and exclusion pressures do not apply. Fuhrmann, for one, makes such a move:

Although we have assumed, with good reason, that the explanantia  $a$  and  $a^*$  must be rendered distinct, there is not the slightest reason to suppose that therefore the causes  $c_a$  and  $c_{a^*}$ , which ground  $a$  and  $a^*$ , must also be distinct (Fuhrmann, 2002, p. 195).

Fuhrmann goes on to say that the identity between the causes renders the two explanations dependent (Fuhrmann, 2002, p. 196. See also Campbell, 2008, p. 86). The strength of this move is obvious. Kim says there can be no more than a single independent explanation of an event, but these explanations are now ontologically dependent, so Kim's principle has not been violated.

There are a number of reasons to prefer not to use the triangular dual explananda reply. First of all, Marras' solution relies on redefining the principle of explanatory exclusion. Although there may be legitimate reason for this re-construal, it is still clear that Kim's original principle of explanatory exclusion continues to be violated. After all, there is still one event which has two different explanations, and according to Kim's version, this means that the mental explanation can be excluded. As for Fuhrmann's solution, it posits the conceptual independence of the explanations with the ontological dependence of the underlying causes. If we insist that there are two explanations because of their conceptual independence, then we face the exclusion pressure that comes from having two explanations for one event. If we reply that there is one explanation because of their ontological dependence, then this invokes an extensional model of explanatory individuation, which is problematic (Section 6.3).

There are certain general worries about the triangular dual explananda reply as well. As we saw in Section 5.4, it would be better to need mental explanations, rather than to simply add them. However, if we already have a complete physical explanation of a given event, we don't really need to add a mental explanation of the same event as well. The triangular dual explananda reply, which posits both a mental and a physical explanation for the same causal relation between events, faces this difficulty. The models presented below will make mental explanations a necessity; which counts as an argument in their favour.

Furthermore, in Section 6.3 we discussed the desire to achieve as strict a correspondence between descriptions and reality as possible. There were certain causal and explanatory benefits to insisting on a mirroring or tracking relation between descriptions and reality. Thus, imagine that there is one world. Let us say that this one world, in one case, acts as follows: event  $P$  caused event  $P^*$ . We can achieve a strict correspondence between this reality and our language by simply describing this causal relation between events as ‘event  $P$  caused event  $P^*$ ’. We lose some of the strength and simplicity of this correspondence by describing this causal relation as ‘event  $P$  caused event  $P^*$ ’ and ‘event  $M$  caused event  $P^*$ ’ and ‘event  $P$  caused event  $M^*$ ’ and ‘event  $M$  caused event  $M^*$ ’. If we seek to preserve as strict a correspondence between the world and our language as possible, then we should not prefer a triangular model which posits two or more explanations for one causal relation between events.

Given these difficulties with the triangular dual explananda reply, it is best to consider certain other versions of the dual explananda reply. Another popular variety of the dual explananda reply is rectangular in shape, and it posits two different explananda which refer to two different objective states (Dretske, 1988; Dretske, 1993; Dretske, 1995; Sturgeon, 1998, p. 424; Baker, 1998; Pereboom, 2002). The central proposal here is that when Kim's principle states that there can be no more than one complete and independent explanation of any one event, this reply suggests that there is no problem because there are actually two independent states that require explanation.

Fred Dretske, one of Kim's first respondents, deploys this tactic as outlined in Box 2 of Figure 7. In various writings, Dretske marks a difference between a triggering cause and a structuring cause. The triggering cause denotes the fact that  $C$  caused  $E$ , while the structuring cause points to "what caused  $C$  to cause  $[E]$  rather than something else" (Dretske, 1988, p. 42). Using Dretske's example, the triggering cause for Clyde to rise to his feet is the fact that the queen entered the room. However, the structuring cause explains why he stands to his feet rather than, say, claps his hands. In terms of the mental causation debate, the triggering cause is the neurological event in the brain. However, the structuring cause invokes an intentional relation with the external environment. Thus, in Kim's words,

Dretske's solution to the exclusion problem is an instance of what may be called the 'two explananda' strategy: rationalizations and biological explanations do not share the same explananda, and therefore there need be no explanatory competition between them (Kim, 1995, p. 133).

Since rationalization explanations refer to structuring causes while biological explanations refer to triggering causes, no exclusion pressures emerge.

Jaegwon Kim critically examines Dretske's proposal in "Explanatory Exclusion and the Problem of Mental Causation", and concludes that there is a serious problem:

More concretely, consider Dretske's [structuring] causings, the supposed explananda of psychological explanations. We can put the issue in a simple and stark way: are these causings physical entities or are they not? If they are not, we have an overt dualism. If they are in the physical domain, are they susceptible of physical causal explanations or are they not? If they are, then these explananda, special though they might be, cannot serve to separate psychology from physical theory, and the exclusion problem arises again. If they are not, we would again have a form of psycho-physical dualism (Kim, 1995, p. 134).

In essence, if Dretske's solution is to double the number of explananda, then Kim's reply is to double the problem by imposing fresh exclusion pressures on both explananda individually. The original physical event, everyone agrees, has a physical causal explanation. The original mental event, although it is a different physical state from the first, is still a physical state, and so this second physical state will have a physical causal explanation as well. If this is granted, then the mental causal explanation is again a second explanation that stands in danger of exclusion.

Dretske responds to Kim's argument by acknowledging that, "Kim is right. As he often is ... as long as what reasons are called upon to explain is a physical phenomenon of some sort ... we will have questions about explanatory exclusion" (Dretske, 1995, p. 142). Dretske agrees that intentionality can be naturalized (Dretske, 1995, p. 148), thus his structuring causes have a complete physical explanation, and the mental explanation is excluded again.

Several authors have attempted to offer up a more nuanced version of the rectangular dual explananda reply that avoids Kim's rebuttal (Sturgeon, 1998; Baker, 1998). Central to this tactic is the suggestion that the mental state is a higher level physical state, which is ultimately irreducible to microphysics. Thus, Kim is not able to create the problem anew by stating that the mental will have a complete causal story in terms of microphysics, for it is not clear that it straightforwardly does. Scott Sturgeon, for example, distinguishes between the bottom quantum level and the broadly physical level that includes such things as handshakes and the felling of trees (Sturgeon, 1998, p. 416). He then argues that mental states cause broadly physical effects such as handshaking, but they do not cause quantum effects. Thus there is no competition between the mental and the microphysical due to the fact that they cause different effects.

Sturgeon anticipates Kim's attempt to double the problem when he notes the possibility that "broadly physical macro events are composed of quantum events...hence [they] causally compete" (Sturgeon, 1998, p. 417). The mental is once again in danger of exclusion since the broadly physical states are composed of quantum events, which have

a full causal story themselves. Whereas Dretske fell at this point, Sturgeon dodges the problem by invoking the irreducibility of the macrophysical domain:

The conceptual divide between quantum and macro reality is great. We should be cautious in accepting unrestricted causal flow between the two. For we cannot see how quantum events build into the splendour of macroreality. We cannot see, for example, how quantum tunnelling *could* build into jalapeños peppers (Sturgeon, 1998, p. 424. See also Sturgeon, 1999, p. 378-379).

Although it may be a valid assumption, there is as of yet no way to be certain that the quantum reality takes up all of the causal power of the macrophysical event. Kim's problem, therefore, rests upon a conjecture, and so long as it does there is no reason to fear.

Although Kim has not directly addressed Sturgeon's concern, Agustin Vicente has responded in a way that Kim would most likely find acceptable. Vicente begins by noting that the traditional rectangular dual explananda reply that Dretske invokes will double the problem: "The problem is none other than the exclusion argument again" (Vicente, 2002, p. 78). He then notes that Sturgeon has attempted to avoid this pitfall by pointing to the gap between the broadly physical and the quantum mechanical which makes the one irreducible to the other. The problem with this move, according to Vicente, is that a physicalist must accept the ontological reduction of the higher level to the lower level:

Now given that physicalism is the thesis that all entities are reducible to the entities postulated by a true future physics, this possibility would undermine Sturgeon's argument. So he needs a further argument motivating constant concept separation (Vicente, 2002, p. 84).

Vicente thinks that it may be reasonable to admit of a conceptual gap between the higher level and the quantum level, but it is not possible for a physicalist to admit of this ontological gap between the two. Kim, in his insistence that an 'overt dualism' would follow if Dretske were to deny that the mental was physical, seems to demonstrate a

similar perspective. Not surprisingly, Vicente suggests that Sturgeon is not a physicalist (Vicente, 2002, p 77-79). For his own part, Sturgeon does not deny physicalism, but he acknowledges that his views would lead to “a truly modest physicalism … physicalism can be weaker than previously supposed” (Sturgeon, 2001, p. 205).

The structural model presented here is similar to Sturgeon’s model in some respects. Namely, it suggests two different states. The physical explananda refers to a microphysical state of particles and their properties. The mental explananda refers to a specific mereological structure of these parts. The strengths of this model are that it preserves all of the things lost on the triangular model. For example, it preserves the necessity of mental explanations, and the strict correspondence between descriptions and reality.

The weakness of this model, according to Kim and Vicente, is that it may not be physicalism anymore. They argue that an ‘overt dualism’ follows for anyone who endorses a rectangular dual explananda reply. But this conclusion only follows if the debate is between whether the world is made up of only microphysical particles, or the world is made up of microphysical particles and ghostly souls. However, these are not the only ontological options under consideration here. We have considered a layered ontology which posits irreducibly structured macro objects made up of micro parts. From within this ontology it is natural to expect to find microphysical parts and irreducible structures of these parts.

Let us consider how the objection that Kim and Vicente raise fares against this backdrop. First of all, I have suggested that there is a physical explanation and a mental explanation, both of which refer to different states in the world. Nobody objects that the physical state has a complete physical explanation. The mental state, however, is a distinct state from this physical state, so it requires its own mental explanation. At this point Kim will object that this second state also has a complete physical explanation, so we can exclude the mental explanation. There are a number of responses here. First of all, given that macro level mental states are included within the physical realm, it is

obviously true that the mental explanation has a complete physical explanation – the mental explanation is itself this physical explanation. Consider the following explanation: the bird flew because it saw a lion. This is a macro level physical explanation, which refers to the macro level structures of being a bird, and being a lion. Should we say that this explanation is excluded because there is a complete physical explanation of it? No. Presumably birds, lions, and flying are physical states, so it is already a physical explanation. In the same way, we cannot exclude the mental explanation on account of the fact that there is a complete physical explanation for the same state, since the mental explanation is already a physical explanation.

Vicente (and, possibly, Kim) may have a slightly different concern in mind. Namely, if the mental state has a complete microphysical explanation, then we can exclude the mental explanation. It is surely true that if the mental state has a complete microphysical explanation then we can exclude the mental explanation. And, as I have argued, it is surely the case that the mental state does have a complete microphysical causal explanation, since all of the causation occurs on the microphysical level. However, microphysical explanations, given solely in terms of microphysical particles and their properties, which do not mention the specific structure of those parts, are not complete explanations. A reason for this, again, is that these same parts and properties could be arranged differently, which would lead to them not generating the effect in question. Therefore, in order to have a complete explanation for the effect in question, we need to include both the parts and the structure of the parts. The mental explanation refers to the structure of the parts, which determines that the parts will act as they do, so it is a necessary ingredient in a complete explanation.

Perhaps, however, we can give a complete microphysical explanation of the structure of the parts, thereby excluding the mental explanation. The structure of the parts, however, is a specific configuration, it is not many individual relations. Microphysics deals with individual relations between particles, it does not deal with macro level configurations. Consider, once again, the explanation: the bird flew because

it saw a lion. The configuration of ‘being a bird’ is not directly mentioned in the vocabulary of microphysics, nor is the configuration of ‘being a lion’, or ‘flying’. If we give a long microphysical explanation in terms of one billion particles and their individual gravitational and magnetic forces, then we have given an explanation of one billion particles and their individual gravitational and magnetic charges, we have not given an explanation of being a lion. In so far as being a lion shaped structure of these billion parts is necessary in the explanation for why the bird flies, the microphysical explanation that deals with individual particles in individual relations will not be complete.

### 10.6 – Conclusions

What conclusions can be drawn from our investigation? We started out with four fixed assumptions: physical causal completeness, mental causation, irreducibility and causal exclusion. These four assumptions seemed to form an inconsistent tetrad. We tried to resolve this tension via some currently popular schools of thought such as supervenience, emergentism and constitution. As it turned out, these models either violated the principle of causal exclusion, or could not deliver a necessary distinct role for mental states. We then tried to reject the principle of causal exclusion, but we could not do so because it led to massive coincidence and the failure of the mental (and, the physical) to be necessary. We tried to finesse through the more nuanced version of dependent overdetermination, but this rendered the mental (and, the physical) independently insufficient, which violated physical causal completeness. We then attempted to follow Kim and reject irreducibility. As it turned out, however, there were problems with making the reductive identity hold. Moreover, at best this model provided us with reduced mental causation, and at worst we found reason to believe that Kim’s reductionism actually slid into eliminativism, which failed to secure mental causation. Mental causation could not be rejected because it provided a foundation for ethics and epistemology, and it made sense of our everyday

experience. The argument for physical causal completeness was supported by the success of science argument, as well as various conservation laws in the sciences.

Having ruled out these popular resolutions, and the rejection of any one intuition, we had to turn to more exotic responses to the causal exclusion problem rooted in metaphysical considerations. We soon saw that we could simply reject Kim's model of events, and the problem disappears. However, this move involves a rejection of properties, which in turn diminishes the amount of realism we can achieve by rejecting the causal model and endorsing a sort of pragmatism. Alternatively, we could simply reject Kim's model of causation, and the problem once again disappears. This move, however, gives rise to counterfactual or nomological accounts of causation which, among other things, cannot demonstrate that epiphenomenal shadows are not actual causes.

On the positive side, in our discussion on necessity and sufficiency in chapter four we saw that it was possible for a mental state to be necessarily determinative while another physical event was still a sufficient cause. This possibility was repeated in chapter seven where we saw that the object in which an event occurs was a necessary and influential factor in the occurrence of any event. We further realized that the distinction implicit within irreducibility implied that mental states played a different sort of role than physical events. These three insights prepared us for the conclusion that mental states can be necessarily determinative while physical events remain causally sufficient.

This potential for a determinative role for mental states while simultaneously endorsing the sufficient causal role of physical events came to fruition when we observed the nature of mereological structure in chapter eight. We saw that mereological relations were not causal, since they were spatiotemporally conjoined. Mereological relations, however, still played an important determinative role. This determinative role flowed from the micro to the macro, as was the case when a subvening base determined that supervening properties would be instantiated. The determinative role also flowed from macro to micro, as was the case when the structure of a whole determined that certain micro properties of the parts would, or would not, be instantiated.

This model was called Structuralism. Structuralism resolves the problem of causal exclusion as follows: the distinctness implicit in irreducibility suggests that mental states may play a different sort of role than physical events. Causal completeness suggests that physical events are sufficient causes, while causal exclusion states that these physical events will exclude any other event or cause. Thus, if mental states play a role, mental states will have to play a non-causal determinative role. The doctrine of mental causation, however, suggests that mental states will play some sort of a role, so mental states must make a non-causal contribution. Mereological relations are one of the few sorts of determining relations that are not causal, so structuralism looks like a plausible way to reconcile all four of these intuitions. Structuralism claims that mental states play a determinative role by triggering or inhibiting certain micro properties of its parts from instantiating, while simultaneously endorsing the view that the micro properties of the parts are causally sufficient. Structuralism also affords an answer to the problem of explanatory exclusion. Mental explanations refer to mental states, which are distinct states which require a distinct explanation from the physical explanation.

Structuralism allows us to endorse all four intuitions, though in a slightly diminished capacity. Mental causation is salvaged, but it takes the form of mental determination. Causal completeness is salvaged, but determinative completeness is only salvaged by adopting an inclusive definition of the physical. Causal exclusion is salvaged, but not determinative exclusion. Irreducibility is salvaged, though mental states are no longer conceived of as ‘over and above’ the physical, but rather are ‘among and between’ the microphysical. Perhaps some of these concessions are not acceptable in the end, but at least it secures our four intuitions and seems like the best option among the many considered.

## References

- Alexander, S. (1920). *Space, Time, and Deity*, Vol. 2. London: Macmillan.
- Antony, L. (1989). "Anomalous Monism and the Problem of Explanatory Force," *Philosophical Review*, 98, p. 153-87.
- Antony, L. (1999). "Making Room for the Mental". *Philosophical Studies*, 95: 37-44.
- Antony, L. (2007). "A Defense of Nonreductive Materialism". *Contemporary Debates in the Philosophy of mind*. McLaughlin, B. and Cohen, J. (eds). Victoria: Blackwell.
- Antony, L. (2008). "Multiple Realization: Keeping It Real". *Being Reduced*, eds., Hohwy, J. and Kallestrup, J. Oxford: Oxford University Press. Pg. 164-175.
- Aranyosi, I. (2008). "Excluding Exclusion". *Philosophical Explorations*, 11, p. 67-78.
- Aristotle. *The Basic Works of Aristotle*. Ed McKeon, R. Random House, New York, 1941.
- Armstrong, D. (1978). *A Theory of Universals*, vol. 2. Cambridge: Cambridge University Press.
- Armstrong, D. (1989). *Universals: An Opinionated Introduction*. Boulder: Westview Press.
- Armstrong, D. (1997). *A World of States of Affairs*. Cambridge: Cambridge University Press.
- Ayer, A. J. (1997). "Freedom and Necessity". *Readings in Philosophy Series*. Pereboom, D. (ed). Hackett Publishing.
- Baker, L. (1993) "Metaphysics and Mental Content". *Mental Causation*, Ed. Heil, John and Mele, Alfred. Oxford: Clarendon Press. Pg. 75–95.
- Baker, L. (1997). "Why Constitution is not Identity". *Journal of Philosophy*, 94, p. 599–621.
- Baker, L. (1998). "What we Do". *Human Action, Deliberation and Causation*, Ed. Bransen, Jan and Cuypers, S. E. Pg. 249-270
- Baker, L. (2000). *Persons and Bodies: A Constitution View*. Cambridge University Press.

- Baker, L. (2009). "Nonreductive Materialism". *The Oxford Handbook for the Philosophy of Mind*, B. McLaughlin, S. Walter and A. Beckermann (Eds.). Oxford: Oxford Univ. Press.
- Baxter, D. (1988). Many-One Identity. *Philosophical Papers*, 17, p. 193–216.
- Bealer, G. (2007). "Mental Causation". *Philosophical Perspectives*, 21, p. 23-54.
- Bedau, M. (1997). "Weak Emergence". *Philosophical Perspectives*, 11, p. 375-399.
- Bennett, J. (1988). *Events and Their Names*. Hackett Publishing: Indianapolis.
- Bennett, J. (1991). "Reply to Reviewers" *Philosophy and Phenomenological Research*, 51, p. 647-662.
- Bennett, K. (2003). "Why the Exclusion Problem Seems Intractable, and How, Just Maybe, to Tract it", *Noûs*, 37, p. 471-49.
- Bennett, K. (2007). "Mental Causation". *Philosophy Compass*, 2, p. 316–337.
- Bennett, K. (2008) "Exclusion Again". *Being Reduced*. Hohwy J. and Kallestrup J. (Eds.). Oxford: Oxford University Press, p. 280-305.
- Berger, A. (2003). "The Quinean Quandary and the Indispensability of Nonnaturalized Epistemology". *Philosophical Forum*, 34, p. 367-382.
- Bickle, J. (1998). *Philosophy and Neuroscience*. Dordrecht: Kluwer Academic Publishers.
- Block, N. (1978). "Troubles with Functionalism". *Minnesota Studies in the Philosophy of Science IX*, ed. C. Wade Savage. Minneapolis: University of Minnesota Press, p. 294–296.
- Block, N. (1990). "Can the mind change the world?" *Meaning and Method: Essays in Honor of Hilary Putnam*. Cambridge: Cambridge University Press.
- Block, N., and Stalnaker, R. (1999). "Conceptual Analysis, Dualism, and the Explanatory Gap". *The Philosophical Review*, 108, p. 1–46.
- Block, Ned (2003). "Do Causal Powers Drain Away?" *Philosophy and Phenomenological Research*, 67, p. 133-50.
- Block, N. (forthcoming). "Functional Reduction". *A Festschrift for Jaegwon Kim*. Terry

- Horgan, Marcelo Sabates and David Sosa (eds.).
- Boghossian, P. (1990). "The Status of Content". *Philosophical Review* 99, p. 157-84.
- BonJour, L. (1992). "Against Naturalized Epistemology". *Midwest Studies in Philosophy*, 9, p. 283-300.
- Bontly, T. (2002). "The supervenience argument generalizes". *Philosophical Studies*, 109, p. 75-96.
- Bontly, T. (2005). "Exclusion, Overdetermination, and the Nature of Causation". *Journal of Philosophical Research*, 30, p. 261-282.
- Bregant, J. (2003). "Mental causation: Jaegwon Kim's dilemma". *Filozofski Vestnik*, 24, p. 219-240.
- Bregant, J. (2003). "The Problem of Causal Exclusion and Horgan's Causal Compatibilism". *Croatian Journal of Philosophy*, 3, p. 305-320.
- Broad, C. (1925). *The mind and its place in nature*. London: Routledge and Kegan Paul.
- Broadbent, A. (2008). "The Difference between Cause and Condition". *Proceedings of the Aristotelian Society*, 108, p. 355 - 364.
- Burge, T. (1993). "Mind-Body Causation and Explanatory Practice". *Mental Causation*. Heil, John and Mele, Alfred (eds.). Oxford: Clarendon Press, p. 97-120.
- Butler, R. (2005). "Causes in Mind: Causal Exclusion's Collapse and the Causal Powers of Content". *Dissertation Abstracts International*, Vol. 66-05A.
- Campbell, D. (1974). "Downward Causation in Hierarchically Organised Biological Systems". *Studies in the philosophy of biology*. Francisco Jose Ayala and Theodosius Dobzhansky (Eds.). London: Macmillan, p. 179-186.
- Campbell, N. (2000). "Supervenience and Psycho-Physical Dependence". *Dialogue*, 39, p. 303-315.
- Campbell, N. (2008). *Mental Causation*. Peter Lang: New York.
- Campbell, N. (2010). "Explanatory Exclusion and the Intensionality of Explanation", *Theoria*, 76, p. 207-220.
- Campbell, R. J and Bickhard, M. (Forthcoming). "Physicalism, Emergence and

Downward Causation.”

Campbell, S. (2004). *Flaws and Fallacies in Statistical Thinking*, Courier Dover Publications.

Chalmers, D. (1996). *The Conscious Mind*. New York, Oxford University Press.

Chisholm, R. (1982). *The Foundations of Knowing*. Minneapolis: University of Minnesota Press.

Chisholm, R. (1989). *Theory of Knowledge*. Englewood Cliffs: Prentice-Hall.

Churchland, Patricia. (1986). *Neurophilosophy*, Cambridge: MIT Press.

Churchland, Paul. (1981). “Eliminative Materialism and the Propositional Attitudes”. *Journal of Philosophy* 78, p. 67-90.

Churchland, Paul (1988). *Matter and Consciousness*. Cambridge: MIT Press.

Churchland, Paul (1989). *A Neurocomputational Perspective*. Cambridge: Bradford Books.

Churchland, Paul (1998). *On the Contrary*, Cambridge: MIT Press.

Clapp, L. (2001). “Disjunctive Properties: Multiple Realizations”. *Journal of Philosophy*, 98, p. 111-136.

Clayton, P. (2006). “Conceptual Foundations of Emergence Theory”. *The Re-Emergence of Emergence*. Clayton, P. and Davies, P. (Eds). Oxford: Oxford University Press, p. 1-34.

Cornford, F. (1935). *Plato's Theory of Knowledge*. London: Routledge.

Crane, T. and Mellor, D (1990), “There is no Question of Physicalism”, *Mind* 99, p. 185-206.

Crane, T. (1995). “The Mental Causation Debate”. *Proceedings of the Aristotelian Society Supplementary*, 69, p. 211-236.

Crane, T. (2001). “The significance of emergence”. *Physicalism and its discontents*. B. Loewer and C. Gillett (Eds.). Cambridge: Cambridge University Press.

Crane, T (2008). “Causation and Determinable Properties”. *Being Reduced*. Jakob Hohwy and Jesper Kallestrup (eds.). Oxford: Oxford University Press.

- Craver, C. and Bechtel, W. (2007). "Top-Down Causation Without Top-Down Causes". *Biology and Philosophy*, 22, p. 547-563.
- Crisp, T. and Warfield, T. (2001). "Kim's Master Argument". *Noûs*, 35, p. 304-316.
- Cunningham, B. (2001). "The Re-emergence of 'Emergence'". *Philosophy of Science*, 68, p. 62-75.
- Darrow, C. (1924) "The Trial of Compulsion". *Attorney for the Damned: Clarence Darrow in the Courtroom*, Weinberg, A. (ed.), Chicago: The University of Chicago Press.
- David, M. (1997). "Kim's Functionalism". *Nous*, 31, p. 133-148.
- Davidson, D. (1963). "Actions, Reasons and Causes". *Journal of Philosophy*, 60, p. 685-700.
- Davidson, D. (1970). "Mental Events". *Experience and Theory*, Foster, L. and Swanson, J. (eds).
- Davidson, D. (1990). "The Structure of Content and Truth". *The Journal of Philosophy*, 87, p. 279-328.
- Davidson, D. (1993). "Thinking Causes". *Mental Causation*. Heil, John and Mele, Alfred (Eds.) Oxford: Clarendon Press, p. 3-18.
- Davidson, D. (1993b). "Reply to Richard Schantz". *Reflecting Davidson*. Ed. Ralf Stoecker. New York: De Gruyter, p. 36-39.
- Davidson, D. (1995). "Laws and Cause". *Dialectica*, 49, p. 263-279.
- Davidson, D. (1996). "Davidson, Donald". *A Companion to the Philosophy of Mind*. Samuel Gutternplan (Ed.). Wiley-Blackwell, p. 231-236.
- Davidson, D. (2001). *Essays on Actions and Events*. Oxford: Clarendon Press.
- Davidson, D. (2004) *Problems of Rationality*. Oxford: Oxford University Press.
- Davies, P. (2006). "The Physics of Downward Causation". *The Re-emergence of Emergence*. P. Clayton & P. Davies (Eds.). Oxford: Oxford University Press, p. 151-165.
- Descartes, R. (1988). *The Philosophical Writings Of Descartes*, 3 Vols., translated by

- John Cottingham, Robert Stoothoff. Cambridge: Cambridge University Press.
- Dretske, F. (1988). *Explaining Behaviour*. Cambridge: MIT Press.
- Dretske, F. (1993). "Mental Events as Structuring Causes of Behaviour". *Mental Causation*. Ed. Heil, John and Mele, Alfred. Oxford: Clarendon Press, p. 121-136.
- Dretske, Fred. (1995). "Reply: Causal Relevance and Explanatory Exclusion". *Philosophy of Psychology*. Ed. MacDonald, C., and MacDonald, G. Oxford: Blackwood, p. 142-155.
- Ehring, D. (1996). "Mental Causation, Determinables, and Property Instances". *Nous*, 30, p. 461-480.
- Elder, C. (2001). "Mental Causation Versus Physical Causation." *Philosophy and Phenomenological Research*, 62, p. 110-127.
- El-Hani, C. (2002). "On the Reality of Emergents". *Principia*, 6, p. 51-87.
- El-Hani, C. and Emmeche, C. (2000). "On Some Theoretical Grounds for an Organism-Centred Biology". *Theory in Biosciences*, 119, p. 234-275.
- El-Hani, C. and Pereira, A. (1999). "Understanding Biological Causation" .. (ed.) Where Biology Meets Psychology. Hardcastle, B. G. (ed.) Cambridge: MIT Press, p. 333-356.
- El-Hani, C. and Pereira, A. (2000). "Higher-level Descriptions?" *Downward Causation*. Andersen, P. B.; Emmeche, C.; Finnemann, N. O.; Christiansen, P. V. (eds.). Aarhus: Aarhus University Press, p. 118-142.
- El-Hani, C. and Pihlström, S. (2002) "Emergence Theories and Pragmatic Realism". *Essays in Philosophy*, 3.
- El-Hani, C. and Quiroz, J. (2005). "Downward Determination." *Abstracta*, 1, p. 162-192.
- Emmeche, C., et al. (1997). "Explaining Emergence: Towards an Ontology of Levels". *Journal for General Philosophy of Science*, 28, p. 83-119.
- Esfeld, M. (2005) "Mental Causation and Mental Properties". *Dialectica*: 59, p. 5-18.
- Evnine, S. (1991). *Donald Davidson*. Stanford: Stanford University Press.

- Ezquierro, J. (2000). "Explanatory Exclusion, Over-Determination, and the Mind-Body Problem". *The Proceedings of the Twentieth World Congress of Philosophy, Volume 9: Philosophy of Mind*. Elevitch, Bernard (Ed.).
- Farnell, R. (1971). "Metaphysical Necessity is Not Logical Necessity". *Philosophical Studies*, 39, p. 141-153.
- Feigl, H. (1958). "The 'Mental' and the 'Physical'". *Concepts, Theories and the Mind-Body Problem*. Feigl, H., Scriven, M. and Maxwell, G. (eds.), Minneapolis: Minnesota Studies in the Philosophy of Science, Vol. 2.
- Feldman, F. (1971). "Kripke on the Identity Theory". *Journal of Philosophy*, 71, p. 665-676.
- Feyerabend, P. (1981). *Philosophical Papers Vol. 1*, Cambridge: Cambridge University Press.
- Fine, A. (1984). "The Natural Ontological Attitude". *Scientific Realism*. Leplin, J. (Ed.). University of California Press, p. 83-107.
- Flanagan, O. (1992). *Consciousness Reconsidered*. Cambridge: MIT Press.
- Fodor, J. (1974). "Special Sciences, or the Disunity of Science as a Working Hypothesis". *Synthese*, 28, p. 77-115.
- Fodor, J. (1987). "Making mind matter more". *The Journal of Philosophy*, 84, p. 642-653.
- Fodor, J. (1990). *A Theory of Content and Other Essays*, Cambridge: MIT Press.
- Fodor, J. (1997). "Special Sciences: Still Autonomous After All These Years". *Philosophical Perspectives*, 11, p. 149-63.
- Foster, J. (1989). "A Defence of Dualism". *The Case for Dualism*. J. Smythies and J. Beloff, (eds). University of Virginia Press.
- Francescotti, R. (2007). "Emergence". *Erkenntnis*, 67, p. 47–63.
- Frankfurt, H. (1969). "Alternate Possibilities and Moral Responsibility," *Journal of Philosophy* 66, p. 829-839.
- Funkhouser, E. (2002). "Three Varieties of Overdetermination", *Pacific Philosophical Quarterly*

- Quarterly*, 83, p. 335-351.
- Funkhouser, E. (2006). "The Determinable-Determinate Relation". *Noûs* 40, p. 548–569.
- Fuhrmann, A. (2002). "Causal Exclusion without Explanatory Exclusion". *Manuscrito Vol. 25*, p. 177-198.
- Fuhrmann, A. and Mendonca, W. (2002). "Explanatory Exclusion and Causal Relevance". *Facta Philosophica*, 4, p. 287-300.
- Gettier, E. (1963). "Is Justified True Belief Knowledge?" *Analysis*, 23, p. 121-123.
- Gibb, S. (2004). "The Problem of Mental Causation and the Nature of Properties". *Australasian Journal of Philosophy*, 82, p. 464-75.
- Gibb, S. (2006). "Why Davidson Is Not a Property Epiphenomenalist". *International Journal of Philosophical Studies*, 14, p. 407-22.
- Gibb, S. (2009). "Explanatory Exclusion and Causal Exclusion". *Erkenntnis*, 71, p. 205-221.
- Gibbons, J. (2006). "Mental Causation Without Downward Causation". *Philosophical Review*, 115, p. 79-103.
- Gillett, C. et al. (2001). "Does the Argument from Realization Generalize?" *The Southern Journal of Philosophy*, 39, p. 79-98.
- Gillett, C. (2001). "Why Physics Alone Cannot Define the Physical". *Canadian Journal of Philosophy*, 31, p. 333-360.
- Gillett, C. (2006). "Samuel Alexander's Emergentism". *Synthese*, 153, p. 261-296.
- Gillett, C. and Rives, B. (2005). "The Nonexistence of Determinables". *Noûs*, 39, p. 483–504.
- Ginet, C. (1990). *On Action*. Cambridge: Cambridge University Press.
- Goldman, A. (1979). "What is Justified Belief?" *Justification and Knowledge: New Studies in Epistemology*, Dordrecht: Reidel, p. 1-23.
- Gozzano, S. (2003). "Second Order Properties: Why Kim's Reduction Does Not Work". *Logic and Philosophy of Science*, 1, p. 1-18.
- Greenwood, J. (1991). "Reasons to Believe". *The Future of Folk Psychology*: John D.

- Greenwood (ed.). Cambridge: Cambridge University Press.
- Guglielmo, S. and Monroe, A. and Malle, B. (2009). "At the Heart of Morality Lies Folk Psychology". *Inquiry*: 52, p. 449-466.
- Haug, M (*forthcoming*). Realization, Determination, and Mechanisms. *Philosophical Studies*.
- Harbecke, J. (2008). *Mental Causation: Investigating the Mind's Powers in a Natural World*, Frankfurt: Ontos Verlag.
- Hardcastle, V. (1996). *How to Build a Theory in Cognitive Science*. Suny Press.
- Hardcastle, V. (1998). "On the Matter of Minds and Mental Causation". *Philosophy and Phenomenological Research*, 58, p. 1-25.
- Harte, V. (2002). *Plato on Parts and Wholes*. Oxford: Oxford University Press.
- Hedman, C. (1972). "On When There Must be a Time-Difference Between Cause and Effect". *Philosophy of Science*, 39, p. 507-511.
- Hempel, C. (1965). *Aspects of Scientific Explanation and other Essays in the Philosophy of Science*. New York: Free Press.
- Hooker, C. (1979). "Critical Notice". *Dialogue* 18, p. 81-99.
- Hooker, C. (1981). "Towards a General Theory of Reduction", *Dialogue* 20, p. 38-59, 201-236, 496-529.
- Honderich, T. (1983). "Anomalous Monism: Reply to Smith". *Analysis*, 43, p. 147-149.
- Honderich, T. (1984). "Smith and the Champion of Mauve". *Analysis*, 44, p. 86-89.
- Horgan, T. (1991). "Actions, Reasons, and the Explanatory Role of Content". *Dretske and His Critics*. McLaughlin, B. (ed.). Oxford: Basil Blackwell.
- Horgan, T. (1997). "Kim on Mental Causation and Causal Exclusion". *Nous Supplement: Philosophical Perspectives*, 11, p. 165-184.
- Horgan, T. (2001). "Causal Compatibilism and the Exclusion Problem". *Theoria*, 16, p. 95-116.
- Horgan, T. (2007). "Mental Causation and the Agent-Exclusion Problem". *Erkenntnis* 67, p. 183-200.

- Hornsby, J. (1997). *Simple Mindedness*. Harvard University Press: Cambridge.
- Hume, D. (1888). *A Treatise of Human Nature*, L. Selby Bigge (ed.), Oxford: Oxford University Press.
- Humphreys, P. (1997). "How Properties Emerge". *Philosophy of Science*, 64, p. 1–17.
- Humphreys, P. (1997b). "Emergence, not supervenience". *Philosophy of Science*, 64 p. 337–345.
- Humphreys, P. (2008). "Synchronic and Diachronic Emergence". *Minds and Machines*, 18, p. 431-442.
- Huttemann, A. (2004). *What's Wrong with Microphysicalism*? Routledge.
- Jackson, F. and Pettit, P. (1990). "Program Explanation: A General Perspective". *Analysis*, 50, p. 107-117.
- Jackson, F. (1995). "Essentialism, Mental Properties and Causation". *Proceedings of the Aristotelian Society*, 95, p. 253-68.
- Jackson, F. (1996). "Mental Causation". *Mind*, 105, p. 377-413.
- Jackson, F. (2002). "From Reduction to Type-Type Identity". *Philosophy and Phenomenological Research*, 65, p. 644-647.
- Jacob, P. (2002). "Some Problems for Reductive Physicalism". *Philosophy and Phenomenological Research*, 65, p. 648-654.
- Jacquette, D. (1987). "Kripke and the Mind-Body Problem". *Dialectica*, 41, p. 293-301.
- Jacquette, D. (1994). *Philosophy of Mind*. Englewoods Cliffs: Prentice Hall.
- Johnston, M. (1992). "Constitution is not Identity". *Mind*, 101, p. 89–105.
- Jones, D. (1972). "Emergent Properties, Persons, and the Mind-Body Problem". *The Southern Journal of Philosophy*, 10, p. 423–433.
- Joos, E. (2006). "The Emergence of Classicality from Quantum Theory". *The Re-emergence of Emergence*. P. Clayton and P. Davies (Eds.). Oxford: Oxford Univ. Press, p. 151-165.
- Kallestrup, J. (2006). "The Causal Exclusion Argument". *Philosophical Studies*, 131, p. 459-485.

- Kane, R. (1999). "Responsibility, Luck, and Chance". *Journal of Philosophy*, 96, p. 217-240.
- Kane, R. (1996). *The Significance of Free Will*. New York: Oxford University Press.
- Kant, I. (1781). *Critique of pure reason*. Tr. by John Miller Dow. Courier Dover Publications.
- Katz, B. "Kim on Events". *The Philosophical Review*, 87, p. 427-441.
- Khalidi, M. (2005). "Against Functional Reductionism in Cognitive Science". *International Studies in the Philosophy of Science*, 19, p. 319-333.
- Kim, J. (1966). "On the Psycho-Physical Identity Theory". *American Philosophical Quarterly*, 3, p. 227-235.
- Kim, J. and Brandt, R. (1967). "The Logic of the Identity Theory". *Journal of Philosophy*, 66, p. 515-537.
- Kim, J. (1969). "Events and their Descriptions". *Essays in Honor of Carl G Hempel*. Rescher, N. (Ed.) Dordrecht: Reidel Publishing, p. 198-215.
- Kim, J. (1971). "Causes and Events". *Journal of Philosophy*, 68, p. 426-441.
- Kim, J. (1972). "Phenomenal Properties, Psychophysical Laws, and the Identity Theory". *The Monist*, 56, p. 177-192.
- Kim, J. (1973). "Symposium: Causation, Causes and Counterfactuals". *The Journal of Philosophy*, 70, p. 570-572.
- Kim, J. (1976). "Events as Property Exemplifications". *Action Theory*, Ed. Brand, M and Walton, D. Boston: Reidel Publishing Complany, p. 159-178.
- Kim, J. (1981). "Causes as Explanations: A Critique". *Theory and Decision*. 13, p. 293-309.
- Kim, J. (1984). "Epiphenomenal and Supervenient Causation". *Midwest Studies in Philosophy*, 9, p. 257-70.
- Kim, J. (1988). "Explanatory Realism, Causal Realism, and Explanatory Exclusion". *Midwest Studies in Philosophy*, 12, p. 225-239.
- Kim, J. (1988b). "What Is Naturalized Epistemology?" *Philosophical Perspectives*, 2, p.

381-405.

- Kim, J. (1989). "Mechanism, Purpose, and Explanatory Exclusion". *Nous-Supplement: Philosophical Perspectives*, 3, p. 77-108.
- Kim, J. (1989b). "The Myth of Nonreductive Materialism". *Proceedings of the American Philosophical Association*, 63, p. 1-27.
- Kim, J. (1991). "Events: Their Metaphysics and Semantics". *Philosophy and Phenomenological Research*, 51, p. 641-646.
- Kim, J. (1992). "Multiple Realization and the Metaphysics of Reduction". *Philosophy and Phenomenological Research*, 52, p. 1-26.
- Kim, J. (1992b). "Downward Causation in Emergentism and Nonreductive Physicalism". *Emergence or Reduction? Essays on the Prospects of Nonreductive Physicalism*. Flohr, Hans, Beckermann, Ansgar, and Kim, J. (Eds.). Berlin: Walter de Gruyter.
- Kim, J. (1993). *Supervenience and Mind*. Cambridge: Cambridge University Press.
- Kim, J. (1993b). "Can Supervenience and 'Non-Strict Laws' Save Anomalous Monism". *Mental Causation*. Heil, John and Mele, Alfred (Eds.). Oxford: Clarendon Press, p. 18-26.
- Kim, J. (1994). "Explanatory Knowledge and Metaphysical Dependence". *Philosophical Issues*, 5, p. 51-69.
- Kim, J. (1995) "Explanatory Exclusion and the Problem of Mental Causation". *Philosophy of Psychology*. MacDonald, G. and MacDonald, C. (Eds) Cambridge: Basil Blackwell, p. 120-139.
- Kim, J. (1995b). "Mental Causation in Searle's Biological Naturalism". *Philosophy and Phenomenological Research*, 55, p. 189-194.
- Kim, J. (1995c). "Mental Causation: What? Me Worry?" *Philosophical Issues: Content*, 6, p. 123-51.
- Kim, J. (1995d). "Reduction and Properties: Response to Merricks". *Philosophical Books*, 36, 161-164.
- Kim, J. (1997). "Does the Problem of Mental Causation Generalize?" *Proceedings of the*

- Aristotelian Society*, 87, p. 281-297.
- Kim, J. (1998). *Mind in a Physical World*. Cambridge: MIT Press.
- Kim, J. (1999). "Making Sense of Emergence". *Philosophical Studies* 95, p. 3-36.
- Kim, J. (1999b). "Supervenient Properties and Micro-Based Properties: A Reply to Noordhof". *Proceedings of the Aristotelian Society*, 99, p. 115-118.
- Kim, J. (2002). "The Layered Model: Metaphysical Considerations". *Philosophical Explorations*, 5, p. 2-20.
- Kim, J. (2002b). "Responses to Mind in a Physical World". *Philosophy and Phenomenological Research*, 65, p. 671-680.
- Kim, J. (2003). "Blocking Causal Drainage and Other Maintenance Chores with Mental Causation". *Philosophy and Phenomenological Research*, 67, p. 151-167.
- Kim, J. (2005). *Physicalism, or Something Near Enough*. Princeton: Princeton University Press:
- Kim, J. (2006). "Emergence: Core Ideas and Issues". *Synthese*: 151, p. 547-559.
- Kim, J. (2006). *Philosophy of Mind*. Cambridge: Westview Press.
- Kim, J. (2007). "Causation and Mental Causation". *Contemporary Debates in Philosophy Of Mind*. McLaughlin, B. and Cohen, J. (eds). Victoria: Blackwell.
- Kim, J. (2008). "Reduction and reductive explanation". *Being Reduced*. Jakob Hohwy, Jesper Kallestrup (Eds.). Oxford: Oxford University Press.
- Kim J. (2009). "Mental causation". *Oxford Handbook of Philosophy of Mind*. McLaughlin, B. and Beckermann, A. and Walter, S. (Eds.). Oxford: Oxford University Press, p. 29-52.
- Kistler, M. (2009). "Mechanisms and Downward Causation". *Philosophical Psychology*, 22, p. 595-609
- Kitcher, P. (1978). "Theories, Theorists, and Theoretical Change". *The Philosophical Review*, 87, p. 519-547.
- Kitcher, P. (1992). "The Naturalists Return". *Philosophical Review*, 101, p. 53-114.
- Kitcher, Patricia (1984). "In defense of Intentional Psychology". *The Journal of*

- Philosophy*, 81, p. 89-106.
- Koksvik, O. (2007). "Conservation of Energy is Relevant to Physicalism". *Dialectica*, 61, p. 573-582.
- Kornblith, H. (1993). *Naturalizing Epistemology*. Cambridge: MIT Press.
- Kornblith, H. (1993b). *Inductive Inference and its Natural Ground*. Cambridge: MIT press.
- Koslicki, K. (2008). *The Structure of Objects*. Oxford: Oxford University Press.
- Kripke, S. (1980). *Naming and Necessity*. Cambridge: Harvard University Press.
- Kroedel, T. (2008). "Mental Causation as Multiple Causation". *Philosophical Studies*, 139, p. 125-143.
- Lahav, R. (1992). "The Amazing Predictive Power of Folk Psychology". *Australasian Journal of Philosophy*, 70, p. 99-105.
- Leiter, B. and Miller, A. (1994). "Mind Doesn't Matter Yet". *Australasian Journal of Philosophy* 72, p. 220-228.
- Leiter, B. and Miller, A. (1998). "Closet Dualism and Mental Causation". *Canadian Journal of Philosophy*, 28, p. 161-181.
- LePore, E, and Loewer, B. (1987). "Mind Matters." *Journal of Philosophy*, 84, p. 630-642.
- Lewis, D. (1973). "Causality", *The Journal of Philosophy*, 70, p. 556-567.
- Lewis, D. (1979). "Counterfactual Dependence and Time's Arrow". *Noûs*, 13, p. 445-476.
- Lewis, D. (1981). "Are We Free to Break the Laws?" *Theoria*, 47, p. 113-21.
- Lewis, D. (1983). "New Work for a Theory of Universals". *Australasian Journal of Philosophy*, 61, p. 343-377
- Lewis, D. (1986). *On the Plurality of Worlds*, Oxford: Blackwell.
- Lewis, D. (1986b). *Philosophical Papers: Volume II*. Oxford: Oxford University Press.
- Lewis, D. (1991). *Parts of Classes*. Cambridge: Basil Blackwell.
- Lewis, D. (2000). "Causation as Influence". *The Journal of Philosophy*, 97, p. 182-197.
- Levin, M. (1977). "The Extensionality of Causation and Causal-Explanatory Contexts".

- Philosophy of Science*, 43, p. 266-277.
- Levine, J. (2001). *Purple Haze: The Puzzle of Consciousness*. Oxford: Oxford University Press.
- Liu, J. (2001). "A Nonreductionist's Solution to Kim's Explanatory Exclusion Problem". *Manuscrito*, 24, p. 7-47.
- Loar, B. (1981). *Mind and Meaning*. Cambridge: Cambridge University Press.
- Loewer, B. (2001). "Mind in a Physical World". *The Journal of Philosophy*, 98, p. 315-324.
- Loewer, B. (2002). "Comments on Jaegwon Kim's Mind and the Physical World". *Philosophy and Phenomenological Research*, 65, p. 655-662.
- Loewer, B. (2007). "Mental Causation, or Something Near Enough". *Contemporary Debates in Philosophy of Mind*. McLaughlin, B. and Cohen, J. (Eds). Malden: Blackwell Publishing, p. 243-264.
- Lombard, L. (1985). *Events: A Metaphysical Study*. Routledge & Kegan Paul.
- Longworth, F. (2006). "Causation, Counterfactual Dependence and Pluralism Thesis". University of Pittsburg School of Arts and Sciences.
- Lowe, E. (1993). "The Causal Autonomy of the Mental". *Mind*, 102, p. 629-644.
- Lowe, E. (2001). *The Possibility of Metaphysics*. Oxford: Oxford University Press.
- Lowe, E. (2003). "Physical Causal Closure and the Invisibility of Mental Causation". *Physicalism and Mental Causation*. S. Walter & H. Heckmann, (eds.) Imprint Academic
- Lowe, E. (2006). "Non-Cartesian Substance Dualism and the Problem of Mental Causation", *Erkenntnis*, 65, p. 5-23.
- Lyons, J. (2006). "In Defense of Epiphenomenalism". *Philosophical Psychology*, 19, p. 767-794.
- MacDonald, C. (1995). "Causal Relevance and Explanatory Exclusion". *Philosophy of Psychology*. MacDonald, C. and MacDonald, G. (Eds.) Oxford: Basil Blackwell.
- MacDonald, C. (2005). *Varieties of Things*. Cambridge: Blackwell.

- MacDonald, C., and MacDonald, G. (1986). "Mental Causes and Explanation of Action". *Mind, Causation and Action*. L. Stevenson, R. Squires, and J. Haldane (Eds.). Oxford: Basil Blackwell.
- MacDonald, C. and MacDonald, G. (2006). "The metaphysics of mental causation". *The Journal of Philosophy*, 103, p. 539-576.
- Malcolm, N. (1968). "The Conceivability of Mechanism". *Philosophical Review*. 77, p. 45-72.
- Marcus, E. (2001). "Mental Causation: Unnaturalized but not Unnatural". *Philosophy and Phenomenological Research*, 63, p. 57-83.
- Marcus, E. (2005). "Mental Causation in a Physical World". *Philosophical Studies*, 122, p. 27-50.
- Markic,O. (2002). *Nonreductive Materialism and the Problem of Causal Exclusion*. Atlanta: Rodopi.
- Markic, O. (2004). "Nonreductive Materialism and the Problem of Causal Exclusion". *Acta Analytica*, 19, p. 65-81.
- Marino, P. (2008). "Toward a Modest Correspondence Theory of Truth". *Dialogue*, 47, p. 81-102.
- Marras, A. (1997). "The Causal Relevance of Mental Properties". *Philosophia* 25, p. 389-400.
- Marras, A. (1998). "Kim's Principle of Explanatory Exclusion". *Australasian Journal of Philosophy*, 76, p. 439-451.
- Marras, A. (1998b). "Critical Notice of Kim (1998)". *Canadian Journal of Philosophy*, 76, p. 439-451.
- Marras, A. (2003). "Methodological and Ontological Aspects to the Mental Causation Problem". *Physicalism and Mental Causation*. Sven Walter, Heinz-Dieter Heckmann (Eds.) Imprint Academic.
- Marras, A. (2007). "Kim's Supervenience Argument and Nonreductive Physicalism". *Erkenntnis*, 66, p. 305-327.

- Marras, A. and Yli-Vakuri, J. (2008). "The Supervenience Argument". *Tropes, Universals and the Philosophy of Mind*, Gozzano, S. and Orilia, F. (Eds.). Frankfurt: Ontos Verlag, p. 101-134.
- Martin, C. and Heil, J. (1999). "The Ontological Turn". *Midwest Studies in Philosophy*, 23, p. 34–60.
- McDaniel, K (2008). "Against Composition as Identity". *Analysis*, 68, p. 128-133.
- McGinn, C. (1989). *Mental Content*. Oxford: Blackwell.
- McGinn, C. (1989b). "Can We Solve the Mind-Body Problem?" *Mind*, 98, p. 349-366.
- McGinn, C. (1991). *The Problem of Consciousness*. Cambridge: Basil Blackwell.
- McGinn, C. (1999). *The Mysterious Flame*. New York: Basic Books.
- McGivern, P. (2008). "Reductive Levels and Multi Scale Structure". *Synthese*, 165, p. 53-75.
- McIntyre, L (2002). "Supervenience and Explanatory Exclusion". *Critica*, 34, p. 87-101.
- McLaughlin, B. (1992). "The Rise and Fall of British Emergentism". *Emergence or Reduction?* A. Beckermann, H. Flohr, & J. Kim (Eds.). Berlin: De Gruyter.
- McLaughlin, B. (1993). "On Davidson's Response to the Charge of Epiphenomenalism." *Mental Causation*. Heil, J. and Mele, A. (eds.). Oxford: Oxford University Press.
- McLaughlin, B. (1997). "Emergence and Supervenience". *Intellectica*, 25, p. 25-43.
- Melden, A. (1961). *Free Action*. London: Routledge and Kegan Paul.
- Mellor, D. (1999). "Properties and Predicates". *The Nature of Properties*. Tooley, M. (Ed.) Cambridge: Taylor and Francis, p. 25-39.
- Menzies, P. (1996). "Probabilistic Causation and the Pre-Emption Problem". *Mind*, 105, p. 85-117.
- Merricks, T. (1995). "Review of Supervenience and Mind". *Philosophical Books*, 36, p. 156-164.
- Merricks, T. (1999). "Composition as Identity, Mereological Essentialism, and Counterpart Theory". *Australasian Journal of Philosophy*, 77, p. 192–195.
- Merricks, T. (2005). "Composition and Vagueness". *Mind*, 114, p. 615-637.

- Meixner, U. (2008). "New Perspectives for a Dualistic Conception of Mental Causation". *Journal of Consciousness Studies*, 15, p. 17-38.
- Mele, A. (1992). *Springs of Action*. New York: Oxford University Press.
- Mellor, D. (1995). *The Facts of Causation*, London: Routledge.
- Melnyk, A. (1996). "Testament of a Recovering Eliminativist". *Philosophy of Science*, 63, p. S185-S193.
- Melnyk, A. (2003). *Physicalist Manifesto*. Cambridge: Cambridge University Press.
- Melnyk, Andrew (2005) "Review of Jaegwon Kim's Physicalism, Or Something Near Enough". <http://ndpr.nd.edu/review.cfm?id=3321>
- Menzies, P. and List, C. (2010). "The Causal Autonomy of the Special Sciences". *Emergence in Mind*. Cynthia McDonald & Graham McDonald (eds.). Oxford: Oxford University Press.
- Miller, A. and Leiter, B. (1994). "Mind Doesn't Matter Yet". *Australasian Journal of Philosophy*, 72, p. 220-228.
- Mill, J. (1843). *A System of Logic*. London: Longman.
- Mills, E. (1996). "Interactionism and overdetermination". *American Philosophical Quarterly*, 33, p. 105–117.
- Montero, B. (2003). "Varieties of Causal Closure". *Physicalism and Mental Causation*. S. Walter and H. Hackmann (Eds.). Exeter: Imprint Academic, p. 173-187.
- Montero, B. (2006). "What does the Conservation of Energy Have to Do with Physicalism?" *Dialectica*, 60, p. 383-396.
- Moreno, A. and Umerez, J. (1999). "Downward Causation at the Core of Living Organization". *Downward Causation*. Peter Bogh Anderson et al. (eds.). Oakville: Aarhus University Press, p. 99–117.
- Morgan, C. (1923). *Emergent evolution*. London: Williams and Norgate.
- Murphy, N. (2006). "Emergence and Mental Causation". *The Re-Emergence of Emergence*. Clayton, P. and Davies P. (Eds.). Oxford: Oxford University Press, p. 227-243.

- Nagel, E. (1961). *The Structure of Science*. New York: Harcourt, Brace, and World.
- Nagel, T. (1974). "What it is Like to Be a Bat". *Philosophical Review*, 83, p. 435-450.
- Nahmias, E. et al. (2005). "Surveying Freedom". *Philosophical Psychology*, 18, p. 561-584.
- Newen, A. and Cuplinskas, R. (2002). "Mental Causation: A Real Phenomenon in a Physicalistic World without Epiphenomenalism or Overdetermination". *Grazer Philosophische Studien*, 65, p. 139-167.
- Ney, A. (2007). "Can an Appeal to Constitution Solve the Exclusion Problem?" *Pacific Philosophical Quarterly*, 88, p. 486-506.
- Noordhof, P. (1998). "Do Tropes Resolve the Problem of Mental Causation?" *Philosophical Quarterly*, 48, p. 221-226.
- Noordhof, P. (1999). "Micro-Based Properties and the Supervenience Argument". *Proceedings of the Aristotelian Society*, 99, p. 109-114.
- Nosek, J. (1995). "An Eliminative Attempt at the Resolution of the Psychophysical Problem". *Filosoficky Casopis* 43, p. 225-248.
- O'Connor, T. (1994). "Emergent properties". *American Philosophical Quarterly*, 31, p. 91-104.
- O'Connor, T. (2000). "Causality, mind, and free will". *Philosophical Perspectives*, 14, p. 105-117.
- O'Connor, T and Jacobs, J. (2003). "Emergent Individuals". *The Philosophical Quarterly*, 53, p. 540-555.
- O'Connor, T., and Wong, H. (2005). "The Metaphysics of Emergence", *Noûs*, 39, p. 658-678.
- Odling, W. (1871). "The Revised Theory of Phlogiston". *Chemical News*, 1871 p. 243-245, 256-258.
- Papineau, D. (2001). "The Rise of Physicalism". *Physicalism and its discontents*. Gillett, C. and Loewer, B. (Eds.) Cambridge: Cambridge University Press, p. 3-36.
- Papineau, D. (2008). "Must a Physicalist be a Microphysicalist?" *Being Reduced*.

- Hohwy, J. and Kallestrup, J. (Eds.). Oxford: Oxford University Press, p. 125-148.
- Paul, L. (2000). "Aspect causation". *Journal of Philosophy*, 97, p. 235-56.
- Paul, L. (2007). "Constitutive Overdetermination," *Topics in Contemporary Philosophy, Volume 4*. Keim Campbell, M. O'Rourke, and D. Shier (eds). Cambridge: MIT Press. Page numbers correspond to online version.
- Pereboom, D. (2002). "Robust Nonreductive Materialism". *Journal of Philosophy*, 99, p. 499-531.
- Pereboom, D. and Kornblith, H. (1991). "The Metaphysics of Irreducibility". *Philosophical Studies*, 63, p. 125-145.
- Pianesi, F and Varzi, A. (2000). *Speaking of Events*. New York: Oxford University Press, p. 3-47.
- Pietroski, P. (1994). "Mental Causation for Dualists". *Mind and Language*, 9, p. 336-366.
- Pineda, D. (2001). "Functionalism and nonreductive physicalism". *Theoria*, 16, p. 43-63.
- Pineda, D. (2002). "The Causal Exclusion Puzzle". *European Journal of Philosophy*, 10, p. 26-42.
- Pineda, D. (2005). "Causal Exclusion and Causal Homogeneity". *Dialectica*, 59, p. 63-66.
- Place, U. (1956). "Is Consciousness a Brain Process?" *British Journal of Psychology*, 47, p. 44-50.
- Plato. *Complete Works of Plato*. John Cooper (Ed.). Indianapolis: Hackett Publishing.
- Poirier, P. (2006). "Finding a Place for Elimination in Inter-level Reductionist Activities". *Synthese* 151, p. 477-483.
- Poland, J. (1994). *Physicalism*. Oxford: Clarendon Press.
- Popper, K. (1957). "The Aim of Science", *Ratio* 1, p. 24-35.
- Pritchard, D. (2005). *Epistemic Luck*. Oxford: Oxford University Press.
- Psillos, S. (2000). "The Present State of the Scientific Realism Debate". *British Journal for the Philosophy of Science*, 51, p. 705-728.
- Putnam, H. (1967). "Psychological Predicates". *Art, Mind and Religion*, Capitan, W. and Merrill, D. (Eds.). Pittsburgh: University of Pittsburgh Press, p. 37-48.

- Putnam, H. (1979). "Mind, language, and reality." *Philosophical papers (of Hilary Putman Volume 2)*. Cambridge: Cambridge University Press.
- Putnam, H. (1984) "Is the Causal Structure of the Physical Itself Something Physical". *Midwest Studies in Philosophy*, 9, p. 3-16
- Raymont, P. (2003). "Persons and Causes". *Review of Metaphysics*, 57, p.170-172.
- Raymont, P. (2003). "Kim on Overdetermination, Exclusion and Nonreductive Physicalism". *Physicalism and Mental Causation*, Walter, S. and Heckmann, H. (Eds.). Exeter: Imprint, p. 225-242.
- Rey, G. (1997). *Contemporary Philosophy of Mind*: Wiley-Blackwell.
- Riggs, W. (2007). "Why Epistemologists are so Down on their Luck". *Synthese*, 158, p. 329-344.
- Robb, D. (1997). "The Properties of Mental Causation". *The Philosophical Quarterly*, 47, p. 178-194.
- Robb, D, (2001). "Reply to Noordhof on Mental Causation". *Philosophical Quarterly*, 51, p. 90-94.
- Robb, D. (2008). "Review of Harbecke". *Notre Dame Philosophical Reviews*, 9.
- Robb, D. and Heil, J. (2008). "Mental Causation". *The Stanford Encyclopedia of Philosophy*. <http://plato.stanford.edu/entries/mental-causation/>
- Rorty, R. (1970). "Incorrígibility as the Mark of the Mental". *Journal of Philosophy*, 67, p. 399-424.
- Rothschild, L. (2006). "The Role of *Emergence* in Biology". *The Re-emergence of Emergence*. P. Clayton and P. Davies (Eds.). Oxford: Oxford University Press, p. 151-165
- Rueger, A. (2000). "Robust supervenience and emergence". *Philosophy of Science*, 67, p. 466-489.
- Rueger, A. (2000). "Physical Emergence, Diachronic And Synchronic". *Synthese*, 124, p. 297-322.
- Rueger, A. (2004). "Reduction, Autonomy, and Causal Exclusion among Physical

- Properties". *Synthese*, 139, p. 1-21.
- Rueger, A. (2006). "Functional Reduction and Emergence in the Physical Sciences". *Synthese*, 151, p. 335-346.
- Ryle, G. 1949: *The Concept of Mind*, London: Hutchinson.
- Sabates, M. (1996). "Kim on the Metaphysics of Explanation". *Manuscrito*, 19, p. 93-110.
- Sabates, M. (2002). "Review of Mind in a Physical World". *Philosophy and Phenomenological Research*, 65, p. 663-670.
- Sachse, C. (2007). *Reductionism in the Philosophy of Science*, Frankfurt: Ontos.
- Sawyer, R. (2004). "The Mechanisms of Emergence". *Philosophy of the Social Sciences*, 34, p. 260–282.
- Schiffer, S. (1987). *Remnants of Meaning*, Cambridge: MIT Press.
- Schroder, J. (2002). "The Supervenience Argument and the Generalization Problem". *Erkenntnis*, 56, p. 319-328.
- Schwartz, J. (1991). "Reduction, Elimination and the Mental". *Philosophy of Science*, 58, p. 203-220.
- Seager, W. (2006). "Emergence, Epiphenomenalism and Consciousness". *Journal of Consciousness Studies*, 13, p. 21-38.
- Searle, J. (1984). "Intentionality and its Place in Nature". *Synthese* 38, p. 87-100.
- Searle, J. (1992). *The Rediscovery of the Mind*. Cambridge: MIT Press.
- Searle, J. (2004). *Mind: A Brief Introduction*. Oxford: Oxford University Press.
- Segal, G. (2009). "The Causal Inefficacy of Content", *Mind and Language*, 24, p. 80-102.
- Shapiro, L. and Sober, E. (2007). "Epiphenomenalism the Do's and Donts". *Thinking about Causes*. Peter K. Machamer, Gereon Wolters (Eds.) Pittsburg: Univ. of Pittsburgh Press.
- Shoemaker, S. (1999). "Self, Body, and Coincidence". *Proceedings of the Aristotelian Society*, 73, p. 287-306.

- Shoemaker, S. (2001). "Realization and Mental Causation". *Physicalism and Its Discontents*. Gillett, C. and Loewer, B. (Eds.). Cambridge: Cambridge University Press, p. 74-98.
- Shoemaker, S. (2002). "Kim on emergence". *Philosophical Studies*, 108, p. 53–63.
- Schroder, J. (2002). "The Supervenience Argument and the Generalization Problem". *Erkenntnis*, 56, p. 319-328.
- Shaffer, J. (2003). "Overdetermining Causes". *Philosophical Studies*, 114, p. 23-45.
- Shrader, W. (*forthcoming*). "Shoemaker on Emergence". *Philosophical Studies*, online first.
- Sider, T. (2003). "What's So Bad About Overdetermination". *Philosophy and Phenomenological Research*, 67, p. 719 - 726.
- Sider, T. (2007). "Parthood". *Philosophical Review*, 116, p. 51-91.
- Silberstein, M. (1998). "Emergence and the mind-body problem". *Journal of Consciousness Studies*, 5, p. 464–482.
- Silberstein M., and McGeever, J. (1999). "The search for ontological emergence". *The Philosophical Quarterly*, 49, p. 182–200.
- Silberstein, M. (2001). "Converging on emergence". *Journal of Consciousness Studies*, 8, p. 61-98.
- Silberstein, M. (2006). "In Defence of Ontological Emergence and Mental Causation". *The Re-Emergence of Emergence*. Philip Clayton and Paul Davies (eds.). Oxford: Oxford University Press.
- Slors, M. and Walter, S. (2002). "Introduction". *Mental Causation, Multiple Realization, and Emergence*, Slors, M. and Walter, S. (eds). Rodopi.
- Smart, J. (1959). "Sensations and Brain Processes". *Philosophical Review*, 68, p. 141-156.
- Smith, P. (1983). "On Identifying the Mental With the Physical". *Canadian Journal of Philosophy*, 13, p. 227-238.
- Smith, P. (1984). "Anomalous Monism and Epiphenomenalism". *Analysis*, 44, p. 83-86.

- Sosa, E. (1984). "Mind-Body Interaction and Supervenient Causation". *Midwest Studies in Philosophy*, 9, p. 271-281.
- Soto, A. and Sonnenschein, C. (2006). "Emergentism by Default". *Synthese*, 151, p. 361-376.
- Sperry, R. (1969). "A Modified Concept of Consciousness". *Psychological Review*, 76, p. 532–536.
- Sperry, R. (1975). "Mental Phenomena as Causal Determinants in Brain Function". *Process Studies*, 5, p. 247-256.
- Sperry, R. (1986). "Macro- Versus Micro-determination". *Philosophy of Science*, 53, p. 265-275.
- Sperry, R. (1991). "In Defense of Mentalism and Emergent Interaction". *The Journal of Mind and Behavior*, 12, p. 221–245.
- Stephan, A. (1992). "Emergence – A Systematic View on its Historical Facets". *Emergence or Reduction?* A. Beckermann, H. Flohr, and J. Kim (Eds.). Berlin: De Gruyter.
- Stephan, A. (1999). *Emergenz: von der Unvorhersagbarkeit zur Selbstorganisation*. Dresden: Dresden University Press.
- Stern, R. (2004). "Does 'Ought' Imply 'Can'? And Did Kant Think It Does?" *Utilitas*, 16, p. 42-61.
- Stich, S. (1983). *From Folk Psychology to Cognitive Science*. Cambridge: MIT Press.
- Strawson, G. (1985). *Skepticism and Naturalism*. London: Columbia University Press.
- Strawson, G. (1986). *Freedom and Belief*. Oxford, Oxford University Press.
- Stueber, K. (2005). "Mental Causation and the Paradoxes of Explanation". *Philosophical Studies*, 122,p. 243-277.
- Sturgeon, S. (1998). "Physicalism and Overdetermination". *Mind*, 107, p. 411-432.
- Sturgeon, S. (1999). "Conceptual Gaps and Odd Possibilities". *Mind*, 108, p. 377-380.
- Sturgeon, S. (2000). *Matters of Mind*. Routledge, London.
- Sturgeon, S. (2001). "The Roots of Reductionism". *Physicalism and Its Discontents*.

- Gillett, Carl and Loewer, Barry (eds.). Cambridge: Cambridge University Press.
- Swoyer, C. (1996). "Theories of Properties: From Plenitude to Paucity". *Philosophical Perspectives*, 10, p. 243-264
- Symons, J. (2002). "Emergence and Reflexive Downward Causation". *Principia*, 6, p. 183-202.
- Tammelleo, S. (2008). "A Limited Defense of Epiphenomenalism". *South African Journal of Philosophy*, 27, p. 40-51.
- Thalberg, I. (1980). "Can We Get Rid of Events?" *Analysis*, 40, p. 25-31.
- Thomasson, A. (1998). "A Nonreductivist Solution to Mental Causation". *Philosophical Studies*, 89, p. 181-195.
- Toner, P. (2008). "Emergent Substance". *Philosophical Studies*, 141, p. 281-297.
- Unwin, N. (1996). "The Individuation of Events". *Mind*, 105, p. 315-330.
- Van Cleave, J. (1985). "Three Versions of the Bundle Theory". *Philosophical Studies*, 47, p. 95-107.
- Van Gulick, R. (1992). "Three Bad Arguments for Intentional Property Epiphenomenalism". *Erkenntnis*, 36, p. 311-332.
- Van Gulick, R. (1993). "Who's in Charge Here? And Who's Doing all the Work?" *Mental Causation*, Heil, John and Mele, Alfred (Eds.). New York: Clarendon Press.
- Van Gulick, R. (2001). "Reduction, Emergence and Other Recent Options on the Mind/Body Problem". *Journal of Consciousness Studies*, 8, p. 1-34.
- Van Inwagen, P. (1975). "The Incompatibility of Free Will and Determinism" *Philosophical Studies*, 27, p. 185-99.
- Van Inwagen, P. (1983). *An Essay on Free Will*. Oxford: Clarendon Press.
- Van Inwagen, P. (1994). "Composition as Identity". *Philosophical Perspectives*, 8, p. 207-220.
- Vicente, A. (2002). "The Dual Explanandum Strategy". *Critica*, p. 73-96.
- Vicente, A. (2002). "How Dispositions Can be Causally Relevant". *Erkenntnis*, 56, p.

- 329–344.
- Vicente, A. (2004). “The role of Dispositions in Explanation”. *Theoria*, 19, p. 301–310.
- Walter, S. (2005). “Program Explanations and Causal Relevance”. *Acta Analytica*, 20, p. 32–47.
- Walter, S. (2006). “Causal Exclusion as an Argument against Nonreductive Physicalism”. *Journal of Consciousness Studies*, 13, p. 67–83.
- Walter, S. (2008). “The Supervenience Argument, Overdetermination, and Causal Drainage”. *Philosophical Psychology*, 21, p. 673–696.
- Walter, S. (*forthcoming*). “Taking realization seriously”. *Philosophical studies*, page references to online first version.
- Wegner, D. (2002). *The Illusion of Conscious Will*. Cambridge: MIT Press.
- Welshon, R. (2002). “Emergence, Supervenience, and Realization”. *Philosophical Studies*, 108, p. 39–51.
- Whittle, A. (2007). “The Co-Instantiation Thesis”. *Australasian Journal of Philosophy*, 85, p. 61–79.
- Wierenga, E. and Feldman, R. (1978). “Identity Conditions and Events”. *Canadian Journal Of Philosophy*, 11, p. 77–90.
- Wiggins, D. (1967). *Identity and Spatio-Temporal Continuity*. Oxford: Blackwell.
- Wilkes, K. (1995). “Losing Consciousness”. *Conscious Experience*. Metzinger, T. (ed.). Ferdinand Schoningh.
- Wilson, M. (1982). “Predicate Meets Property”. *Philosophical Review*, 91, p. 549–89.
- Wilson, J. (1999). “How Superduper Does a Physicalist Supervenience Need to be?” *The Philosophical Quarterly*, 49, p. 33–52.
- Wilson, J. (2009). “Determination, Realization and Mental Causation”. *Philosophical Studies*, 145, p. 149–169.
- Witmer, G. (2003). “Functionalism and Causal Exclusion”. *Pacific Philosophical Quarterly*, 84, p. 198–215.
- Worley, S. (1993). “Mental Causation and Explanatory Exclusion”. *Erkenntnis*, 39, p.

333-358.

Yablo, S. (1992). "Mental Causation". *Philosophical Review*. 101, p. 245-280.

Yablo, S. (2000). "Textbook Kripkeanism and the Open Texture of Concepts". *Pacific Philosophical Quarterly*, 81, p. 98-122.

Yi, B. (1999). "Is Mereology Ontologically Innocent?" *Philosophical Studies*, 93, p. 141-160.