Emergence is the metaphysical buzzword of the last decade. This dissertation investigates ontological emergence (OE), a concept prevalent in contemporary philosophy of mind and science. An ontologically emergent property (hereafter, emergent property) is causally efficacious, irreducible to microphysical properties, and yet still dependent upon microphysical properties. Though this basic idea is clear, there is considerable disagreement in the literature as to the exact nature of emergent properties, and this motivates this dissertation.

Chapter One introduces some metaphysical assumptions and concepts regarding properties and causation. Particular attention is given to the frequently invoked yet under-analyzed notion of causal power contribution. An analysis distinct from Shoemaker’s and Armstrong’s is suggested.

Chapter Two situates OE among solutions to the Problem of the Special Sciences, the problem of identifying the referents of special science predicates. It is suggested that recent formulations of OE invoke too many necessary conditions upon an emergent property. It is contended that a more modest formulation of OE is sufficient to
distinguish OE from its main rivals—epiphenomenalism, micro-determinism, and standard nonreductive physicalism.

Chapter Three reviews two Kimian arguments against British Emergentism, one of which is shown to be unsound and the other not injurious to the more modest formulation of OE suggested in Chapter Two.

The fourth chapter houses a taxonomy of theories of OE. Such theories can be classified along two dimensions, the first being the type of dependence relation holding between emergent properties and microphysical properties and the second being the way in which an emergent property is causally efficacious. Mere supervenience, causal dependence, and two distinct types or realization are the examined dependence relations (with original analyses of mere supervenience and realization proffered). Novel causal contribution and synchronic non-causal determination are the competing understandings of causal efficacy. This suggests eight distinct, logically possible theories of OE. The prospects for each are briefly discussed.

The dissertation concludes with a review of theories proffered by Derk Pereboom and Carl Gillett. It is argued that both suffer from internal problems that preclude judging them to be true. Nonetheless, much is to be learned by noting where they go wrong.
CONTENTS

PREFACE ....................................................................................................................... iv

CHAPTER 1 ONTOLOGICAL EMERGENCE: THE BASIC CONCEPTS .......... 1
  1.1 Introduction ......................................................................................................... 1
  1.2 Origins of the Concept of Ontological Emergence ........................................... 3
  1.3 What Is Ontological Emergence? ...................................................................... 6
  1.4 Clarifying Some Important Assumptions and Concepts ................................. 15
    1.4.1 The Microphysical ...................................................................................... 15
    1.4.2 Properties, Property Instance, and Causal Relations ............................... 17
    1.4.3 Causal Powers ......................................................................................... 17
      1.4.3.1 Conditional vs. Manifest Powers ....................................................... 19
      1.4.3.2 The Ontological Status of Causal Powers ......................................... 22
      1.4.3.3 An Analysis of Property Causal Power Contribution ..................... 27
    1.4.4 Causal Efficacy and Causal Autonomy ................................................... 32
    1.4.5 Causal Power Types Versus. Causal Power Tokens ................................. 34

CHAPTER 2 ONTOLOGICAL EMERGENCE AND THE PROBLEM OF
  THE SPECIAL SCIENCES .......................................................................................... 36
  2.1 Introduction to the Problem ............................................................................. 37
  2.2 Categorizing Responses to the Problem of the Special Sciences ................. 42
    2.2.1 Eliminativism and Unqualified Micro-determinism .................................. 45
    2.2.2 Qualified Micro-Determinism .................................................................. 47
    2.2.3 Non-Reductive Physicalism and Epiphenomenalism .............................. 52
    2.2.4 Ontological Emergence vs. Serious Dualism .......................................... 57
  2.3 Minimal Ontological Emergence ..................................................................... 58

CHAPTER 3 CLASSIC ONTOLOGICAL EMERGENCE AND KIM’S
  CHALLENGE ............................................................................................................. 61
  3.1 Introduction ....................................................................................................... 61
  3.2 A Quick Tour of British Emergentism ............................................................. 62
  3.3 Jaegwon Kim’s Formulation of COE ............................................................... 69
    3.3.1 Non-functionalizability .......................................................................... 70
    3.3.2 Novel Causal Power Contribution ........................................................... 71
    3.3.3 Synchronic Physical Determination and Strong Supervenience ............ 75
  3.4 Some Reasons to Think K-COE Is Too Restrictive .......................................... 78
  3.5 Kim’s Challenge ............................................................................................... 80

CHAPTER 4 ALTERNATE CONCEPTIONS OF ONTOLOGICAL
  EMERGENCE ........................................................................................................... 90
  4.1 Introduction ....................................................................................................... 90
4.2 Options for the Microphysical Determination Condition ........................................... 95
  4.2.1 Mere Supervenience ................................................................................................. 95
  4.2.2 Realization ............................................................................................................. 101
    4.2.2.1 Distinguishing Two Realization Relations: R^{THEO} vs. R^{CAUS} ............... 103
    4.2.2.2 Formulating R^{THEO} .................................................................................... 107
    4.2.2.3 Formulating R^{CAUS} .................................................................................... 115
  4.2.3 Causation ............................................................................................................. 117
  4.3 An Alternate Interpretation of Causal Efficacy and Autonomy ................................. 118
  4.4 A Taxonomy of Possible Theories of Ontological Emergence ................................. 121
    4.4.1 Theories Incorporating SCPD (Boxes V, VI, VIII, and VIII) ............................... 121
    4.4.2 Box III: R^{CAUS} Plus Novel Causal Powers .................................................... 123
    4.4.3 Box II: R^{THEO} Plus Novel Causal Powers ...................................................... 123
    4.4.4 Box IV: Causation plus Novel Causal Powers ..................................................... 124
    4.4.5 Box I: Mere Supervenience plus Novel Causal Powers ...................................... 128
  4.5 Conclusion ............................................................................................................. 130

CHAPTER 5 PEREBOOM’S ROBUST NON-REDUCTIVE PHYSICALISM .......... 131
  5.1 Introduction ............................................................................................................. 131
  5.2 An Overview of Robust Non-Reductive Physicalism ............................................. 135
  5.3 Pereboom’s Constitution Relation .......................................................................... 142
  5.4 Token Causal Powers and Revised Novel Causal Power Contribution .................... 150
  5.5 Analyzing Pereboom’s Case for R-NRP ................................................................. 153
  5.6 Conclusion ............................................................................................................. 162

CHAPTER 6 GILLET’S STRONG EMERGENCE ..................................................... 163
  6.1 Introduction ............................................................................................................. 163
  6.2 How Gillett Develops His Account ........................................................................ 165
  6.3 Sharpening Gillett’s Account .................................................................................... 170
    6.3.1 Gillett’s Causal Powers—Conditional or Manifest? ............................................ 170
    6.3.2 A Technical Issue with DVR .............................................................................. 173
    6.3.3 Non-Causal Determination ................................................................................ 174
  6.4 Raising Some Concerns with Strong Emergence ................................................. 176
  6.5 Conclusion ............................................................................................................. 184
  6.6 Some Final Thoughts on Ontological Emergence ................................................. 185

BIBLIOGRAPHY ........................................................................................................... 187
PREFACE

As in life so it often is in philosophy, that it is hard to see the forest for the trees. This dissertation is a tour through a very dense forest, with particular attention being paid to some very specific, and often unwieldy, pieces of metaphysical lumber. Since once having embarked upon this tour the reader may find it difficult (as the author himself often did) to find a vantage point from which to see the broader woodlands, I feel it is incumbent upon me to first point out some very general features of the territory this dissertation shall be exploring.

Jerry Fodor once commented that one of the great mysteries of our world is why there are any sciences other than physics. The question does seem pertinent—why are there truths to be found, or explanations to be given, that don’t make reference, either directly or indirectly, via some kind of reduction, to microphysical entities? Broadly speaking, this is the question that serves as the backdrop to my dissertation. More specifically, this dissertation concerns the proposal that the reason there are sciences other than physics is that there are certain features of the natural world which, though ultimately dependent for their existence on the attributes and behavior of microphysical entities, contribute something causally to the world beyond what those microphysical entities contribute. Proposals like this one are generally classified as “theories of emergence.” The specific proposal I deal with here has come to be referred to via the moniker ‘ontological emergence’, and it entails that the emergent entities are properties, more specifically, those properties dependent for
their existence on microphysical properties, but which still make a causal contribution beyond that of these microphysical properties.

All that being said, in this dissertation I will not really answer the question of why there are sciences beyond physics. And, I will not tell you whether there are any ontologically emergent properties, nor even come to any conclusion about which properties would be ontologically emergent were there to be any (though of course I have my suspicions). What I do here is to investigate, dissect, and hopefully clarify the concept of ontological emergence. Thus, this dissertation is more an investigation into the internal challenges such a concept faces rather than external challenges that may be placed on such a theory by other facts or theories. That this is the aim of this dissertation is unsurprising (at least to me) given that this project grew out of my desire to understand just what is being claimed when someone says that a certain feature of the world is “emergent”.

So there is the forest—do your best to be mindful of it as you traverse through the “trees” of supervenience, realization, property constitution, causal efficacy and causal power contribution. These are all concepts central to understanding ontological emergence and to each I devote a great deal of space in this dissertation. I intend this project to be a springboard to future research and theorizing. I hope to have made some of these concepts more clear and sorted out the options that other philosophers have in developing theories of ontological emergence. I myself hope to do just this in the future, as I do in fact believe that some theory of ontological emergence is probably correct in describing the relationship between certain aspects of human mentality and human physicality.
But before turning my eyes toward any future project, there are several people I would like to thank, people without whose assistance, encouragement, and support the completion of this project would not have been possible. First of all, I should like to express my thanks to my directors, Alvin Plantinga and Fritz Warfield, who have provided me with very timely and valuable feedback throughout this entire process, even as deadlines were getting tight and the pace of their work in reading my material greatly increased. The many conversations I had with them certainly helped to sharpen my focus on this project as well as the expression of my ideas. I am also thankful to Mike Loux, Mike Rea, and Leopold Stubenberg for serving as my readers and for the comments they were able to provide.

My interest in the philosophy of mind, and ultimately this project, is undoubtedly attributable to my interactions with two outstanding philosophers, J.P. Moreland and Jaegwon Kim, whom I have tried to emulate both in their passion for the subject and their genuine concern for those with whom they work. I thank both of them for their philosophical and personal inspiration that led to the beginnings of this project. Without J.P.’s pursuit of me for the master’s program at Biola University, I never would have left work as an engineer to study philosophy, and without his continued encouragement, along with the encouragement of another Biola mentor to whom I am indebted, Doug Geivett, I never would have continued along this path. Professor Kim’s work inspired me take a serious look at the metaphysics of mind, and specifically of emergence, and his serving on my proposal committee and commenting upon some of my early work confirmed my commitment to this endeavor and my belief that I could accomplish my goals. I thank all of these men.
I’d also like to thank Montey Holloway and Catherine Robertson for the prompt and always cheerful advice they provided me regarding the procedures required for the ending this project—scheduling of meetings, reminders of deadlines, and the like. Without these two things certainly would not have gone as smoothly. Thanks to Mike Loux for serving as my advisor for my Ph.D. candidacy exams, and who taught me a whole lot about metaphysics. As for others who helped me throughout the long haul of this project, the list is great and yet my appreciation for them is greater. Thanks to Dean Zimmerman, Timothy O’Connor, and Carl Gillett for numerous conversations about the concept of emergence which proved invaluable to my work. I am also grateful to the audience at the 2005 Southern Society of Philosophy and Psychology meeting in Durham, NC, and to my commentator Nick Power, who provided very constructive feedback on an ancestor of Chapter Five.

Thanks beyond what I could possibly express go to my fellow students and friends whose willingness to talk philosophy (or, just as often, willingness not to talk philosophy) was critical to my success, and from whom I benefited both as a scholar and a person. Thad Botham, E.J. Coffman, Tom Crisp, Robert Garcia, Jeff Green, Stephen Grimm, Jennifer and Mark Jensen, Cristian Mihut, Christian Miller, and Donald Smith all aided me in this project in more ways than they could imagine.

Finally, I extend my heartiest thanks to my family. To my parents, who have always encouraged me not only in my academic pursuits but in all my life’s pursuits, I thank you for your love and support. I thank my mom for those many days she came and spent time with Joey so I could have some extra time to work on my dissertation. I am grateful to my son Joey who came along at such a perfect time in our lives and brought such joy and love. I thank Joey for his understanding and acceptance of the
lack of “Papa Time” over the last few months. Finally, my deepest thanks go to my wife Laura. She married an engineer, but wound up with a philosopher. She has been a constant source of faith, hope, and love to me and she has sacrificed much to support me both in this endeavor and in my personal life. As I write this, it is ten years to the day that I asked her to marry me, and other than the day Joey was born, it is still the happiest day of my life. Without Laura’s love and support over the last ten years I never could have come close to reaching my goals. I thank her with all my love.
CHAPTER 1

ONTARIOLOGICAL EMERGENCE: THE BASIC CONCEPTS

1.1 Introduction

The last several years have been witness to an incredible growth in interest across a wide variety of disciplines in what is basically a technical, philosophical concept—that of “emergence”. Writers in fields ranging from philosophy of mind to chaos theory to theology have deployed the concept to categorize certain features of the world that are ostensibly irreducible or non-identical to any more fundamental features and yet are still in some sense dependent upon or determined by the more fundamental features.¹ As often happens with a relatively new term of art, the term ‘emergence’ is not employed univocally in the literature; it is clear that there are numerous distinct concepts captured by the term in different contexts.² The purpose

¹ Examples of those who have deployed the concept in developing theories in the philosophy of mind include Carl Gillett (2002b, 2002c), William Hasker (1999), David Newman (2001), Timothy O’Connor (2000 and 2001), Karl Popper and John Eccles (1977) and Roger Sperry (1969, 1991). In the science of complexity and chaos theory, see for example Kauffman (1995), Newman (1996), and Rueger (2000). And in theology, where the concept has primarily been used in an attempt to understand divine action, see papers in Russell, et. al (1999). Theories advanced in these latter two areas are primarily examples of theories of “epistemological emergence” (EE) whereas most theories that have been discussed under the rubric of philosophy of mind, including the theories advanced by all those listed earlier in this note (David Newman excepted) are instances of “ontological emergence”, the concept of emergence I am dealing with in this dissertation. For more on theories of EE, see footnote 2.

² Robert van Gulick (2001) distinguishes ten different meanings of the term ‘emergence’ and does an admirable job distinguishing them. The primary distinction that van Gulick highlights is that between theories of epistemic or epistemological, emergence (EE), of which he identifies four different types, and theories of ontological emergence, of which he identifies six different types. Theories of
of this dissertation is to identify, elaborate upon, and critically evaluate several versions of one particular concept of emergence. Theories falling under this concept I will be referring to using the moniker ‘Ontological Emergence’, or the more economical ‘OE’. OE bears special philosophical interest, especially to those interested in the metaphysical issues arising in the sciences (first and foremost in psychology, but also in other special sciences like biology and chemistry). The basic tenet of any theory of OE is that there exist some composite systems that instantiate ontologically emergent properties, properties that are distinct from and irreducible to more fundamental properties and yet which are exhibited by the system only when and because it attains a certain complex microphysical state. Exactly what these emergent properties might be, of course, is a matter of controversy, and even among those who

---

3Epistemological emergence generally entail that there are certain facts about the nature of complex systems that cannot be, perhaps even in principle by an idealized knower, predicted given knowledge of the facts governing the basic, lower-level properties and laws of the system. However, theories of EE generally entail nothing about the ontological status of the entities that make these facts true and in some forms are consistent with ontological reduction. For instance, while some, such as Schroder (1998) and Batterman (2001), take epistemic emergence to be a synchronic phenomena, others claim it is only to be found when viewing systems diachronically (i.e. current states of a system are epistemologically emergent if and only if they are not predictable from past states (Bedau 1997, Rueger 2000). But of course that is consistent with every property of a system being reducible to a microphysical property (imagine that certain properties of systems, though ultimately reducible to microphysical properties, were governed by indeterministic laws, so that the state of the system at time t2 could not be predicted from the state of the system at time t1—such a scenario would make the feature at t2 epistemologically emergent, but not ontologically emergent). Theories of EE are much more common in the literature than theories of OE. Virtually all of the authors writing outside of philosophy (even more specifically, those not developing a theory that is being applied to philosophy of mind) develop theories that are best characterized as versions of EE, not OE. An excellent discussion of how both current theories of OE and EE are connected to the work of the British Emergentists can be found in O’Connor and Wong (2002) Stephan (1992) and O’Connor and Wong (2002) also provide valuable taxonomies. My use of the term ‘ontological emergence’ is a bit narrower than van Gulick’s use of the term. I am talking about theories that would fall under the categories he calls “Modest Kind Emergence” or “Radical Kind Emergence.” What he calls “Specific Value Emergence” is really no theory of emergence at all, by my lights.

---

3Others have different labels for the view that I am describing. Van Gulick (2001), in his thorough taxonomy of emergentist views, labels the kind of view I am considering “Modest Kind Emergence” (although it could be that what he calls “Radical Kind Emergence” is consistent with the theory I am discussing as well). Carl Gillett (2002b, 2002c) refers to the type of view discussed in this dissertation as “Strong Emergence” (reserving the term ‘Ontological Emergence’ for a different kind of view entirely). But ‘Ontological Emergence’ seems to be the most common and appropriate name.
espouse Ontological Emergence there is disagreement as to what kinds of properties would qualify as ontologically emergent. However, the properties that are most commonly put forth as candidates for members of the class of ontologically emergent properties are mental properties. More specifically, the suggestion is usually that it is conscious mental properties (e.g., occurrent beliefs, awarenesses, and experiences), or even more specifically qualitative mental properties, the “what it is like” phenomenal character of our experience, that qualify as ontologically emergent properties.

In this first chapter my aim is to, first of all, provide a rough characterization of theories of OE (and thus what is involved in a property’s being ontologically emergent) and, secondly, to review some of the central metaphysical concepts that are involved in or importantly related to OE. This will lead up to Chapter Two in which I take on the task of situating OE along the spectrum of possible solutions to what I shall be calling the “Problem of the Special Sciences”, a problem that has, both historically and in the present day, been one of the primary driving forces behind the development of OE.

1.2 Origins of the Concept of Ontological Emergence

Ontological Emergence has its origins in the work of a cadre of philosophers who have come to be known collectively as the “British Emergentists.” Its roots stretch back to the mid-nineteenth century work of John Stuart Mill (1843) and G.H. Lewes (1875) with interest in the view reaching its zenith in the early part of the twentieth century through the works of philosophers and scientists such as C.D. Broad

---

4 For a thorough and discerning overview of the work of the British Emergentists, see McLaughlin (1992).
(1923, 1925), Lloyd Morgan (1923), and Samuel Alexander (1920). These authors all wrestled with the Problem of the Special Sciences, the task of expounding the relationship between the entities referred to (or expressed) by predicates of lower-level theories (e.g., microphysical theory) and the entities referred to (or expressed) by the predicates of the special sciences. As these authors noted, it seems fairly clear that the laws and theories of the special sciences, inasmuch as they provide robust and indispensable explanations of higher-level phenomena not attainable by reference to physical laws and theories alone, are irreducible to the laws and theories of physics.

Thus, the challenge taken up by the British Emergentists was to draft an ontological framework that supported the intuitive irreducibility of both the theories and properties of the special sciences. This framework included the notion that some properties of complex systems were *ontologically emergent*. Among the properties that the British Emergentists hypothesized to be emergent were certain chemical properties (like the bonding relations between certain atoms), biological properties (like the property being a digestive system, or even the property being alive itself), mental properties, and even the property being a deity. The project, though briefly abandoned in the

---

5 I will leave the concept of a “level” as unexplained and intuitive, except to say that special science properties and theories (e.g., of chemistry, biology, etc.) are at a “higher-level” than microphysical theory. For more on the metaphysics of the ideas of “levels of nature” might be sorted out, see Carleton (1985), Wimsatt (1995), Emmeche, Koppe, and Stjernfelt, (2000), and Kim (2002).

6 I count among the progeny of the British Emergentists contemporary proponents of OE but also many who reject OE but still classify themselves as proponents of Non-Reductive Physicalism (NRP), perhaps the dominant view in philosophy of mind today. In Chapter Two I take some steps toward distinguishing OE from different versions of NRP, but it’s no easy task for views that have been labeled as instances of NRP do not fit nicely into any easily definable category. It’s also clear that some theories consistent with OE are consistent with NRP. More on this in Chapter Two.

7 Lloyd Morgan suggested that the emergence of different properties could be represented by a pyramid with the properties of space-time at the bottom, moving up through material properties, chemical properties, biological properties, mental properties, and culminating at its apex with the property of deity, “an emergent quality that characterizes only certain persons at the highest and latest
middle of the last century, \(^8\) continues on today in the work of those who defend a certain type of non-reductive physicalism. \(^9\) Like their philosophical ancestors, contemporary non-reductive physicalists (whose work is normally focused on explaining the non-reducibility of mental properties) seek a *via media* between ontological reductionism and the view that there are special substances—Cartesian

---

\(^8\) That emergentism fell out of favor and into a period of relative obscurity in the middle part of the last century is due to many factors—the two most prominent being the rise of quantum mechanical explanations, which rendered it apparent that many of the features (primarily of chemistry and biology) that the British Emergentists had taken to be as emergent were in fact not, and the rise of logical positivism, which cast a rather substantial pall upon metaphysics in general (for more on both of these causes of the waning of emergentism, see McLaughlin (1992 and 1997)). But over the last fifteen to twenty years emergence has, one might say, re-emerged as a pivotal player in attempts by philosophers and scientists alike to understand the relationship between the properties, laws, and theories uncovered by physics and those revealed by the special sciences.

\(^9\) Contemporary expositors of OE who attempt to assimilate it to an overall nonreductive physicalist position include Van Cleve (1990), Sperry (1991), Searle (1992), Humphreys (1996, 1997a, 1997b), Crane (2001), Gillett (2002b), and Shoemaker (2002). Not all of these writers call themselves emergentists, however (Searle specifically denies it, for instance, but it seems to me the only way to make sense of his view is as a theory of OE). Jaegwon Kim, perhaps the most prominent and thorough contemporary expositor of Ontological Emergence (though no proponent of the view) has gone as far as equating the rise of nonreductive physicalism/materialism to the rise of emergentism. He writes:

> The fading away of reductionism and the enthronement of nonreductive materialism as the new orthodoxy simply amounts to the resurgence of emergentism – not all of its sometimes quaint and quirky ideas but its core ontological and methodological doctrines. The return of Emergentism is seldom noticed, and much less openly celebrated; it is clear, however, that the fortunes of reductionism correlate inversely with those of Emergentism (*modulo* the rejection of substantival dualism). It is no undue exaggeration to say that we have been under the reign of Emergentism since the early 1970s. (Kim 1999, p. 5)

I imagine this quote will raise the eyebrows of many proponents of non-reductive physicalism, many of who would disavow their views are emergentist (for example, Derk Pereboom, whose position will be discussed in detail in Chapter Four). But it seems clear that Kim’s comments here are motivated by his belief that other, nonemergentist forms of nonreductive physicalism, like the token identity theory or certain forms of functionalism, or “supervenience physicalism” fail as accounts of special science properties. Of course many will disagree with him here, and Kim’s arguments against such theories are complicated and controversial. They appear in numerous places. See for example Kim (1993) and Kim (1998).
souls, entelechies, élan vital, etc—from which higher-level, non-reducible properties (the ostensible referents of special-science predicates) emanate.\textsuperscript{10}

1.3 What Is Ontological Emergence?

Though there is far from total agreement as to what conditions a property must meet to qualify as ontologically emergent, there is a certain core conception common to nearly all formulations of the view. As it turns out, the philosophical concept is not that far off from how the common man uses the term ‘emergent’. For the common man to say that one entity or phenomenon emerges from another is to say that the former thing is distinct from the latter and yet in some way has its origins in the latter. So, for instance, to say that a caterpillar, after spending some time in its cocoon, emerges as a butterfly is to say that being a butterfly is something different than being a caterpillar, but that the butterfly’s existence (or perhaps the instancing of the property \textit{being a butterfly}) somehow depends on the caterpillar (or an instance of the property \textit{being a caterpillar}).\textsuperscript{11} And to say that a widespread panic emerged from a series of outside pressures placed on the stock market is to say that the panic is something above and beyond the pressures, but nevertheless dependent upon them for its existence.

\textsuperscript{10} And none of this is to say that ontological emergence is inconsistent with dualism, at least certain forms of dualism. On the contrary, several authors of late have defended theories of ontological emergence consistent with some form of serous dualism. Jaegwon Kim describes any serious dualism as one that denies the supervenience of mental properties upon physical properties, which Kim takes to be the minimal necessary condition on physicalism, the view that all concrete substances and properties are physical (Mind in a Physical World,15). Cartesian dualism would be an example of such a dualism, but less radical forms incorporating ontological emergence have recently been defended by people such as Hasker (1999), Lowe (2000), O’Connor (2000), and O’Connor and Jacobs (2004).

\textsuperscript{11} Throughout my dissertation I will place names of properties (like \textit{being a caterpillar} and \textit{being a butterfly}) in italics.
I do not intend here to launch into a lengthy analysis of the everyday concept of “emergence”, but merely to point out that the philosophical concept has much in common with the ordinary. Ontologically emergent properties, for one, are dependent upon other properties, but not reducible to them. In speaking of the philosophical concept, Robert Van Gulick writes:

The basic idea of emergence is more or less the converse of that associated with reduction. If the core idea of reduction is that Xs are ‘nothing more than Ys’ or ‘just special sorts of Ys’, then the core idea of emergence is that ‘Xs are more than just Ys’ and that ‘Xs are something over and above Ys’. (2001, p. 16).

So one component of the concept of ontological emergence is that an ontologically emergent property is something “over and above”. But over and above what? Note that implicit in this core notion of the philosophical conception of emergence is that if a certain feature is emergent it is emergent relative to some other feature—it does not really make sense to say of something that it is emergent simpliciter. Ontologically emergent properties are emergent relative to properties of some lower-level, and this is normally assumed to be the microphysical level (more on what is meant by ‘microphysics’ and ‘microphysical’ in the next section). Here is where the philosophical concept begins to diverge from the everyday concept. It would not be out of the ordinary to say that, for instance, when red paint is mixed with yellow paint, an orange paint “emerges”. But it would not be proper to say that the property being orange is, in the philosophical usage of the term, emergent from the properties being red and being yellow. This is because being yellow, being red, and being orange are all properties “of the same level”, or, to put it another way, are all properties of the
same kind (namely, colors).\(^\text{12}\) An emergent property cannot be of the same level or kind as the property/properties from which it emerges.\(^\text{13}\)

Another way that the philosophical concept of emergence departs from the ordinary is that ontologically emergent properties must be in some sense “novel” in their causal contribution. As Lloyd Morgan puts it:

In an organism within which consciousness is emergent a new course of events depends on its presence. In a person in whom reflective thought is emergent behaviour is sustained at a higher level…Strike out reflective consciousness and action is of a lower impulsive order. Strike out all guiding consciousness and behaviour is that appropriate to the level of life. Strike out life and the course of events drops down to the physical level. (1923, p. 17)

Morgan is noting that because objects exhibit properties like \textit{being alive} and \textit{being conscious}, the subsequent course of causally produced events in the world is different. I will be reflecting upon and refining this idea of novel causal contribution throughout the dissertation. As with the term ‘emergent’ the locution ‘novel causal contribution’ certainly denotes different ideas in different contexts, but for now I simply note it is an important aspect to a property’s being ontologically emergent that goes beyond the ordinary concept of emergence. Whereas one might say that a statue “emerges” from the clay as it is molded by the sculptor, it is not the case that \textit{being a statue} is an

\(^{12}\) Of course I am speaking somewhat loosely here. Clearly \textit{being yellow} and \textit{being orange} are both members of several kinds—the kind of “being a property” comes to mind—but clearly the kinds to which they both belong are not the most \textit{relevant} kinds in this situation. If we were to try to make this idea more precise, we could perhaps say that a property P cannot be emergent from a property Q if the kind of property of which P is most immediately determinate is identical to the kind of property of which Q is most immediately determinate. Since \textit{being yellow} and \textit{being orange} are both most immediately determinate of the property/kind \textit{being colored}, one cannot be emergent from the other.

\(^{13}\) One familiar with the literature might be pondering whether this example reveals another sense in which the philosophical concept is different from the vulgar, namely the former is indifferent to whether emergence is a synchronic or a diachronic relation, whereas the latter entails that emergence is only a synchronic phenomenon. But I demur over whether synchronicity is part of the core conception of an ontologically emergent property. Certainly most philosophers have included synchronic dependence on the microphysical as a touchstone of emergence, but not everyone agrees that emergent properties are so dependent. It is one of those controversial topics that will be revisited later.
ontologically emergent property, for this property contributes nothing causally beyond the properties of the objects that compose it.\textsuperscript{14}

So the core elements of the concept of an ontologically emergent property are that an instantiation of such a property is dependent upon yet not reducible to instantiations of lower-level (e.g., microphysical) properties and is somehow novel in its causal contribution. But there has been significant disagreement on how exactly to flesh out the notions of non-reduction, microphysical dependence, and novel causal contribution. There has also been significant disagreement as to whether the conjunction of these notions alone is sufficient to characterize an ontologically emergent property. Now some of this disagreement is clearly attributable to disagreement about the nature of the concept itself, and I shall return to that in a moment. But the disparity is equally attributable, I believe, to significant disagreement over some of the key metaphysical concepts employed in the debate. That there is a need for a thoroughgoing investigation into the metaphysics of ontological emergence is evidenced by the fact that contemporary expositions of the nature of an ontologically emergent property vary widely—even to the point of being logically inconsistent with one another. Let me give just one example that should help bring out the importance of analyzing some of these key metaphysical concepts.

There is disagreement as to whether a property’s being ontologically emergent is consistent with its being a realized property, a disagreement rooted not so much in the core conception of ontological emergence, but in a disagreement over the nature of

\textsuperscript{14} This is not to say that the property contributes nothing causally. Maybe what it contributes causally is identical to what properties of its constituents, or perhaps a property of the clay, contributes. If so then there may be some type of causal overdetermination involved.
realization itself, specifically, whether a property’s being realized is sufficient for its being ontologically reducible. For instance, in describing the ontological emergence relation, Sydney Shoemaker presents the relation an emergent property bears to its base properties as pretty much the opposite of that a realized property bears to its base, writing that, “clearly this [emergence] will not be a relation of realization—that we have emergence here is precisely the fact that the emergent properties are not realized in the properties they supervene on” (2001, p. 57). Jaegwon Kim agrees with Shoemaker on this point. Kim equates a property’s being realized with its being functionalizable, or being a functional property. According to Kim, a property is a functional property if and only if its essence consists in the having of one or another property out of a set of properties \( P_1 \ldots P_n \), where each of \( P_1 \ldots P_n \) are such that in circumstances \( B \) it plays a certain functional role.\(^{15} \) In \( B \) instances of the \( P \)’s are such that each, when instantiated, is the causal product of certain set of events and causes another set of events.\(^{16} \) An example of a functional property would be the property of \textit{being a mousetrap}. Many different structural configurations could realize this property, i.e., there are many different “structural properties” that are such that they are the causal product of a certain set of events (in this case, probably, the events

\(^{15} \) Throughout the dissertation, I will be using capital letters to represent both properties and sets of circumstances. However, I will use capital letters in bold lettering to refer to the latter.

\(^{16} \) Functionalization is one type of realization relation. The exact nature of realization will be discussed in more detail in subsequent chapters—a precise characterization is not necessary to my point here. Generally speaking, though, realization is a dependence relation that is a step-up from mere supervenience—to say that a property \( A \) realizes a property \( B \) is to say roughly that \( B \) belongs to a set of properties supervenient on the set of properties to which \( A \) belongs, AND that \( B \) is instantiated in virtue of \( A \) being instantiated, where the in virtue of is something stronger than mere nomological dependence. For example, the property of being red realizes the property of being colored, and, as already mentioned, a functional property is realized by the property that plays its causal role.
involved in the design and construction of the mousetrap) and produce certain other events (the trapping of mice among them). Kim concludes that:

Functionalization of a property is both necessary and sufficient for reduction. This accords well with the classic doctrines of emergentism...However, emergentism may yet be an empty doctrine...Many scientists have argued that certain “self-organizing” phenomena of organic, living systems are emergent. But it is not clear that these are emergent in our sense of nonfunctionalizability. (1998, p. 18)

Kim’s proposal here more or less reflects the standard assumption in the literature that a property P’s being realized by another property (or set of properties) is sufficient either for P’s being reducible to that property, or for the denial of the proposition that P makes any novel causal contribution beyond that of its realizers. But there is at least one current player in the discussion on ontological emergence, Carl Gillett, who disagrees with this assessment of the realization relation. Gillett thinks that once one is clear on the exact nature of the realization relation, one will see that X’s being realized by Y does not preclude X’s making a novel causal contribution. He thus concludes that a property’s being ontologically emergent is consistent with its being realized—in fact, his preferred characterization of an ontologically emergent property incorporates as a necessary condition that such a property be realized by microphysical properties. So here we have a case where differences in opinion as to the nature of an ontologically emergent property are rooted in a difference of opinion about the metaphysics of the realization relation. Clearly deeper investigation into the nature of that relation would be helpful, and this is a project I shall undertake in Chapter Four, and continue in Chapters Five and Six as I look at two accounts of OE that incorporate the idea that ontologically emergent properties are realized properties.
There are many other examples of confusion and/or disagreement about metaphysical concepts that leads to divergent characterizations of an ontologically emergent property. For example, there is disagreement as to whether a property P’s being ontologically emergent is consistent with P’s being a structural property, or whether P’s being ontologically emergent requires P’s contributing a novel causal power, issues that both hang on how one understands the metaphysics of these entities, and issues that will both resurface in Chapter Three. But an equally important factor accounting for the disparity in analyses of an ontologically emergent property is disagreement over the components of the concept of ontological emergence itself, and whether, and how far, those components go beyond nonreduction, microphysical dependence, and novel causal contribution. For instance, there has been significant disagreement as to whether the very concept of a property P’s being ontologically emergent should include the idea that instances of P are nomologically supervenient on microphysical properties, or that instances of P are not theoretically predictable given a complete knowledge of the distribution microphysical properties.\(^{17}\) Disagreement over these issues I think is largely traceable to a methodological consideration in formulating a theory of OE. The more or less standard procedure has been to diligently mine the works of the British Emergentists and from there extrapolate a set of necessary conditions on a property’s being ontologically emergent.

\(^{17}\) Those who suggest these necessary conditions include Kim (1999), McLaughlin (1992, 1997), O’Connor (1994), Shoemaker (2002), and Stephan (1992). Those who disagree include O’Connor (2000 and 2001), Gillett (2002b, 2002c), and Silberstein and McGeever (1999). Note the disagreement here is not over whether a property’s being ontologically emergent entails any of these things, but rather the concept itself includes these features. It may turn out, say, if physicalism is true, that all ontologically emergent properties are nomologically supervenient on microphysical properties, but it seems questionable that such a condition should be built in to the concept of an ontologically emergent property.
This strategy certainly has its merits (one certainly wants one’s version of OE to be as faithful as possible to those who originated the view), but it also faces limitations. First of all, the British Emergentists did not employ in their writings many of the metaphysical concepts we take for granted today—concepts crucial to the contemporary debate including “ontological reduction”, “causal efficacy” and “causal powers.” Thus, formulating the views of the British Emergentists in modern terms often requires imposing our concepts upon their terminology in ways it is not clear they intended. Second, reliance on the ideas of the British Emergentists has led to a conception of an ontologically emergent property that is too strong, i.e. one that disqualifies by definition certain properties that may be good candidates for in fact being emergent properties. For instance, the idea that emergent properties must be nomologically supervenient upon microphysical properties, as opposed to bearing some other type of dependence relation microphysical properties, like causal dependence, is clearly a central aspect of ontological emergence as conceived by the British Emergentists, and thus has made its way into the most common contemporary formulations. But it seems to me adding this to the concept of ontological emergence is unnecessary as the concept can be distinguished from rival concepts without the inclusion of this condition (as I hope to show in Chapter Two).

The alternate strategy that has been employed acknowledges the value in gaining an understanding of these historical works, but places more emphasis on

---

18 Chapter Three is devoted to the British Emergentist conception of ontological emergence, what I call **Classic Ontological Emergence**, and the way that has influenced Jaegwon Kim’s development of, and criticisms of, the view.
framing OE as a position distinct from other contemporary nonreductionist views. It is this camp I find myself falling in line with, and I believe what is needed is a more modest characterization of OE (what I will be referring to as Minimal Ontological Emergence, or MOE) than what dominates the literature today. This will be a characterization which, though still in line with the spirit and the core ideas of the British Emergentists, does not build so many necessary conditions into a property’s being emergent. By my lights the best way to understand what MOE should look like is to examine the main problem that ontological emergence is supposed to solve—the Problem of the Special Sciences. By examining this problem, we can see more clearly what OE is not, and thus learn the space in which we have to maneuver developing OE as a unique solution to this problem. I thus will devote the second chapter of this dissertation to situating OE with respect to other theories of higher-level, special-science properties on offer today. In particular, I want to compare OE to what I take to be its three chief rivals—qualified micro-determinism, epiphenomenalism, and serious substantival dualism. But before delving into the Problem of the Special Sciences in the next chapter, it will be important to review a few assumptions and explain and analyze some key metaphysical concepts that are relevant to any discussion of Ontological Emergence. That is the task of the remainder of this first chapter.

---

19 I include in this group people like Humphries (1997a, 1997b), Gillett (2002b, 2002c), and the more recent work by O’Connor (2000) and O’Connor and Wong (2002 and forthcoming).
1.4 Clarifying Some Important Assumptions and Concepts

1.4.1 The Microphysical

First, as indicated earlier I shall be assuming that the basic features of our physical world are those characterizing the most fundamental physical particulars of the universe, the entities of microphysics, the properties of such particulars, the laws characterizing the interactions of such particulars, etc. I shall refer to such basic properties, theories, and laws as the microphysical.\(^2\) I take microphysics to be the science of the most fundamental particles of the spatiotemporal universe and the microphysical level to be the objects and properties that characterize microphysics. As Lynne Rudder Baker notes, the properties we find in the set of microphysical properties need not be, as she puts it, “well-behaved.” Some properties in this set may not be mentioned in any of the laws of microphysics. Some may be such that they are “hideously complex, a Boolean combination of properties” and “not necessarily a natural kind” (Baker 1993, p. 81). It could also be that many members of the set of microphysical properties are multi-placed relational properties instantiated by a group of microphysical particles (for instance, the shapes and structures of combinations of microphysical particles might be relational microphysical properties).

Although this characterization of microphysics is certainly incomplete, and probably contestable, and although there is certainly some work to be done in sorting out what it means, for instance, for an entity to be microphysical or for a property to

---

\(^2\) The reason I speak here of the basic features of the physical world, and not of basic features simpliciter is that if something like Cartesian dualism is true, then there would be basic features of the world exemplified by nonphysical objects (souls or the like). These features, though fundamental, would not properly be classified as microphysical.
be a microphysical property (for instance, whether microphysical theory should be identified with current science, or a theoretically completed future science), what I have said about microphysical properties is more or less taken for granted in the literature on emergence, and thus I find it a fair place to start my discussion. The reason that ontologically emergent properties are normally characterized as being (ultimately at least) emergent from the microphysical is that the relation of dependence that links emergent properties and their lower-level “base” properties is almost always understood to be transitive. Thus if A is an emergent property that is emergent relative to another set S of emergent, nonmicrophysical properties, which are in turn emergent relative to set of the microphysical properties, then A is ultimately emergent relative to the microphysical as well. Of course there is the possibility (the epistemic possibility, at least) that there are no basic features of the spatiotemporal universe, that there may be no bottom level at all (if, for instance, matter is infinitely divisible, and every property instance owes its existence to some more basic property instance or instances). But this possibility is not of great concern to those who espouse a theory of Ontological Emergence. The central tenet of a theory of Ontological Emergence is that there are properties emergent relative to other properties. That ‘relative to’ is usually not tacked on to the word ‘emergent’ in the literature simply reflects the assumption that what is being spoken of is emergence relative to the microphysical. However, even if were is no bottom microphysical level, there could still be properties at the n+1 level emergent relative to properties at the n level, and everything said in this dissertation could sill apply to this scheme.
1.4.2 Properties, Property Instance, and Causal Relations

Second, I shall be assuming that at least one type of feature of our world that makes a causal difference is properties. If a property does make a causal difference it does so via one of its instances, where an instance of a property is not the object that exemplifies it, but the particular manifestation of the property, i.e. a property token. A property token might be an event or state of affairs, in which case it would be something like s-instantiating-P, or a property instance could be a trope. One’s overall metaphysics of properties would determine the choice here. It is property instances that I take to be the relata of causal relations. Throughout the dissertation, unless otherwise specified, properties will be represented by capital letters and property instances by small letters.

1.4.3 Causal Powers

I will also assume that if (but not necessarily only if) an instance of a property Q stands in—or is involved in an event that stands in—a causal relation to some other property instance or event, that Q possesses, or contributes, a causal power. Sydney Shoemaker, perhaps the best known expositor of causal powers, describes a causal power by saying that “for something to have a power…is for it to be such that its presence in circumstances of a particular sort will have certain effects. One can think of such a power as a function from circumstances to effects” (1980, p. 113). The

21 If property instances are tropes, then the following is true: For any property P such that p is one of P’s instances, then p = P. On most formulations of trope theory, there is no distinction between a property and its instance, since properties just are individual particulars.

22 I insert this caveat for if a property instance is a trope it may be that the true relata of the causal relation is not the trope itself but rather an event involving the trope (if p is a trope, then s-instantiating-p may be the relata of the causal relation).
primary reason I will employ causal power talk, and will generally choose to speak of causal powers rather than instances of causation or causal laws, is that it reflects the standard parlance in the literature on ontological emergence. Furthermore, the term ‘causal power’ is often, perhaps we could even say usually, employed not merely as shorthand for a property’s being causally efficacious (or potentially causally efficacious), but as a reflection of the belief that causal powers are part of the basic ontological structure of the universe. Though it is possible to employ the term ‘causal power’ without much ontological import, the fact of the matter is that ‘causal power’ has been used by many theorists to denote entities of a unique ontological category whose possession or contribution by certain properties marks those properties off as irreducible, and that is how I shall be using the term in this dissertation.23 It is important then to examine more carefully the metaphysics of causal powers and what it means for a property to possess or contribute a causal power.

Now above I said that if an instance of a property Q stands in—or is involved in an event that stands in—a causal relation to some other property instance or event, that Q possesses, or contributes, a causal power. But which is it? This is a difficult question to answer, for it is safe to say that there is no agreement in the literature to which entities causal powers should be attributed. One can find in the literature references to properties possessing causal powers as well as substances possessing causal powers—even to events possessing causal powers. Usually, if it is said that a substance or particular possesses a causal power, it is because that power is

23 For instance, that a causal power is a genuine ontological feature of the world plays a central role in at least one of the accounts of OE I will be looking at in this dissertation, that of Derk Pereboom (2002). Others, like Sydney Shoemaker and Carl Gillett, though never explicitly committing themselves to this understanding of a causal power, seem to make use of it in their accounts as well.)
contributed to it by a property. What is fairly clear in all this is that properties do play a central role in the possession of causal powers—either by possessing causal powers themselves or by contributing them to their proper possessors (be it substances or events). By my lights, it is a bit odd to say that a property “contributes” a causal power, as it seems to imply that the causal power is transferred from one thing to another, i.e. from the property to a substance or event. It seems better to think of a property “bringing a causal power along with it”, either by its actually being a constituent of a substance or event, or by its being instantiated. But if it brings the property along with it, it must possess the power. So I think ‘possession’ is probably a better term in describing the relationship between properties and their causal powers. Nevertheless, it is ‘contribution’ that has gained more currency in the literature, particularly by authors (e.g. Carl Gillett) who I will be discussing later in the dissertation. Therefore, I shall generally speak of causal power contribution, but of course the reader should understand that, by my lights, the way a property contributes a causal power is by first possessing it.

1.4.3.1 Conditional vs. Manifest Powers

The *locus classicus* of the notion of a causal power, at least when it comes to contemporary discussion, is Sydney Shoemaker’s 1980 work “Causality and Properties.” Therein Shoemaker defends the view that properties are to be identified with clusters of what he calls “conditional causal powers.”

---

24 In his more recent work relating to the causal theory of properties—Shoemaker (1997)—Shoemaker defends the position that properties are clusters of both, as he calls them, “forward-looking” and “backward-looking” causal powers. The “forward-looking” powers of a property P are the ones more traditionally understood to be causal powers, those that are responsible for P’s causing
given causal powers a fairly rigorous analysis.\textsuperscript{25} Both authors distinguish two different ways an object may have a causal power: conditionally or manifestly. Shoemaker gives the following sufficient condition for an object’s having a causal power conditionally:

Let us say that an object has power $P$ conditionally upon the possession of the properties in set $Q$ if it has some property $r$ such that having the properties in $Q$ together with $r$ is causally sufficient for having $P$, while having the properties in $Q$ is not by itself causally sufficient for having $P$…When a thing has a power conditionally upon the possession of certain properties, let us say that this amounts to its having a \textit{conditional power}. (1980, p. 115)

Elsewhere, Shoemaker interprets his previous self as meaning that “something has a conditional power if it is such that it would have a certain causal power if it had such and such other properties, where the possession of those properties is not itself sufficient to bestow that power” (1997, p. 63). An example (used by Shoemaker) should help illustrate. Suppose $P$ is the property of \textit{being knife-shaped} and $C$ is the power to cut wood. Shoemaker notes that an object $O$ that has $P$ does not, simply by having $P$ alone, have $C$. Having $C$ is conditional upon instantiating not only $P$, but numerous other properties as well, e.g. \textit{being larger than a water molecule, not being made of butter}, etc. $C$ here is an example of what Shoemaker calls a causal power \textit{simpliciter}—the power $O$ manifests when it is actually capable of cutting. But, there is another power that $O$ has, call it $C^*$, a conditional power to cut wood if knife-sized and made of steel. $C^*$ is not the same power as $C$. To paraphrase Shoemaker’s 1997 quote above, something has $C^*$ if it is such that it would have $C$ upon the possession

\begin{footnotesize}
\begin{itemize}
\item some effects. The “backward-looking” powers are powers associated with $P$ that are powers to be \textit{caused} by certain other properties. Shoemaker does this to block a couple of objections to his general theory of properties. But here I am only concerned with “forward-looking” powers.
\item Armstrong (1997, especially pp. 70-73). Armstrong (1968) also contains a fair amount on this issue.
\end{itemize}
\end{footnotesize}
of some other properties (in this case *being larger than a water molecule, not being made of butter*, and many others).

So C* is a conditional power, and C is a causal power *simpliciter*. Shoemaker does not say much more about powers *simpliciter*, other than to say that they are a special case of conditional powers. This is somewhat odd. If all powers are ultimately conditional why take pains to mark out the category of causal powers *simpliciter*? It seems Shoemaker is wrong in saying that causal powers *simpliciter* are a special case of conditional power; I think we can understand them as a completely separate category of power. Armstrong can be of assistance here in distinguishing *conditional causal powers* from a separate class of causal powers, what I will call *manifest causal powers*. Armstrong has this to say about a causal power being had manifestly:

A particular may have a disposition or power, but may fail to *manifest* that disposition or power. This, indeed, is the normal thing...When a particular has an unmanifested power, then the particular cannot be *related* to the potential manifestation of this power because the instantiation of a relation demands that all its terms exist...Suppose, however, that a particular does manifest a power that it has. This always is a case of *causation*, singular or token causation. For a brittle object to manifest its brittleness, the striking of the object, in conjunction no doubt with suitable standing conditions or other circumstances together with the brittleness itself, must *bring about* the shattering of the brittle object. (1997, pp. 70-71)

A causal power is had by an object manifestly, then, only if it is somehow involved in an instance of causation.\(^2^6\) In such a situation, we could say that the property P that is

\(^{26}\) I think it is probably most natural to interpret Shoemaker’s use of the term ‘object’ and Armstrong’s use of the term ‘particular’ here as referring to substances, but I suppose one could understand these terms as referring to events. In the literature on emergence (and mental causation in general), analyses of causal powers and causal power contribution in terms of powers being had by objects has not gained much currency. This is primarily due to the fact that philosophers are more likely to speak of *properties* having or contributing causal powers, and not *objects*. I have no grand explanation as to why causal powers are usually attributed to properties, but it is most likely due to the fact that it follows from two very commonly held beliefs—first, that the relata of causal relations are events or property instances, and not substances, and, second, that the things that have causal powers or to which causal powers are contributed should be the things that are the relata of causal relations. From
involved in the instance of causation possesses a *manifest causal power.*\(^\text{27}\) We therefore have two types of causal powers: conditional causal powers, which are had by objects (substances or events) whenever they exemplify a property that has such a conditional power, and manifest causal powers, exemplified by objects only when involved in an instance of causation. In the section after the next, I want to try to formalize the ideas of property causal power contribution, as this will serve us well throughout the rest of the dissertation. But first, let me turn to the question of the ontological status of causal powers—do causal powers exist as genuine entities in the world, and, if so, what kind of entities are they?

### 1.4.3.2 The Ontological Status of Causal Powers

We *could* interpret the aim of Shoemaker and Armstrong in their introduction of causal power talk to be that of simply defining certain terms like ‘x has power P’ these two beliefs it follow that causal powers are had by or contributed to events or property instances, and not substances. In any case more recent analyses of causal power contribution have focused on the role properties play in such contribution.

\(^\text{27}\) At this point the reader may be straining toward an objection. Isn’t it the case that all causal powers are possessed conditionally? After all, isn’t the fact that an object O possesses a power (say the power to cut) conditional on O’s possessing the property of *being such that there is a material world.* My response is that it is clear that P’s contributing the manifest power to cut is conditional upon O’s possession of the aforementioned property, but not any conditional property. As I will out the notion of a conditional power below, *no* conditional power is contributed by a property conditionally, rather conditional powers are part of the essence of a property. What is conditional about conditional powers is not there contribution (or possession) by properties, but that they result in an instance of causation. For a conditional power to result in an instance of causation, certain other properties must be instantiated. Perhaps the term ‘conditional causal power’ is unfortunate here, but that is what’s been handed to us and I want to change a piece of terminology that is so frequently employed in the literature.

Another reason that I have for asking the readers indulgence with this distinction is that one of the accounts I will be reviewing later in this paper, that of Carl Gillett, seems to suffer a serious blow if the only types of causal powers at Gillett’s disposal are conditional causal powers. So I am introducing the idea of a manifest causal power here with hopes of possibly remedying Gillett’s account in Chapter Six. Whether the remedy works or not will of course be an open question, but it is one of the reasons I have for making the distinction between two different types of causal powers.
without taking those terms to have any ontological import. For example, a claim such as “x has a power C” could just be read as elliptical for something like “when instances of x are in circumstances K, they cause instances of E” without implying there is some genuine entity referred to by the term ‘power C’. This interpretation of Shoemaker and Armstrong attributes to them a deflationary view of causal powers. On such an interpretation, the truth maker for the statement ‘x has power C’ is not x’s having or possessing some entity—a causal power—for there really are no such things; rather, the truth maker for the statement ‘x has power C’ is something like x’s having some property P combined with certain truths (probably counterfactual) about the causal relations in which instances of P stand (or could stand).

Though I think there may be some advantages to interpreting causal power talk in this fashion (especially talk of manifest causal powers, as I shall explain shortly), this is clearly not how most people in the emergence literature understand causal powers, nor how Shoemaker himself seems to understand causal powers. Most are realists about causal powers, or at least about conditional causal powers, which are the ones most frequently discussed in the literature. Shoemaker thinks properties are

---

28 However, though he is a bit cagey on the issue (he never comes right out and says it), the most straightforward reading of D.M. Armstrong is that he advocates a deflationary view of causal powers, a view he refers to it as the Soft doctrine of powers. Armstrong defends Categoricalism about properties, the thesis that all true properties are non-dispositional. Armstrong takes the opposite thesis, Dispositionalism, to be the view that each property is either identical to a causal power or constituted of causal powers. So, the defender of Categoricalism, Armstrong says, must for true statements employing causal power talk “find truthmakers, entities in the world that correspond to these truths, without the aid of properties conceived as the Dispositionalist conceives properties.” (Armstrong 1997, p. 81) In other words, the Categoricalist must find a way of explaining the truth of a statement like ‘x contributes causal power C(K,E)’ without making reference to causal powers as genuine entities in the world. He then goes on to say what these truthmakers would be states of affairs including S-instantiating-P, there being a law that P causes E in circumstances K, it’s being possible that K occur, and perhaps some others. I am rather attracted to this deflationary view of causal powers, mostly because I am attracted to Categoricalism about properties. But that is a side issue I do not have the space to explore now.
“clusters” of conditional causal powers, somehow composed of these powers, and since properties are genuine features of the world, so are causal powers. And as we shall see in later chapters, theories of ontological emergence presented by Derk Pereboom and Carl Gillett depend upon the idea that causal powers are genuine features of the world.

If causal powers are part of the fundamental ontological building blocks of the universe, then many interesting questions arise. What ontological category do they fall under? Which of the causal power concepts (conditional or manifest) match up to a real world entity, i.e., are conditional causal powers part of the ontological structure of the world? Are manifest causal powers? Are both? As to the first question, there seems to be four options for what a causal power C contributed by property P could be:

(a) a property exemplified by the same object S that exemplifies P
(b) a property exemplified by P itself
(c) a constituent of P
(d) Some ontological category completely different than those described in (a), (b), or (c).

Consider first (a). This view might initially appear quite attractive. We might even suppose that causal powers are properties that supervene upon regular, causally efficacious properties, with the type of necessity entailed by the supervenience relation

---

29 But actually Shoemaker has recently softened his stance on this issue somewhat. He now formulates the Causal Theory of Properties not as entailing that properties are clusters of causal powers, but that the causal powers of properties are essential to them, and that properties having all the same causal powers are identical. Saying that the causal powers of properties are essential to them probably does not commit him to quantifying over causal powers, he could adopt a deflationary sense of causal powers.
being either nomological or metaphysical, depending on one’s view of the modal relationship between properties and their causal powers. However, a significant problem looms. If a causal power C is a property exemplified by S, then we have the following two property instances: S-exemplifying-P-at-t, which we can call ‘p’, and S-exemplifying-C-at-t, which we shall call ‘c’. Suppose then that circumstances K are actualized and an instance of property E results. What is the cause of this instance of E? Is it p or c? Of course based on our definition of causal power contribution p has to be a cause. Assuming p and c don’t causally overdetermine the effect, then c is otiose—though it is a causal power it contributes nothing causally. This seems very odd—it sure seems that if one were to include causal powers in one’s ontology then one thing they would have to be is causally efficacious. But it seems that c has absolutely nothing at all to do. Of course one might say that c does have one thing to do, namely, serve as the truthmaker for the statement ‘P contributes C’. But it seems a metaphysical extravagance to posit the existence of entities simply to serve as truthmakers if those entities have no other function.

The same problem arises if we take option (d). No matter what novel ontological category C falls under (and I’m not sure what it might be), the instances of the causal power will either not be causally efficacious or it will causally overdetermine the instance of E. I take it that neither option is acceptable.

So, we are left with (b) and (c) as the most likely options. Both suggest that a causal power is a type of entity that characterizes the property contributing it, either as

---

30 If this were the approach we took, then perhaps conditional causal powers could simply be identified with dispositions, and perhaps something like the functional view of dispositions could be upheld, perhaps similar to the one advocated by David Lewis (see 1986a, pp. 223-4).
a property of that property or as a constituent of it (and this seems to be what Shoemaker is getting at in calling properties “clusters” of causal powers). I can think of no conclusive evidence to sway us one way or the other, although I do think the problems that arise from viewing the relationship between causal powers and properties along the lines of (c) may be too daunting for many to embrace the view. For instance, it seems that a given causal power could be contributed by different properties. For instance, the causal power to crush a metal can in certain circumstances may be contributed by the property having mass of 500 kg and having mass of 5000 kg. Is the causal power a constituent of both of these properties? Does that mean the properties overlap in some fashion? Furthermore, if causal powers fail to be essentially contributed by properties (as those who eschew dispositionalism believe) and if causal powers are related to properties as described in (c), then a property could be constituted by different entities in different worlds, which many may find too unseemly. Of course, a similar problem afflicts (b). If it is true that it is a contingent matter what powers a property contribute (a thesis one might think follows from the contingency of the natural (causal) law) then (b) would imply that a property would have certain properties—its causal powers—contingently. Now in general it’s not a problem to suppose that a property has some properties contingently; for instance, a relational property like being Al’s favorite property, which the property of being scalable by 70-year old humans has in the actual world, is had contingently. But if causal powers are intrinsic properties, then it seems odd that they could be had contingently. However, I think there is a way to understand property causal power contribution that is consistent both with the laws of nature being contingent and a
property’s conditional causal powers being essential features of the property. I propose this understanding in the next section.

1.4.3.3 An Analysis of Property Causal Power Contribution

Suppose we understand what it is for a property to contribute (or possesses) a conditional causal power as follows:

**Conditional Causal Power Contribution:** A property P contributes a conditional causal power C(K,E) just in case an instance of P, were it to occur in circumstances K, would cause (or contribute to causing) an instance of E.\(^{31}\)

Note that just as an object’s possession of a conditional power is dependent upon certain circumstances, so a property’s contributing a conditional causal power is referenced to circumstances. In **Conditional Causal Power Contribution,** K is the “domain”, so to speak, in which the conditional causal power is “manifested”, and E is the property whose instances are caused (or partially caused) by instances of P when

---

\(^{31}\) This way of formulating causal power contribution is inspired by Jessica Wilson (2002). She advances a principle she calls Causal Powers Bestowal:

A property P contributes causal power C(K,E) just in case (i) instances of P, in circumstances K, cause (or contribute to causing) instances of E, and (ii) the holding of K alone does not cause (or contribute to causing) instances of E. (2002, p. 61)

My principle involves a reformulation of (i) to make it clear that it is a counterfactual condition, and also drops (ii). Condition (ii) seems extraneous, and it actually leads to a problem. The inclusion of condition (ii) in Wilson’s analysis suggests that K should entail all of the facts relevant to E’s being caused other than the facts about P, for the property instances that K entails to exist are not supposed to be sufficient for P. But making condition (ii) part of the analysis of causal power contribution would rule out genuinely overdetermining causes from contributing causal powers. Suppose an instance of E is overdetermined—it is caused both by an instance of Q and an instance of R. Does property Q contribute causal power C(K,E) where K entails all facts relevant to E’s instantiation other than facts about Q? No, for the obtaining of K alone entails that R exists, and R is a cause of an instance of E. But the solution here is straightforward—simply drop (ii) from the analysis of what it is for P to contribute a causal power. The reason (ii) was there in the first place was simply to mirror Shoemaker’s suggestion that an object has a power conditionally only if it has some property r such that having the properties in Q together with r is causally sufficient for having the power, while having the properties in Q is not by itself causally sufficient for having the power. But Wilson is not talking about causal sufficiency to have powers, but about causal sufficiency to produce certain effects.
**K** is actualized. I will return to the question of exactly what **K** should include momentarily, but it does seems fairly clear that conditional causal powers need to be referenced to some circumstances **K**. The property of being scissors-shaped (call it Pss) contributes on every occasion it is instantiated a conditional causal power relevant to the cutting of paper, but it does not, on every occasion it is instantiated, contribute the manifest causal power to cut paper *simpliciter*. Rather, Pss contributes the following conditional power: the power to cut paper were it to be instantiated in circumstances **K**, where **K** includes Pss being co-instantiated with the property being *larger than a water molecule* (among a host of other properties), its being instantiated in a world in which paper is in fact sliceable, etc.

However, manifest causal powers do not need to be referenced to any circumstances. We can analyze manifest causal power contribution as follows:

**Manifest Causal Power Contribution**: Where p is an instance of a property P, P contributes a manifest causal power M(E) to p just in case p causes (or contributes to causing) an instance of E. M(E) does not need to be referenced to circumstances, for a property contributes a manifest causal power simply when it is involved in an instance of causation, regardless of the circumstances.

**Conditional Causal Power Contribution** helps us to make sense of conditional causal powers as intrinsic, essential features of the properties that contribute them, and thus, by my lights, makes their existence more palatable. Suppose **K** is as a complex state of affairs that includes all of the states of affairs relevant to the causal production of E that do not involve P. **K** will include states of affairs involving all other properties whose presence are relevant to the production of E (in that they also contribute causal powers relevant to E’s production), states of
affairs involving properties whose absence is crucial to the production of E, and, most importantly for the point I am making here, states of affairs about the laws of nature. In doing so, K will mark out a set of possible situations such that if P were to be instantiated in any of these situations, an instance of E would result. But it is easy to see that on this understanding of what K includes, the conditional causal powers that a property contributes will be essential to it. Suppose one of the conditional causal powers that the property of having mass of 1 kg contributes is the causal power to depress scale S a distance of x millimeters when in circumstances K, where K includes, among its contingent states of affairs, the laws of gravitational attraction being exactly what they are. In the actual world, K obtains and an instance of having mass of 1 kg causes scale S to depress x millimeters. In worlds with different laws of nature, having mass of 1 kg still contributes exactly the same conditional causal powers, including the power to cause scale S to depressed x millimeters in K. However, in some other world in which K does not obtain, conditional power C(K,E), though present, will never be “manifest”.

So here we have an understanding of causal power contribution in which the conditional causal powers a property contributes are essential without also implying that the laws of nature are essential. I take this to be a strong virtue of the account. The only possible perceived negative I can foresee with this analysis of Conditional Causal Power Contribution (combined with our understanding of what is to be included in K) is that two properties (say M and N) instantiated at the same time could never contribute exactly the same causal power. For even if the instance of M and the instance of N are overdetermining causes of an instantiation of E, the instance of M and the instance of N occur in different circumstances, and thus contribute different
causal powers—M contributing $C(K, E)$ where $K$ includes an instantiation of $N$, but not of $M$, and $N$ contributing $C(K^*, E)$ where $K^*$ includes an instantiation of $M$. Some may this as a vice of the analysis, but it seems to me it may be another virtue. It allows us to distinguish the causal power contributed by two simultaneous, non-overdetermining causes of a given event—for instance, given **Conditional Causal Power Contribution** the causal powers contributed by the property *being a struck match* are different than the causal power contributed by the property *being oxygen*, even though instances of both properties are partially causally responsible for an instance of *being on fire*. On an analysis of causal power contribution wherein $K$ simply involved every state of affairs relevant to the causal production of *being a fire*, the circumstances in which *being a struck match* and *being oxygen* are instantiated would be the same, and thus both would contribute the same causal power. I take it as a virtue of the account that we get the former result and not the latter. True, given **Conditional Causal Power Contribution** two properties causally overdetermining an effect will have to be analyzed in some way other than their contributing the same causal power, but it’s not clear that the notion of **Conditional Causal Power Contribution** is designed for such an analysis.

So we now have a handle on what it means for a property to contribute the different types of causal powers it does and how we might go about identifying which properties possess which causal powers.\(^3\) Conditional causal powers can be included in our ontology as intrinsic, essential properties of the properties that contribute them

---

\(^3\) Of course figuring out what causes what is mainly an empirical matter. What I mean here is that once we’ve figured out what causes what, we can then know, based on **Causal Power Contribution**, which causal powers to assign to which properties.
(or as constituents of such properties). So understood, it seems fair to say that a property is causally efficacious (or at least potentially causally efficacious) in virtue of the conditional causal power(s) it possesses or contributes. Thus a conditional causal power does not “compete” with its bearer for causal efficacy—strictly speaking causal powers are not causally efficacious, they are merely the features of causally efficacious properties by which such properties are causally efficacious.

However, manifest powers do not have as comfortable a fit into an ontology. If a manifest power is a property at all, it would seem to be a relational property and certainly not one that characterizes the essence of a property, nor even of a property instance. Furthermore, there seems to be nothing for a manifest power to do—the property is causally efficacious, efficacious via the conditional causal power that it contributes, and it is the property instance that actually causes the effect. Manifest powers are thus epiphenomenal. For this reason it seems best to take a deflationary view of manifest powers. To say that a property contributes a manifest causal power is just to say that an instance of that property is causing an effect and nothing more. Given that it is only conditional causal powers and not manifest causal powers that are important to include in one’s ontology, from here on when I use the term ‘causal power’, I shall be referring to a conditional causal power (unless otherwise stated).

But it is not clear that all philosophers agree with this conclusion. As I will argue in Chapter Six, Carl Gillett’s theory of ontological emergence seems to depend on the notion that some of the powers a property contributes are in fact manifest powers. So the issue of the ontological status of manifest powers will resurface later in this dissertation. I will leave any further commentary until then.
Lastly, let me clarify two terms that I will use to describe the causal influence properties have (via their instances). First, let us say that a property instance \( p \) is causally efficacious with respect to some other property instance \( q \) iff \( p \) is at least partially responsible for determining that an instance of causation resulting in \( q \) occurs. Though I will usually speak of property instances being causally efficacious, we can also understand properties as being causally efficacious. We can say that a property \( P \) is causally efficacious \textit{simpliciter} just in case it is possible that \( P \) have an instance that is causally efficacious. But more often we might speak of a property’s causal efficacy in a given set of circumstances \( L \). If that’s how we’re speaking, then we can say that \( P \) is causally efficacious if the instance of \( P \) present in \( L \) is causally efficacious.

Why not just say that a property instance \( p \) is causally efficacious with respect to \( q \) iff \( p \) causes \( q \), or contributes a causal power that, in the right circumstances, results in \( q \)? Though this is certainly a sufficient condition upon causal efficacy, defining causal efficacy in such a fashion will rule out another way that has been proposed for a property to be causally efficacious, a proposal that plays a central part of a theory of OE delineated by Carl Gillett. According to Gillett’s proposal, a property is causally efficacious not only if it contributes a causal power, but also if it synchronically determines (or partially determines) that a \textit{distinct} property contributes a causal power. Again, the details of Gillett’s proposal will be analyzed in Chapter Six.

Finally, allow me to touch upon the concept of causal autonomy. A property \( P \) is causally autonomous iff \( P \) is causally efficacious and \( P \)’s causal efficacy is non-
derivative, i.e. wholly grounded in the nature of P alone, and not derived, inherited, or otherwise obtained from any other property instance. Admittedly the notions of “derived”, “inherited”, and “obtained” are somewhat vague and opaque—perhaps the best I can do is give an example of a property instance upon which it is almost universally agreed lacks causally autonomy. This would be an instance of a functional property. Let ‘F’ denote some functional property. Let us suppose F is defined by its causal role, in this case, say, its causal production of one and only one property E (this is for simplicity sake only—F could be a functional property that is defined by a much more complex causal role, it would not make a difference). In other words, F is just the property of having some property out of the set S of properties P1…Pn that is such that an instance of every member of S causes an instance of E. What then of a particular instance f of F that is instantiated in virtue of an instance p of property P1 (a member of S) being instantiated? It seems clear that whatever causal work f does will be wholly derivative of that of p, i.e. if f causes anything it is so solely because p causes something. We might even say that f’s causal work consists in p’s causal work. So, for example, the causal efficacy of the property being a mousetrap, a functional property, just consists in the causal efficacy of the particular structural property that realizes it, or, more colloquially, whatever a mousetrap does in virtue of being a

33 This is not to say that the instantiation of a causally autonomous property will not be determined by the instantiation of some other property or properties—this will be the case in all but the most robust forms of property or substance dualism which deny even the supervenience of higher-level properties upon fundamental properties. The point is that the causal efficacy of instances of the property does not come in virtue of its being so determined or in virtue of its being constructed from other property instances.
mousetrap it does simply in virtue of having the structure that it does. Thus being a mousetrap is not causally autonomous.

1.4.5 Causal Power Types Versus. Causal Power Tokens

Might we not just say in such a case that the causal power contributed by F is identical to the causal power contributed by P?34 We might, but if we do we should be sure to distinguish between causal power types and causal power tokens. If causal powers are properties of properties, we would expect such a distinction to apply, and expect that a causal power token will just be a particular instance of a causal power type. Now based on our definition of a causal power, we could say that our functional property F is the property of having some property in set S such that each member of set S contributes causal power $C(K,E)$. So P1, which is a member of set S, contributes $C(K,E)$ through its instance p. But F does not contribute this causal power, because F is not instantiated in the same circumstances in which P1 is instantiated. P1 is instantiated in K, which includes f (an instantiation of F), but the circumstances in which F are instantiated do not include f. So F is instantiated in different circumstances—call them $K^*$. Thus if F contributes a causal power in this case, it

34 Jaegwon Kim, among others, has argued for this conclusion, which he formulates as a general principle he calls “The Causal Inheritance Principle”, which states that “if a functional property F is realized on a given occasion in virtue of the instantiation of one of its realizers, P, the causal powers of this instance of F are identical with the causal powers of P.” (2003b, p. 580). But as I show above this principle, if it applies at all, applies only to causal power tokens, and not causal power types. One might even wonder why one should think that functional properties contribute any causal powers at all, i.e., why they are not simply epiphenomenal. I am sympathetic to this worry, but I can’t discuss it here, as it is not really critical to my project. For more on whether functional properties are epiphenomenal, one could consult the ongoing debate between Kim and Block on what is called the “Causal Drainage Problem”, most recently continued in Kim (2003a) and Block (2003).
contributes $C(K^*,E)$, a causal power that is at least of a different type than that contributed by P1.

But the question remains as to whether in this case the instance of $C(K,E)$ contributed by P1 is identical to the instance of $C(K^*,E)$ contributed by F. And the answer to that question is probably parasitic on the answer to the question of whether the instance of P1 is identical to the instance of F. If one thinks $p = f$, one is likely to think that the instances of the causal powers they contribute are identical. I have nothing more to say about this question here, but it will resurface in Chapter Five as we consider Derk Pereboom’s attempted solution to the Problem of the Special Sciences. Pereboom contends there are certain situations where a causal power contributed by a realized property is not even token identical to any causal power of the property realizing it. So there will be ample room to ponder this question more in Chapter Five.

This concludes the introduction to the concept of Ontological Emergence and some of the key metaphysical concepts crucial to understanding Ontological Emergence. In the next chapter, I turn to the task of situating the Theory of Ontological Emergence among various solutions to the Problem of the Special Sciences.
CHAPTER 2

ONTOLUTIONAL EMERGENCE AND THE PROBLEM OF THE SPECIAL SCIENCES

In this chapter I will situate Ontological Emergence along the spectrum of possible solutions to the Problem of the Special Sciences. We already have a rough characterization of OE—comparing OE to alternate solutions to the Problem of the Special Sciences will help us gain further insight into the position by showing exactly what types of theories (and implications of these theories) OE is designed to avoid. Once we see what OE is not, we will have more information with which to construct a positive characterization of the view, one that is relatively modest, building only as much into the theory as is needed to distinguish OE from other candidate solutions to the Problem of the Special Sciences. This modest characterization of OE is what I call Minimal Ontological Emergence, or MOE.

This chapter is divided into three main sections. In the first section, I develop the Problem of the Special Sciences. In the second, I identify various candidate solutions to the problem, showing along the way how they rival Ontological Emergence. Finally, in the last section, I set forth what I think are the minimal necessary conditions on a property’s being ontologically emergent. Having these conditions in hand will allow us to compare MOE to the stronger, yet more traditional
conception of OE that will be discussed in Chapter Three, as well as preparing the way for my presentation of a taxonomy of different theories of OE in Chapter Four.

2.1 Introduction to the Problem

The Problem of the Special Sciences concerns entities whose existence and role in the world is, on the surface of things at least, so obvious that the problem they confront us with is often overlooked. These entities are the referents of what have come to be known as “special science predicates”, predicates such as ‘having an electronegativity of 3.44’, ‘having the gene for Down’s Syndrome’, or ‘believing that the Cubs will eventually win the World Series’. Such predicates are called special science predicates in that they appear in theories put forth by the special sciences, sciences whose domain includes propositions, facts, or theories that do not fall within the domain of the most fundamental science characterizing our universe—microphysics.35 The special sciences would include, but are obviously not limited to, examples such as chemistry, biology, and psychology. The question then is what the different special sciences such as chemistry, biology, and psychology are describing. To what do predicates like ‘having an electronegativity of 3.44’, or

35 The two ideas that are encompassed in this notion of certain theories or facts being more basic than others are, I think, that science X and its theories are more basic than science Y and its theories if (a) the entities that fall under the domain of science Y are more organizationally complex, and, in most cases, composed by the entities which fall under the domain of science X, but not vice versa, and/or (b) the properties, laws, and facts of that are characteristic of science Y are in a robust sense determined by the properties, laws, and facts of science X, but no vice versa. Both of these ideas can be summed up in the notion that reality is broken up into levels, and that different sciences have different levels of reality as their domain (chemistry has the chemical level, biology the biological level, and so on). There have been numerous attempts to make this notion of levels more precise, but for my purposes in this dissertation, it is sufficient to note that is good reason to believe that there are different levels of reality (generated either by the part-whole relation holding between substances or a determination relation holding between property instances that different sciences describe). For more on the metaphysics of “levels of nature” see Carleton (1985), Wimsatt (1995), Emmeche, Koppe, and Stjernfelt (2000), and Kim (2002).
‘believing that the Cubs will eventually win the World Series’ refer? The most obvious answer is that they refer to properties. But what kind of property? Does, for example, ‘having an electronegativity of 3.44’ refer to a single property, or rather to a collection of properties? If to a single property, does it refer to a property that is uniquely chemical, and novel from the standpoint of microphysics, i.e., distinct from property denoted by a predicate of microphysics? Or does it simply refer to a more complicated (perhaps multi-place relational) microphysical property, one that is only exemplified by things achieving a certain degree of complexity (e.g., complex enough to be called a chemical element)? Or are these questions just misguided? Isn’t it just a matter of semantics whether the property in question is a microphysical property or a chemical property?

I think this last question, as deceptively attractive its rhetorical force might be, at root portrays a bias, or perhaps just an unreflective assumption, about the metaphysics of properties. If one holds to a certain kind of nominalism about properties, then one will think that all that it is required for a certain object to be of a certain kind is for a certain predicate to truly apply to that thing. For example, in order for it to be true that oxygen has an electronegativity of 3.44 is simply for the predicate ‘has an electronegativity of 3.44’ to correctly apply to oxygen. In this case of course there is nothing outside of the predicate itself that grounds its application to oxygen. But this really is just to deny that there are properties in the sense I have been assuming, and is just to dissolve the Problem of the Special Sciences. Central to the Problem of the Special Sciences, at least as I am framing it, is the assumption that

\[\text{36 I am thinking here of an idealized microphysics, one that would include all predicates necessary to capture all the truths about the microphysical realm.}\]
there are non-linguistic entities to which predicates refer. My point in all this is that there being any sort of problem in this vicinity depends on a construal of properties according to which properties are an important aspect of the entities that exemplify them—that they are “in” the entities that have them (either in a spatial sense, or in the sense of being predicated of these entities), and it is there being “in” the entities that have them that grounds similarity.37

So I shall assume that if a predicate like ‘having an electronegativity of 3.44’ refers, that it refers to a property in the sense just specified. Of course it is possible that it does not refer at all. Clearly electronegativity is a concept utilized in chemistry (specifically, it is a concept used to quantify the relative ability of atoms of a given element to attract bonding electrons) and when one says of oxygen that it has an electronegativity of 3.44, chemists would admit that one speaks truly. Of course just because an entity falls under a concept does not alone imply that the entity possesses a single property whereby it falls under that concept.38 But let us assume that ‘having an electronegativity of 3.44’ does refer; to what, then? It seems clear that it must refer to

37 Note that assuming what I am here about properties does not amount to assuming what is traditionally called “realism” about properties. Although realism about properties is consistent with this assumption, so is at least one prominent nominalist theory, namely, trope theory.

38 One may recall Wittgenstein’s well-known example of the concept ‘game’. Wittgenstein argues that there is no one feature common to all and only the things that we refer using the concept ‘game’. He argues that all entities that fall under the concept ‘game’ share a “family resemblance” to each other, but no genuine property (Wittgenstein 1953: secs. 66 and 67). D.M. Armstrong offers other examples such as disjunctive concepts like ‘having charge C or mass M’ where it seems questionable that there is a unique property corresponding to the concept (Armstrong 1989, p. 84). The question these examples beget is this: how do we distinguish predicates that pick out genuine properties over against those which truly apply only because, as in the case of ‘game’ or ‘having charge C or mass M’, they refer to some set of properties which can vary on different occasions? And the answer I think that is generally adopted in the literature, and which I concur with, is that a special-science predicate picks out a genuine property only if it picks out something that has causal efficacy, as I discuss above. I leave it open as to whether there exist properties that do not fall within the domain of any science per se, in that they have no causal efficacy (I am personally quite sure that there are, but that is really neither here nor there for the purposes of my discussion).
a property whose instances are both (a) causally efficacious and (b) wholly determined by instantiations of microphysical properties. As to (a), these are special science predicates after all, and science is in the business of categorizing the things that make a causal difference in the world. Clearly, it is because oxygen has an electronegativity of 3.44 (call this property Q) that it attracts an electron pair from a hydrogen atom (resulting in a polar bond). The instance of Q brings it about that the electron pair occupies a certain position, and thus that the water molecule of which the pair is a proper part has a certain configuration, i.e. exemplifies a certain property instance. The common man believes that all manner of special-science properties have causal efficacy. An instance of having the gene for Down’s Syndrome is causally efficacious in that it brings about a person’s having Down Syndrome, an instance of believing the Cubs will eventually win the World Series is causally efficacious in that it causes a certain philosopher to persevere in his allegiance to the team and that philosopher’s body to appear frequently at Wrigley Field, etc.\(^{39}\)

As to (b), are not all entities described by the special sciences constructed out of parts whose ultimate constituents are the entities of microphysics? There is a strong intuition then that, for a given entity, there can be no difference in the special science properties such an entity exemplifies without a difference in the microphysical properties of its constituents (or, as Kim puts its, its total microstructural property).\(^{40}\)

In other words, we have the intuition that the following thesis is true:

\[^{39}\text{Of course all of these property instances act in conjunction with others in producing these effects—their being causally efficacious does not entail their being solely responsible for their effects.}\]

\[^{40}\text{Kim defines a total microstructural property as “a macro-property (macro since it belongs to the system as a whole) constituted by the system’s basic micro-constituents, their intrinsic properties, and the relations that structure them into a system with unity and stability as a substance” (1999, p. 7). I}\]
**Mereological Supervenience.** Systems with an identical total microstructural property have all other properties in common. Equivalently, all properties of a physical system supervene on, or are determined by, its total microstructural property. (1999, p. 7).

But how can mereological supervenience be true and there be special science property instances that are causally efficacious? That, in a nutshell, is the Problem of the Special Sciences. It can be formulated as a set of individually plausible yet jointly inconsistent theses:

(i) There are some special science predicates that refer to non-microphysical, special science properties.

(ii) Every instance of every property referred to by a special science predicate is wholly determined by instances of microphysical properties.

(iii) Instances of some properties referred to by special science predicates are causally efficacious.

(iv) If every instance of every property referred to by a special science predicate is wholly determined by instances of microphysical properties, then no instances of properties referred to by special science predicates are causally efficacious.

By substitution it is clearly seen that (iv) is equivalent to “if (ii), then ~ (iii)”.

Thus, (ii), (iii), and (iv) are jointly inconsistent—(i) is listed as a relevant thesis simply because the whole Problem of the Special Sciences is predicated upon it—one can simply avoid the force of the Problem by denying it. Actually, it would be more precise to note that (ii) entails (i), so denying (i) really just amounts to a special case of denying (ii). The *challenge* is to offer an explanation as to the nature of special science properties and their relation to microphysical properties that not only entails

---

don’t think we should read too much into the “constitution” talk here, as if Kim is endorsing an immanent universals or trope theory. All Kim means to say here is that there exist microstructural properties, and any given microstructural property is simply the property of being composed by certain parts that instantiate certain intrinsic properties and bear certain relations to each other.
the denial of one of these theses, but makes that denial plausible. Let’s now turn to a range of possible solutions to this problem, and see where OE falls into the mix.

### 2.2 Categorizing Responses to the Problem of the Special Sciences

As I see it there are basically five options as to the referent of any given special-science predicate ‘P’. It will be most convenient to demarcate the most common positions on the Problem of the Special Sciences by identifying which of these options each position allows. From there we can see which of (i), (ii), (iii) or (iv) each position entails to be false.

(A) ‘P’ fails to pick out any property at all, and therefore fails to refer.

(B) ‘P’ does refer to some property P, but P is simply a microphysical property or the right type of combination of microphysical properties (perhaps some Boolean combination of microphysical properties which itself qualifies as a microphysical property).

(C) ‘P’ refers to a non-microphysical, special-science property P, instances of which are “constructed out of” (but not via Boolean combination, rather, as a “structural property” which will be described below) or “realized in” instances of microphysical properties, and P is causally efficacious but not causally autonomous.

(D) ‘P’ refers to a non-microphysical, special-science property P that is not causally efficacious, but P is still autonomous in that P figures in laws and causal explanations that are distinct from laws and causal explanations governing microphysical properties.

(E) ‘P’ refers to a non-microphysical, special-science property P that is causally efficacious and causally autonomous.

Of course it is possible that all of these options are satisfied by different special-science predicates (i.e. there may be some special-science predicates that do not refer, some that refer only to microphysical properties, etc.). But based on these options, I want to identify what I take to be the five most common solutions to the Problem of the Special Sciences.
Eliminativism about special-science properties entails that (A) is true of all special-science predicates, that all of them fail to refer. The eliminativist thus “solves” the Problem of the Special Sciences by denying that there is a problem, i.e. by denying (i)—and thereby denying (ii). What I shall call “unqualified micro-determinism” entails that either (A) or (B) is true for every special-science predicate. I’ll use this name to refer to the view because it clearly entails that the only causally efficacious properties are microphysical ones. This solution also entails the rejection of (i), as I shall explain presently.\(^4^1\)

Qualified micro-determinism entails that either (A), (B), or (C) is true of every special-science predicate. It is a qualified micro-determinism in that it allows for causally efficacious special science properties, but not ones that have any causal autonomy—all special-science properties will derive their causal powers or causal efficacy entirely from microphysical properties. Examples of such properties would include *having mass of 10 kg* or *being a mousetrap*. If such properties are causally efficacious they are clearly so only as a result of the causal efficacy they derive from the properties that form their microphysical base. This position entails the denial of (iv). Breaking up the pattern somewhat, epiphenomenalism entails that either (A), (B), or (D) is true of every special-science predicate. Epiphenomenalism entails the denial of (iii). Finally, we have what I will call Serious Non-Reduction (SNR), which

\(^{4^1}\) Of course, one could be an eliminativist in a more qualified sense—an eliminativist about only the referents of psychological predicates, for example. But here I am taking eliminativist to represent a more broad sweeping position

\(^{4^2}\) The truth of unqualified micro-determinism would not imply that the causal laws governing the microphysical are completely deterministic. I am not using the word ‘determinism’ in that sense here. It would imply that, insofar as events in the world are causally determined, they are completely determined by instances of microphysical properties.
entails that there is at least one predicate for which (E) is true, allowing of course that there may be some predicates for which (A)-(D) are true. OE, which also entails the denial of (iv), is an example of SNR. Where it differs from qualified micro-determinism, as we shall see, is in its explanation as to why (iv) is false, an explanation that references the causal autonomy of higher-level special science properties. However, SNR is not exhausted by OE. This is due to the fact that (E) is also a thesis compatible with certain varieties of dualism or vitalism, views that entail that there are some nonphysical substances—Cartesian souls, entelechies, etc. from which higher-level, nonphysical properties and/or forces emanate that are capable of effecting the physical world and “disrupting” the natural course of purely physical processes, yet which fail to be dependent upon the physical world.43 Jaegwon Kim refers to these views as “serious dualisms.” Since they involve distinct substances we can call them serious substantival dualisms. More on the distinction between OE and serious substantival dualism will come later in the chapter.

One may note that nowhere in this brief taxonomy of solutions to the Problem of the Special Sciences do I include a view bearing the name ‘Non-Reductive Physicalism’ (hereafter NRP). One may find this odd, given that such a view would normally be considered one of the main rivals of Ontological Emergence. And this is

43 Broad’s objections to theories of this sort, what he called “Substantial Vitalism”, are the most clearly articulated. He suggests we should not believe in such entities in that (i) they cannot be isolated, thus are purely hypothetical entities and not empirically testable, (ii) there composition is not known, nor is there anyway to conceive of a way to test their composition, and (iii) it is hard to see how such things could exist in space, and thus it is hard to see how they could be part of a substance that effects the world (Broad 1925, pp. 57-58). Emergent features, however, were not at all like this to Broad. In fact, Broad thought that in addition to having good indirect evidence for their existence (that we could see their effects in the world, effects not attributable to the effects of fundamental properties) we had direct awareness of some emergent properties, namely, the “secondary qualities” of our senses. In this, he agreed with Mill. Compare Broad (1925, pp. 77) with Mill (1843, p. 346, Bk. 3, Ch. 14, §2).
partially true—there are versions of NRP that do rival Ontological Emergence, but there are also versions of NRP that would qualify as instances of OE. So they are not mutually exclusive categories. The main problem is that NRP encompasses many different theories and it is difficult to neatly classify them all into one group. In fact, theories that have been labeled NRP could, in my taxonomy, fall into three different categories—qualified micro-determinism, epiphenomenalism, or Ontological Emergence. In what follows, I will try to point out which of the more “standard” versions of NRP would fall in each of these categories.

2.2.1 Eliminativism and Unqualified Micro-determinism

Eliminativism simply dissolves the problem, and there is no question about the metaphysics of special science properties. Unqualified micro-determinism also dissolves the problem—thought it’s not as immediately clear why. Allow me to elaborate. Frank Jackson refers to those who hold this position as “transparency theorists” suggesting that when one uses a special science predicate one believes (or perhaps even pretends, depending upon one’s background beliefs) that one is referring to a unique special science property that fails to be identical to any microphysical property, but in actuality one is only “referring through” such a property to the “real” properties at the microphysical level. The not-so-subtle implication on Jackson’s part is that on such a view there really are no special science properties that fail to be identical to microphysical properties (Jackson 1996, p. 378). The only story to be told is the one at the fundamental level by microphysics, and any other story (the chemical story, the biological story, the mental story, etc.) is just a retelling of the physical story in a different language. According to this view, there is no uniquely chemical property
picked out by ‘having an electronegativity of 3.44’. The predicate, if it refers at all, refers only to the microphysical properties exemplified by the most ultimate proper parts of the oxygen atom. There is no causal efficacy or autonomy for the special-sciences, for there really are no special-sciences to be had. This is a truly extreme reductionist view—so extreme in fact that I am fairly confident that there is no one who actually holds the view. This is with good reason. We might be able to accept that ‘having electronegativity of 3.44’ refers to a microphysical property, but it would be a huge stretch to accept this about the predicate ‘believing the Cubs will eventually win the World Series’, for that would be tantamount to accepting the proposition that our minds, the things exemplifying mental properties, are microphysical entities (albeit one’s arranged in an interesting way). While we might accept that a bunch of microphysical entities could work together to attract a pair of electrons without thereby composing one thing that attracts the electrons (i.e., an oxygen atom), it seems beyond the pale to accept that a bunch of microphysical entities could work together to

---

44 If there even is such a thing as an oxygen atom. I take it that unqualified micro-reductionism is most attractive to those who are nihilists about composition (or, more specifically, about what Peter van Inwagen (1990) the “Special Composition Question”), and thus those who would deny the existence of oxygen atoms. Nihilists believe that there are only physical simples. Van Inwagen himself approaches being a nihilist—he thinks the only composite objects that exist are living beings, he thus rejects the existence of almost all composite objects that one would normally recognize. Quantify over everything but living objects and van Inwagen is an unqualified micro-determinist (you might say then that he is a qualified, unqualified micro-reductionist). What a predicated like ‘having electronegativity of 3.44’ refers to is probably a very complicated relational property holding between a host of simples.

45 There may be grounds for thinking that Jaegwon Kim holds this view. In Kim (1998), Kim seems to conclude that there really are no “special-sciences properties”, but merely “special-sciences concepts,” concepts that ultimately refer only to microphysical properties. As suggested in the previous note, though, there are some philosophers who appear to hold this view in a more relaxed sense to apply only to special science properties not predicatable of living or conscious beings. See, for example, van Inwagen (1990) and Merricks (2001).
think, and reason, and feel, without thereby composing one thing that thinks, reasons, and feels.\textsuperscript{46}

2.2.2 \textit{Qualified Micro-Determinism}

If ontological reduction required accepting unqualified micro-determinism then probably no one (or at least no one who fully saw the consequence of the view) would be an ontological reductionist. But there is another option that has been the preferred choice for self-proclaimed ontological reductionists, and this is qualified micro-determinism.\textsuperscript{47} As I stated earlier, qualified micro-determinism entails that there are causally efficacious special science properties, but that all special science properties derive their causal powers from microphysical properties, or the causal efficacy of special science properties is wholly derivative upon the efficacy of microphysical properties. I noted earlier that clear examples of properties whose causally efficacy is so inherited or determined included having mass of 10 kg or being a mousetrap. The former, it could be argued, is a type of structural property, whereas the latter is a realized property. Qualified micro-determinism entails that all special science properties are either structural or realized. How should we understand these types of properties? Well, most who champion the existence of a structural, or as it is sometimes called a micro-based, property seem to accept something like D.M. Armstrong’s formulation. According to Armstrong, to instantiate a structural property

\textsuperscript{46} This echoes van Inwagen’s sentiment and is at least one of his reasons for concluding that nihilism is false, and that there is at least one class of composite objects (thinking beings). See van Inwagen (1990, pp. 118-9).

\textsuperscript{47} Besides self-identified realization physicalists like Poland (1994) and Melnyk (2003), one can add those who have defended type-type identity theory, like Brian McLaughlin (2001), who are at root realization physicalists.
F is to be a particular made up of non-overlapping proper parts related in certain ways, each of which instantiates certain universals.\(^48\) (Armstrong 1997, p. 32) For example, if F is a structural property had by an object x which is composed of just two proper parts \(x_1\) and \(x_2\) instantiating intrinsic properties \(P_1\) and \(P_2\) respectively, then instantiating F is just having nonoverlapping parts \(x_1\) and \(x_2\) such that \(P_1x_1\) and \(P_2x_2\) and \(x_1Rx_2\) (where \(R\) is some relation).\(^49\) Those who advocate an ontology that includes structural properties often suggest that structural properties are causally efficacious in that their instantiations can bring about certain other events, but not causally autonomous because the causal contribution they make is nothing beyond the contribution of the properties “composing” the structural property. Suppose, for instance, that F is a structural property instantiated by a baseball and \(P_1\ldots P_n\) are the intrinsic properties of the microphysical parts of the baseball that are “constitutive” of this structural property. One might think that F is causally efficacious, say, in the production of a shattered window, but if it is, then it would also seem that \(P_1\ldots P_n\), for instances of \(P_1\ldots P_n\) are equally responsible for the shattering of the window. The causal powers of F thus are derivative of those of \(P_1\ldots P_n\).

\(^{48}\) On Armstrong’s ontology, an instance of a structural property is a conjunction of the states of affairs involving the parts and their properties—in this particular case, the instance of F, i.e., x-having-F, is the conjunction of \(x_1\)-having-\(P_1\), \(x_2\)-having-\(P_2\), and \(R(x_1, x_2)\). These property instances are, as Armstrong puts it, a “part of” the structural property instance, and the structural property instances are what Armstrong refers to as an “ontological free lunch”. It is clear then that Armstrong would be of the opinion that structural properties are not emergent.

\(^{49}\) Armstrong (1997, pp. 12-13 and p. 35). The way Armstrong actually puts it is that the state of affairs \(x_1\)-having-\(F\)1 that is part the state of affairs of x-having-F. But it seems fair enough to stay on Armstrong’s ontology that states of affairs are property instances. He speaks of universals as “state-of-affairs types” (1997, p. 28), and this suggests that states of affairs are tokens of universals. One might wonder why structural properties do not qualify as microphysical properties, and thus why what I am calling a version of qualified micro-reductionism does not just reduce to unqualified micro-reductionism. But structural properties are not microphysical properties, because they are not exemplified by microphysical entities; rather, they are exemplified by composite objects composed of the entities of microphysics.
However, it is not a structural property but a realized property that is the type of property entailed to exist by the most common expression of qualified micro-determinism, realization physicalism. A recent thorough treatment of the view has been put forth by Andrew Melnyk (2003). Melnyk sets up the view as a generalization of psychofunctionalism, one of the reigning theories of mind over the last forty years. Melnyk explains that realization physicalism generalizes psychofunctionalism in several respects; the most relevant to our discussion are as follows:

First, it claims that not merely mental properties but all (instantiated) properties not mentioned as such in the theories of fundamental physics – presumably including therefore all chemical, biological, sociological, and folk-physical properties – are to be identified a posteriori with functional properties. Second, it claims that all these properties are physically realized in the sense of being realized by properties that are mentioned as such in the theories of fundamental physics. (2003, p. 8)

Realization/functionalization has been defined in a variety of ways, but those who advocate realization physicalism all assume that a necessary condition is that the realizer(s) play(s) the causal role of the realized property, i.e. that whenever an instance of a realized property causes an event e, it does so in virtue of the instance(s) of its realizer(s) causing e; the instance of the realized property “inherits its causal powers” from its realizer(s). Melnyk’s preferred example of a realized property is that of being a can opener. It can be realized by myriad structural properties (which are in turn realized or constructed out of microphysical properties), and any causal power it has, such as the power to open a can, is, though not identical to any causal power of the properties of the parts of the can opener, still nothing over and beyond these causal
powers. So a realized property can be said to be causally efficacious, but not causally autonomous.50

In any case, I hope it is now at least clear why qualified micro-determinism is to be distinguished from the unqualified variety. While the latter implies that all causally efficacious properties are fundamental, microphysical properties, the former implies there are some non-microphysical properties that are causally efficacious. But

50 But does not their being “constructed out of” or “realized by” microphysical properties preclude there causal efficacy? Those who advocate realization physicalism obviously answer this question in the negative, and many have offered complicated defenses of the notion—see for example Shoemaker (2001) and Melnyk (2003). Sydney Shoemaker defends the causal efficacy of realized properties (using as an example the property of being in pain) as follows:

Suppose physical property P1 is one of the realizers of the property of being in pain, and that in a particular case the causing of a piece of “pain behavior”, say taking an aspirin, involved the exercise of some of the conditional powers conferred by P1…Which of these property instantiations, that of being in pain or that of P1, should we say was a cause or partial cause of the piece of behavior? It seems natural to say that it is the instantiation of the property of being in pain. Of course, the person was in pain by having P1. But…it is qua realization of pain that P1 made its contribution to causing the behavior; for the conditional powers it conferred that are independent of those conferred by the property of being in pain were irrelevant to its making this contribution. (Shoemaker 2001, p. 79).

I think the initial reaction one should have to Shoemaker’s claim that it is qua realization of pain that P1 makes its causal contribution is that this seems to get things backwards. Surely it is qua being realized by P1 that this instance of pain makes its causal contribution—after all, its causal powers are just a subset of this instance of P1’s causal powers. And even if one were to say that the instance of pain is a genuine cause, then one would be forced into either denying the causal closure of the physical level (according to which if a physical event has a cause at time t, then it has a sufficient physical cause at time t—one would reject this by rejecting P1 as a cause of the pain behavior), or embracing causal overdetermination, with the instance of P1 and the instance of pain as overdetermining causes. This would be an instance of what Eric Funkhouser (2002) calls “incorporating causal overdetermination”, according to which “the alleged overdetermining causes ‘work through’ the same mechanism. The causally relevant properties of the incorporating causes are different, but they are predicated of the same mechanism.” This is my point exactly—there is only one “mechanism” responsible for the causation, the “mechanism” that works itself out at the fundamental level, and thus it is difficult to see how realized properties could be viewed as causally autonomous.

So ultimately I think this defense of the causal efficacy of realized properties is questionable. However, it seems to me the best reason for believing in the causal efficacy of structural or realized properties is also perhaps the easiest to understand. It certainly seems possible at least that there is no “bottom-level” to reality, that every object that exists has proper parts and thus that every property that is instantiated is either a structural or realized property. Were this possibility to be actual, then instances of some structural or realized properties would have to be causally efficacious (modulo the acceptance of the thesis that at least some property instances are causally efficacious). This, according to David Lewis, no proponent of Armstrong’s structural universals, was the weightiest reason to believe in them. See Lewis (1986a)
it is still a micro-determinist view in that all effects are ultimately determined solely by the determinative influence or causal powers bestowed by the fundamental, microphysical properties. The reason I have devoted so much exposition to these micro-determinist positions is that they are clearly the type of reductionisms that the British Emergentists and contemporary emergentists alike have sought to avoid. As Brian McLaughlin has persuasively argued, there was a singular line of thought descending from Mill all the way down to Broad (and most evident in Broad) that the British Emergentists believed in “configurational forces, fundamental forces that can be exerted only by certain types of configurations of properties” (McLaughlin 1992, p. 52). And it is easy to infer from his writings that Broad believed the existence of such forces was attributable to the existence of properties, those referred to by at least some of terms of the special sciences, that endow their bearers with certain causal powers “to influence motion in ways unanticipated by laws governing less complex kinds and conditions concerning the arrangements of particles” (1992, p.51).\footnote{\textsuperscript{51} Broad had in mind primarily chemical, biological, and psychological terms as those that referred to irreducible properties responsible for fundamental configurational forces. McLaughlin notes later in his article that given the science of the day, it was eminently reasonable to suppose that chemical and biological properties were such fundamental force-generating (or causal power bestowing) properties. Things changed with the advent of quantum mechanics, statistical thermodynamics, and the discovery of DNA—it now seems, notes McLaughlin, that we can explain the workings of chemical and biological systems without reference to novel causal powers. The most likely place for the presence of genuinely emergent features and forces is in the mind, specifically with qualitative aspects of experience. McLaughlin discusses this possibility more seriously in McLaughlin (1997).} And thus emergent features are the kinds of causally \textit{autonomous} features that would exist were micro-determinism of any variety to be false.
2.2.3 Non-Reductive Physicalism and Epiphenomenalism

But while the British Emergentists may have been aware of the threat of these causal reductionisms in formulating their views, there is a more recent arrival on the scene that in all likelihood they had not considered. This is contemporary non-reductive physicalism (NRP). As I mentioned previously, theories of NRP are not easily grouped into a single category. One type of theory that has been proclaimed by its adherents as non-reductive is the token-token identity theory. According to token physicalism, special-science properties are to be distinguished from, i.e. are not identical to, microphysical properties, nor any properties reducible to the microphysical (i.e. structural properties or realized properties) but tokens (instances) of special-science properties are identical to token of properties reducible to the microphysical. What of the causal contribution of a special-science property Q under the assumption of token physicalism? By my lights, token physicalism is an example of qualified micro-determinism. Even if Q is causally efficacious by having an instance that is identical to an instance of some causally efficacious microphysical property P, Q is not causally autonomous, for clearly Q’s causal efficacy is derivative on the causal efficacy of P.

Other versions of NRP are more difficult to classify. However, what all proponents of NRP agree upon is that there are separate stories to be told about the

---

As I’m thinking of standard NRP, it entails the denial of ontological reduction but the acceptance of epistemological, or explanatory, reduction. For a good description of the differences and why someone might hold this view, see Antony (1999). But there are many other types of NRP, and they are not easily categorized. The difficult think about grasping NRP is how wide the umbrella is—so many different people refer to there views as examples of NRP. For instance, whereas Antony claims that NRP entails the denial of ontological reduction and the acceptance of explanatory reduction, Robert Van Gulick asserts that NRP entails just the opposite—the acceptance of ontological reduction and the denial of explanatory reduction. (Van Gulick 1992 in the Emergence and Reduction volume)
world (in contrast to the “transparency” theory in which there is only one story to be
told), stories that “cross-cut” each other to some degree. As Frank Jackson
summarizes NRP with respect to its understanding of psychological predicates, the
“divisions being effected by psychological language cut across those being effected by
physical language in a way that cannot be captured in physical terms” (1996, p. 378).
So there are psychological predicates that play roles in lawlike statements that can’t be
captured in any reasonable way in physical terms. This is enough to suppose, or so
argues the proponent of NRP, that there are mental properties corresponding to these
predicates, in that the referents of such terms are projectible kinds participating in
interesting regularities. And this is sufficient for the rejection of ontological
reduction. As Louise Antony puts it, disagreeing with Jaegwon Kim’s claim that the
functionalization of a higher-level property P is sufficient to reduce P (in that it shows
that all of P’s causal powers are derivative):

…a thing is ontologically irreducible if it participates essentially in regularities
that are novel from the point of view of the alleged reducing base. Thus, I
contend, the reality of a property – or, as we non-reductive materialists like to
say – its autonomy – is not a matter of whether or not its patterns of causal
relations can be predicted or explained from below, but rather a matter of
whether it figures essentially in non-analytically-guaranteed regularities…As
long as such generalizations are discoverable, we have empirical evidence that
it is the higher-level property that we are projecting. (1999, p. 42)

However, even though the non-reductive materialist like Antony acknowledges the
existence (and in some sense the indispensability) of higher-level properties, it is
unclear whether she acknowledges their causal efficacy. For instance, Antony writes
with co-author Joseph Levine that:

there is reason to think that the “real causal work is being done at a lower
level.”… What we mean by calling mental states “causally explanatory” is
independent of the nature of the mechanisms by which their casual work is
accomplished. (1997, p. 97)
Such a sentiment echoes the work of one of the most ardent and capable defenders of NRP, Terence Horgan. According to Horgan, the only fundamental laws are laws within physics itself. There are no brute, fundamental psychophysical laws, and thus there are no fundamental “force-generating” or causal power bestowing properties above the fundamental microphysical level. More than anything according to Horgan, this is what distinguishes NRP from ontological emergence. Horgan even refers to his position as one that understands higher-level properties to be merely “resultant” and not “emergent”, meaning that their causal capacities are accountable for in terms of and derivative from the causal capacities of the fundamental properties that realize them. (2002, p. 308)

All of this would seem to suggest that Antony and Levine as well as Horgan conceive of special-science (mental) properties as epiphenomenal, but of course they would beg to differ. Still, it is hard to make sense of special-science properties being causally efficacious if “all the causal work is done at the lower-level.” Horgan’s strategy for maintaining the causal relevance of the mental is to adopt a “compatibilist” account of mental causation, according to which there are multiple levels of genuine causal/explanatory efficacy, levels that are not in direct competition. He refers to this as the CP—counterfactual pattern—conception of explanatory relevance. He writes:

53 See for example Horgan (1993a, p. 560).

54 For an excellent criticism of Antony and Levine’s position, as well as distinct, penetrating criticisms of some other versions of NRP (and their compatibility with special-science property causal efficacy), see Jaworski (2002). I am convinced by Jaworski’s arguments that Antony, Levine, and others who advocate their form of NRP are really epiphenomenalists.
The upshot so far is this. Under the CP conception of explanatory relevance, there can perfectly well be multiple levels of causal explanation for a single phenomenon. The phenomenon and its cause can respectively instantiate multiple pairs of properties; and several instantiated pairs at different theoretical levels, can perfectly well each belong to some pattern of counterfactual relations among properties that makes for explanatory relevance. Furthermore, prima facie there are strong grounds for thinking that intentional mental properties figure in CR (counterfactually relevant) patterns of the requisite kind, and thus that explanations at the level of intentional psychology are compatible with explanations at other more theoretical levels such as neurobiology. (1993b, p. 300).

However, he goes on to say that:

For any causal transaction where some higher-level property F is cited as causally explaining the effect, there must be an underlying mechanism in virtue of which the transaction occurs—a mechanism involving a physical property…which, on the given occasion, physically realizes the property F. (Horgan 1993b, pp. 302-3).

Horgan’s view is thus akin to the picture represented in Figure 1, where M1 and M2 are instances of special science properties, and P1 and P2 are instances of microphysical properties:

```
M1 _ _ causally explains _ _ M2
_                   _
_realizes            _ realizes
_                   _
P1 _ causally explains (and causes) _ P2
```

**Figure 1**

But Figure 1 is not a picture consistent with Ontological Emergence, for M1 is not causally efficacious. Consider what Horgan’s picture implies about an instance of the
property referred to by ‘having electronegativity of 3.44’. The pre-philosophical (but post freshman chemistry) intuition that we have is that this property instance is at least partially causally efficacious in bringing about the existence of a water molecule with a certain bonding structure. It is thus causally efficacious with respect to subsequent microphysical properties. But this is not so given the epiphenomenalist picture sketched above—M1’s sole role in this picture is to serve as the causal explanans for the occurrence of M2 (some other chemical property that will be realized by the structure of the water molecule)—but M1 does no causal work and is not causally efficacious with respect to P2. The causal mechanism is completely determined by the properties at the microphysical level, leaving M1 with no role to play in this determination.\footnote{This line of reasoning is reminiscent of the well-known “supervenience argument” against the causal efficacy of higher-level supervenient properties offered in numerous places by Jaegwon Kim. In the first step of this argument Kim contends that “same-level” causation, in this case causation of M2 by M1, requires “downward causation”, in this case, causation of P2 by M1 (assume that now we are talking about property instances and not properties). Suppose that M1 does cause M2. Given that M2 is realized by P2 (and thus is at least nomologically necessitated by P2), M1 must cause M2 by causing its realizer to be instantiated, i.e. by causing P2. But given that P1 is a sufficient cause of P2 (an assumption non-reductive physicalists like Horgan can hardly deny, given it is constitutive of their view that the causal mechanism plays out entirely at the physical level), then either M1 is causally inert, or it is an overdetermining cause of P2. Now Horgan all but grants this conclusion, reverting to his “causal compatibilism” response to preserve the autonomy of properties like M1. But unfortunately this response does not save their causal efficacy. (Horgan actually at one point suggests that the conclusion of Kim’s causal exclusion argument looks good, and that he can offer no easy refutation of it, but that the error involves confusing the concept of physical causation with that of mental causation, as if there are two different concepts of causation in operation (Horgan 1997, p. 180). I think this probably in fact is Horgan’s view, but is one I cannot bring myself to take seriously. It is a view that grows out of a Humean view of causation that I think is inadequate).}

But maybe in the final analysis we can accept this conclusion about the referent of ‘having electronegativity of 3.44’. Maybe instances of this property do not affect the microphysical world via any causal mechanism, and maybe the only reason for believing in the existence of such a property is as a placeholder in lawlike...
regularities that can not be captured by employing the predicates of microphysics alone. Would we really be giving up much if we accepted this? Probably not. But it is a very different matter when we are wondering about the causal efficacy of the referents of the predicates ‘believing that the Cubs will eventually win the World Series’, or ‘having a sharp pain in one’s back’. Giving up on the referents of these predicates being causally efficacious with respect to microphysical events (or at least some lower-level events) would be a huge concession—and one that proponents of OE are not willing to make at all. In fact, I think it is the fact that we must preserve the causal efficacy of mental properties with respect to lower-level events that has been the primary driver of theories of emergence, both in the past and in the present. The notion of causally autonomous emergent properties is also rooted in this idea—once you combine it with the belief that the causal efficacy of special-science properties cannot be had simply by the reduction of them or their causal powers to entities of some more fundamental level.

2.2.4 Ontological Emergence vs. Serious Dualism

The ontological emergentist, then, is committed to (E) and to the proposition that some special science properties are causally autonomous, i.e., causally efficacious by making a causal contribution that is not wholly determined by the powers or properties of the microphysical level. Suppose a property like being in pain were ontologically emergent. This would imply that an instance of the property being in pain makes some causal contribution to the world that is not made by any microphysical property nor is wholly determined by instantiations of microphysical properties, rather, the causal contribution that being in pain made is wholly grounded
in the essence of the property itself. So it is the acceptance of (E) that distinguishes OE from epiphenomenalism and micro-determinisms of all varieties. But it does not distinguish OE from all its rivals. A proponent of serious substantival dualism could also, and most likely would also, agree that *being in pain* is causally efficacious and autonomous in the way just described. Where the proponent of OE parts company with the proponent of serious substantival dualism is in that whereas the former will accept that the occurrence of every instance of the property *being in pain* is completely determined by instances of microphysical properties (and thus “emerges from” microphysical properties), the latter will deny this. To the serious substantival dualist, the occurrence of mental property instances is independent of the occurrence of microphysical properties (perhaps because something like occasionalism or theory of Leibniz’ pre-established harmony is true).\(^5\) So beyond non-reduction, causal efficacy, and causal autonomy, Ontological Emergence must include the idea of determination by the microphysical.

### 2.3 Minimal Ontological Emergence

The purpose of the last several paragraphs has been to situate Ontological Emergence along the spectrum of responses to the Problem of the Special Sciences, specifically to distinguish Ontological Emergence from micro-determinism,

---

\(^5\) I am not suggesting that all forms of dualism are incompatible with emergence. In fact, Thomas Aquinas dualism might possibly be characterized as an emergent dualism. Brian Leftow has defended such an interpretation in Leftow (2001), and William Hasker (1999) has defended a dualistic position consistent with OE which he calls “emergent dualism”. All I am pointing out here is that there are non-reductive views entailing that the irreducible entities are completely dependent from the physical, i.e. do not owe their existence to the physical in any sense. These are the dualist views I have mentioned.
epiphenomenalism, and serious dualisms. The work of this chapter suggests the following minimal necessary conditions upon an ontologically emergent property:

**Non-Reduction:** For all M such that M is an ontologically emergent property, there is no microphysical property P to which M is reducible.

**Microphysical Determination:** For all m such that m is an instance of an ontologically emergent property, the occurrence of m is ultimately determined by an instance or instances of microphysical properties.

**Causal Efficacy:** For all M such that M is an ontologically emergent property, there is at least one instance of M that is causally efficacious.

**Causal Autonomy:** For all M such that M is an ontologically emergent property, M is causally autonomous.

I will refer to any theory of OE that incorporates these conditions upon a property’s being ontologically emergent as an instance of **Minimal Ontological Emergence**, or **MOE**. As we will see in the next chapter (and the chapter that follows), the more or less “received” idea of OE incorporates conditions beyond the above, and is thus stronger than MOE. We will also see that many of the criticisms that have been leveled against OE have been directed at this stronger conception of an ontologically emergent property.

Before moving on to the next chapter, a couple of comments are in order with respect to the conditions that comprise **MOE**. First, **Non-Reduction** obviously entails that no ontologically emergent property is identical to a microphysical property. Whether it entails anything beyond that I will take no position on—so the reader can simply read **Non-Reduction** to be a statement of non-identity. Second, to say that an occurrence of a property instance m is *ultimately determined* by a set of microphysical property instances p1…pn is to say that there is a chain of determination relations linking m to p1…pn, although the determination may not be “direct”—there may be intermediates in the chain, e.g., other emergent property instances. **Microphysical**
**Determination** includes this qualification to allow the possibility of emergent properties whose instances are most directly determined by other emergent properties instances, and thus to allow different “levels” of emergence, so to speak (imagine, for instance, that *having a sharp neck pain* is an emergent property whose instantiation is fully determined by instances of microphysical properties, but that an instance of *having a sharp neck pain* fully determines by itself and in an unmediated fashion the instantiation of *feeling anxious*. We would not want the instance of *feeling anxious* to fail to qualify as being emergent simply because it is not directly determined by microphysical properties). The reader may be wondering what the exact nature of this “determination relation” is, and she is justified in so wondering. This question will be explored in later chapters. Different understandings of what the **Microphysical Determination** amounts to result in different species of MOE.

In the next chapter I turn to an exposition of the most common analysis of an ontologically emergent property at play in the literature—that which has descended from the British Emergentists and has been exposited most eloquently by Jaegwon Kim. I call this **Classic Ontological Emergence**.
CHAPTER 3

CLASSIC ONTOLOGICAL EMERGENCE AND KIM’S CHALLENGE

3.1 Introduction

The purpose of this chapter is threefold. First, I want to clarify the concept of ontological emergence as it was set forth by the British Emergentists and explain how this conception has informed the work of the most prominent and thorough contemporary expositor of the view, Jaegwon Kim. Though there was some variation in the doctrines laid out by Alexander, Broad, Morgan, and others under the rubric of emergence, there are certain core ideas they all upheld. The concept of an ontologically emergent property that can be gleaned from these ideas I will call Classic Ontological Emergence, or COE. Second, after laying out the components of COE I will then turn my attention to Kim’s characterization of an ontologically emergent property. Kim’s analysis serves as a nice departure point into the contemporary literature on ontological emergence not only because it is an excellent example of analytic metaphysics, but also because it serves as an elaboration of nearly all of the necessary conditions upon an ontologically emergent property included in COE.

57 Of course the British Emergentists did not agree on every detail. For some valuable insight into the differences in their views, see O’Connor and Wong (2002).
Though I think Kim incorporates necessary conditions into his analysis of an ontologically emergent property that go beyond what is necessary, and thus disqualify without argument certain properties from possibly qualifying as ontologically emergent, we can still learn much from Kim’s perspective on the issue, especially from two arguments against the existence of ontologically emergent properties that are suggested by Kim’s work (one that is explicitly stated and the other that is merely implied). In the last section of this chapter I will expound upon both of these arguments. Though Kim’s arguments are formulated under the assumption that K-COE is the proper analysis of an ontologically emergent property, and thus unsuccessful in undermining the existence of emergent properties tout court, the arguments, especially the second, which I will be calling Kim’s Challenge, still reveal much about the debate in that they reveal how possible components of a concept of an ontologically emergent property are inconsistent with each other or other highly plausible theses. Kim’s Challenge thus provides a guideline for those trying to formulate conceptions of OE distinct from COE, or K-COE by showing which combination of theses must be avoided.

3.2 A Quick Tour of British Emergentism

Since most contemporary analyses of an emergent property build upon the work of the British Emergentists, it is important to take a brief look at the work of this group to gain a sense of what they intended in speaking of emergence. Obviously, the group as a whole is committed to their being properties that are not reducible to the properties of microphysics. But what did non-reduction mean to them? In formulating the views of the British Emergentists with respect to reduction, things are
complicated by the fact that these authors hardly, if ever, utilized the term ‘reduction’ in their writings. They were more likely to speak of emergent properties as “new qualities” or “new relations” existing at “higher levels” and not belonging to the “lower levels”. For instance, Samuel Alexander claimed that:

The emergence of a new quality from any level of existence means that at that level there comes into being a certain constellation or collocation of the motions belonging to that level, and this collocation possesses a new quality distinctive of the higher-complex. (1920, p. 45)

And Lloyd Morgan wrote that,

Let there be three successive levels of natural events, A, B, and C. Let there be in B a kind of relation which is not present in A; and in C a kind of relation that is not present in B or in A. If then one lived and gained experience on the B-level, one could not predict the emergent characters of the C-level, because the relations, of which they are the expression, are not yet in being. (1923, pp. 5-6)

The notable exception among the British Emergentists in this regard was C.D. Broad, who did specifically address the nature of reduction. In his epic work *The Mind and Its Place in Nature*, Broad distinguishes three types of properties that could be attributed to higher-level, or what he calls high-order, objects, (i) ultimate characteristics of the order, (ii) reducible characteristics of the order, and (iii) ordinarily neutral properties of the. (1925, p. 78) Properties of type (i) are the emergent properties, those that are such that “aggregates of the order possess them, that no aggregate of the lower order does so, and that they cannot be deduced from the structure of the aggregate and the properties of its constituents by any law of composition which has manifested itself in lower orders.” (Broad 1925, p. 78)

Properties of type (ii), the reducible characteristics, are ones that are also such that no aggregates of a lower order possess them, but are such that they can be deduced from the structure of the aggregate, the properties of its constituents, and laws of
composition. And properties of type (iii) are simply properties that could also be possessed by objects of lower orders.\textsuperscript{58}

We see here that Broad is cashing out reduction in terms of an epistemic notion, deducibility. It was part and parcel of the British Emergentist’s worldview that facts about the instantiations of emergent properties were not deducible from the facts about the underlying base properties.\textsuperscript{59} He went on to add that emergent properties and their base properties were linked by trans-ordinal laws (to be discussed below), laws just as genuine and “scientific” as laws describing the relationships of microphysical properties, but such that:

we must wait till we meet with an actual instance of an object of the higher order before we can discover such a law; and that we cannot possibly deduce it beforehand from any combination of laws which we have discovered by observing aggregates of a lower order. (1925, p. 79)

\textsuperscript{58} It appears that Alexander, Morgan and Broad understand levels (or orders, in Broad’s terminology) to be generated by the part-whole relation in combination with some kind of characteristic activity. This is not to say that every type of whole composes its own level of existence, nor even that it belongs to a unique level of existence (say the level of objects composed by exactly thirty-three fundamental particles), but that for object A to be at a higher-level than object B is for A to be more complex than B, or perhaps to have B as a part, and for there to be some characteristic activity or property associated with being of the kind that A belongs to. This understanding of levels does not divide the world very neatly into levels—it simply divides the parts and properties of a given individual into levels. It would not be correct to say that you or I are at a higher level than, say, a rock, because neither one of us has a rock as a part nor does it seem possible for either of us to have a rock as a part. However, you are at a higher level than your heart, because you have it as a part and you possess characteristic properties that your heart does not possess (e.g., being conscious). If you find it startling or disconcerting that this particular understanding of levels does not place you on a higher level than a rock, I understand—but, take heart, at least it does not place you on a lower level than a rock. It is simply a mistake to compare the relative “levels” of a rock and a human being.

\textsuperscript{59} Elsewhere, Broad (1925, p. 59) penned that: “The characteristic behavior of the whole could not, even in theory, be deduced from the most complete knowledge of the behavior of its components, taken separately or in other combinations…This alternative…is what I understand by the ‘Theory of Emergence’”.

64
By “non-deducibility” or as I shall prefer to call it “unpredictability”, Broad is suggesting emergent properties are, as Jaegwon Kim put it, theoretically unpredictable, not inductively unpredictable. As Kim expresses this idea:

[What is being denied by emergentists is the theoretical predictability of E [an emergent property] on the basis of M [a microphysical property]: we may know all that can be known about M – in particular, laws that govern the entities, properties and relations constitutive of M – but this knowledge does not suffice to yield a prediction of E. This unpredictability may be the result of our not even having the concept of E, this concept lying entirely outside the concepts in which our theory of M is couched. (Kim 1999, p.8)

However, ontological emergence is consistent with E’s being inductively predictable from M, meaning that having “observed that an emergent property, E, emerged whenever any system instantiated a microstructural property M, we may predict that this particular system will instantiate E at t, given our knowledge or belief that it will instantiate M at t.” (Kim 1999, p. 8)

We will see below that Kim also incorporates this epistemic thesis of unpredictability into his characterization of OE, but there is also a more thoroughgoing ontological aspect to the non-reduction advocated by British Emergentism. This comes out in their commitment to “configurational forces” or “new causal powers” at the higher-level, which in turn mandated the existence of emergent properties at the higher levels. Brian McLaughlin, in his excellent discussion of the history and doctrines of the British Emergentists, illuminates this idea as follows:

---

60 The locus classicus of the view that there are such effects, what are called heteropathic effects (attributable to the workings of different heteropathic laws at the higher-level) is John Stuart Mill’s System of Logic (1843). Heteropathic effects are distinguished from homopathic effects, which are effects that result from what might be called the mechanical mode of causes. Mill uses an analogy with dynamic forces to drive home his point. Just as two (or more) dynamical forces acting upon a body can be said to have their full effect upon the body, inasmuch as the body behaves as if the vector sum of the forces were acting upon it, so, too, do two (or more) mechanical causes acting upon some entity have their full effect, producing an effect that is, in some sense, the sum of the two causes. These homopathic effects are governed by causal laws, and Mill refers to these as homopathic laws. Mill then
British Emergentism maintains that some special science kinds from each special science can be wholly composed of types of structures of material particles that endow the kinds in question with fundamental causal powers. Subtleties aside, the powers in question “emerge” from the types of structures in question. Chemical elements, in virtue of their minute internal structures, have the power to bond with certain others…And certain kinds of organisms, in virtue of the minute internal structures of their nervous systems, have “the power of cognizing, the power of being affected by past experiences, the power of association, and so on.” (1992, p. 51)

McLaughlin goes on to explain what contributing a “fundamental causal power” amounts to:

It would imply that types of structures that compose certain special science kinds can affect the acceleration of a particle in ways unanticipated by laws concerning forces exerted by pairs of particles, general laws of motion, and the spatial or spatio-temporal arrangements of particles. In a framework of forces, the view implies that there are what we may call “configurational forces”: fundamental forces that can be exerted only by certain types of configurations of particles. (1992, p. 52)

What is being expressed here by McLaughlin is just the British Emergentist’s commitment to the causal efficacy and causal autonomy of emergent properties. Had emergent feature E not obtained, the world would have played out differently. It was this aspect of their conception of emergence that the British Emergentists used to

writes: “I shall give the name of the Composition of Causes to the principle which is exemplified in all cases in which the joint effect of several causes is identical with the sum of their separate effects.” (1843, p. 267, bk. 3, ch. 6, §1). Over against homopathic effects and homopathic laws are set the effects and laws distinctive of what Mill calls the chemical mode. These are labeled heteropathic effects and heteropathic laws. Heteropathic laws are at work in the course of natural events when it is observed that the Composition of Causes fails to hold. Heteropathic effects are most clearly present in the workings of chemical reactions and the production of new chemical compounds. They are also at work in every instance where one finds “extraordinary new uniformities” such as at the biological level. But nothing that Mill says necessarily implies the existence of ontologically emergent properties, and Mill himself never draws this conclusion. He surely does think there are new “forces” at work. He writes: “to produce a bonfire, there must not only be fuel, and air, and a spark, which are collocations, but chemical action between the air and the materials, which is a force.” (Mill 1843, Bk. 3, Ch. 5,§10. But as Alan Ryan (1970) notes in his commentary on Mill, when Mill uses concepts like “force”, he is not picking out some occult entity (Mill wants to avoid such things as badly as he wants to avoid any reference to mystical substances). Rather, Mill uses a word like force to refer to an inference rule, to “our belief that such-and-such a result would follow from such-and-such a cause.” (Ryan 1970, p. 88.) So there is no real ontological punch behind Mill’s use of force. The inference from the presence of heteropathic effects and laws to emergent properties was one made by Broad, not Mill.
account for the explanatory power of the special sciences. The final notion central to British Emergentism is that instances of emergent properties are determined by instances of the properties from which they emerge—that in some sense what is going on at the non-emergent level “fixes” what is going on at the emergent level. It was this aspect of their doctrine that they took to distinguish their view from positions that advocated the existence of Cartesian souls, or entelechies, or the like, things that housed mental phenomena but “floated free” from the physical world. Nearly all the British Emergentists thought this “fixing” was accomplished by the existence of laws relating the emergent properties and the fundamental microphysical properties. C.D. Broad referred to these as “trans-ordinal” laws, which he describes as a “statement of the irreducible fact that an aggregate composed of aggregates of the next lower order in such and such proportions and arrangements has such and such characteristic and non-deducible properties.” (1925, p. 78) What was unique about these laws was that they were not causal laws, but rather laws expressing synchronic relations between

---

61 The quote contained inside the block quote above is from Broad (1925, p. 436).

62 As Lloyd Morgan writes, “But when some new kind of relatedness is supervenient (say at the level of life), the way in which the physical events which are involved run their course is different in virtue of its presence—different from what it would have been if life had been absent.” (1923, pp. 15-16). The idea of causal autonomy comes out in Morgan’s elaboration upon this idea:

In an organism within which consciousness is emergent a new course of events depends on its presence. In a person in whom reflective thought is emergent behaviour is sustained at a higher level…Strike out reflective consciousness and action is of a lower impulsive order. Strike out all guiding consciousness and behaviour is that appropriate to the level of life. Strike out life and the course of events drops down to the physical level. (1923, p. 17).

The idea seems to be that when the emergent feature is eliminated, all of the determinative influence reverts to the lower-level. This is suggestive, it seems to me, that emergent features are causally autonomous—for if they weren’t, it would not make sense to say that removal of the emergent causes a reverting, or “dropping down” to the physical level—the physical level would already have responsibility for determining the entire course of events.
emergent property instances and their base instances. It is hard to say exactly why the British Emergentists rejected the possibility that the emergence relation was a causal one. My speculation is that it was simply their underlying assumption that placing particles in the right configuration is all that was required to bring about the generation of emergent qualities—once you place the particles in the right configuration, there is nothing more to be done, the emergent simply appears. There is no additional force or oomph that must be added. But it seems to me (as I shall point out later in the dissertation) that causal dependence has been given short shrift when it comes to the relation holding between emergent properties and their bases. Perhaps this should be given more of a look.

We can sum up by noting then that British Emergentism encompasses the following theses, the conjunction of which I will refer to as **Classic Ontological Emergence, or COE:**

**Non-Reduction:** For all M such that M is an ontologically emergent property, there is no microphysical property P to which M is reducible.

**Synchronic Microphysical Determination:** For all M such that M is an ontologically emergent property, and for all m such that m is an instance of M, m is synchronically determined by instances of microphysical properties.

**Emergent Property Unpredictability:** for all M such that M is an ontologically emergent property M, instances of M are not theoretically predictable from instances of microphysical properties.

**Novel Causal Powers:** for all M such that M is an ontologically emergent property, M contributes a novel causal power.

---

Lloyd Morgan (1923, p. 28) specifically denies that the relation that holds between an emergent property and its base properties is a causal one—as far as I can tell other British Emergentists merely fail to advance it as an option. But all of their work is suggestive of the idea that the relation is synchronic.
And this last thesis can be seen as the British Emergentists way of capturing the following two theses that ultimately distinguish ontological emergence from the views I have described as qualified micro-determinism and epiphenomenalism:

**Causal Efficacy:** For all M such that M is an ontologically emergent property, there is at least one instance of M that is causally efficacious.

**Causal Autonomy:** For all M such that M is an ontologically emergent property, M is causally autonomous.

From now on I will simply use these boldface titles to refer to these principles.

### 3.3 Jaegwon Kim’s Formulation of COE

Kim begins his investigation into the nature of emergent properties by highlighting what he takes to be the most important ideas associated with the concept:

In trying to make emergence intelligible, it is useful to divide the ideas usually associated with the concept into two groups. One group of ideas are manifest in the statement that emergent properties are “novel” and “unpredictable” from knowledge of their lower-level bases, and that they are not “explainable” or “mechanistically reducible” in terms of their underlying properties. The second group of ideas I have in mind comprises the specific emergentist doctrines concerning emergent properties, and, in particular, claims about the causal powers of the emergents. Prominent among them is the claim that the emergents bring into the world new causal powers of their own, and, in particular, that they have powers to influence and control the direction of the lower-level processes from which they emerge. This is a fundamental tenet of Emergentism, not only in the classic Emergentism of Samuel Alexander, Lloyd Morgan, and others but also in its various modern versions. (Kim 1999, pp. 5-6)

Now Kim claims to have identified two different ideas associated with emergence, but it seems to me he has actually identified three. The first is that instances of emergent properties are not predictable given a complete knowledge of the instances of their base properties (which of course we are taking to be the microphysical properties—an assumption/simplification that Kim employs as well). In other words, Kim is
endorsing **Emergent Property Unpredictability**. But Kim equates this epistemological claim with what I take to be a separate, ontological claim—that emergent property instances are not “mechanistically reducible” to the instances of their base properties. How could Kim become involved in such a confusion? The answer is that this is no confusion. It is a reflection of an underlying assumption of Kim’s, which shall be revisited momentarily, that reduction is both necessary and sufficient for reductive explanation.

### 3.3.1 Non-functionalizability

For now, however, I note simply that Kim sets out to build his analysis of an emergent property upon these two central ideas he has just identified. Kim suggests that the first idea, that an emergent property is neither reducible to nor reductively explainable in terms of underlying microphysical properties can be captured in the notion that an emergent property is not functionalizable. As the reader may recall from Chapter One, a property is functionalizable just in case it is a functional property, or, as Kim says, can “be construed, or reconstrued, as a property defined by its causal/nomic relations to other properties, specifically properties in the reduction base $B$.\)” (1999, p. 10) By the reduction base $B$ Kim simply means a certain set of microphysical properties. If $E$ is functionalizable, then some substance $S$’s having $E$ is defined as $S$’s “having some property $P$ in $B$ such that (i) $C_1, \ldots, C_n$ cause $P$ to be instantiated, and (ii) $P$ causes $F_1, \ldots, F_m$ to be instantiated.”\(^{64}\) (1999, p. 10) Kim argues

---

\(^{64}\) Kim has an endnote (endnote 16 on page 35) at this point in his article indicating that “for brevity we will often speak of a property causing another property – what is meant of course is that an instantiation of a property causes another property to be instantiated.” So he is aware of the distinction between property and property instances and often uses property to be shorthand for property instance.
that E’s being functionalizable is sufficient both for its being reductively explainable and for its being ontologically reducible. He goes on to suggest that for the latter, functionalizability is also necessary. As to why it is sufficient for reductive explanation, Kim reasons as follows:

Why does this system exhibit E at t? Because having E is, by definition, having a property with causal role C, and the system, at t, has property Q, which fills causal role C…Suppose that being in pain could be given a functional definition…Why do people experience pain when they are in neural state N? Because N is implicated in these causal/nomic relations, and being in pain is being in some state with just these causal/nomic relations. It is clear that in this way all our explanatory demands can be met.

This seems clear enough. It does seem fairly straightforward that instances of functional properties can be reductively explained. But is functionalizability really sufficient for ontological reduction? That is not so clear, and I will have more to say about this in the next section. For now, let us just note that Kim captures Non-Reduction by suggesting the following necessary condition on a property’s being ontologically emergent:

Non-Functionalizability: for all E such that E is an emergent property, E is not functionalizable

3.3.2 Novel Causal Power Contribution

The final idea that Kim brings out toward the end of the long block quote at the beginning of this section is that a property is ontologically emergent only if it contributes novel causal powers. Elsewhere, Kim has written that if some feature of our word, say mentality, has emerged then on the emergentist’s view, mental properties

*must make a genuinely new causal difference to the world.* So the following summarizes the heart of the emergentist doctrine on mental causation: mentality must contribute genuinely new causal powers to the world – that is,
it must have causal powers not had by any physical-biological properties, not even those from which it has emerged. (1992, p. 135)

I take it that when Kim says that emergent properties have powers to influence and control the direction of the lower-level processes and make a genuinely new causal difference to the world, he is concurring with what I have tried to make clear earlier—that **Causal Efficacy** and **Causal Autonomy** are necessary conditions on a property’s being ontologically emergent.

As a novel causal power is a central feature of his conception of Ontological Emergence, we might expect from Kim a fairly detailed analysis of this notion; however, he chooses to leave the concept at the intuitive level. However, I think we can give an analysis of this notion. Intuitively, what it is for a property P to contribute a novel causal power is simply for P to contribute a power that no other property does. One might think this could be captured as follows:

**Novel Causal Power Contribution Initial (simple):** A property P contributes a novel causal power C(K,E) just in case P contributes C(K,E) and there is no property Q distinct from P that contributes causal power C(K, E).

If we recall the analysis of causal power contribution that was arrived at in Chapter One, namely:

**Conditional Causal Power Contribution:** A property P contributes a conditional causal power C(K,E) just in case an instance of P, were it to occur in circumstances K, would cause (or contribute to causing) an instance of E.

then the above analysis would amount to this:

**Novel Causal Power Contribution Initial (expanded):** A property P contributes a novel causal power C(K,E) just in case an instance of P, were it to occur in circumstances K, would cause (or contribute to causing) an instance of E and there is no property Q distinct from P such that an instance of Q, were it to occur in circumstances K, would cause (or contribute to causing) and instance of E.
But this won’t do, for as formulated it is both too strong and too weak. It is too strong in that it implies that there could never be two properties that contributed one and the same causal power $C(K,E)$. But this is an unwelcome implication, for it is easy to conceive of a situation where this would be the case. Suppose there are two distinct non-microphysical properties, $R$ and $R^*$, each of which contributes exactly two causal powers: $R$ contributes $C(K, E)$ and $C(L, F)$ while $R^*$ contributes $C(K, E)$ and $C(L, G)$. The two properties are different in that they have different “total causal profiles” so to speak, but they do have one causal power in common. But suppose there is no microphysical property that contributes $C(K,E)$. Here intuitively we have two properties both of which contribute novel causal powers. But Novel Causal Contribution Initial does not allow it—the fact that $R$ contributes $C(K,E)$ disqualifies $R^*$ from contributing a novel causal power when it contributes $C(K,E)$, and vice versa.

This thought experiment illuminates what is really important when it comes to novel causal power contribution—that the causal powers be novel from the perspective of the emergent base, i.e., the microphysical properties. We thus need to stipulate in our analysis of novel causal power contribution that $Q$ is a microphysical property. But even with that amendment, Novel Causal Power Contribution Initial will be too weak. Suppose there is an instance of a structural property that is causally efficacious, or even an instance of what have been called “resultant properties”, properties like having mass of 10 kg. Intuitively, having mass of 10 kg does not contribute a novel causal power—there is nothing that an instance of this property is causally responsible for that the property instances of its parts are not responsible for. However, Novel Causal Power Contribution Initial rules that having mass of 10 kg
does contribute a novel causal power. Even though there will be microphysical properties (properties of the parts of the object that instantiates *having mass of 10 kg*) that contribute a causal power that results in E, these properties will not contribute \(C(K,E)\), because none of these properties are instantiated in circumstances \(K\).

This example reflects the general truth that no property of a composite object will ever contribute the same power as a property of one of the object’s proper parts, even though instances of the properties of those proper parts, acting collectively, could potentially cause the same effect. However, if a property contributes a novel causal power this will not be the case—if a property \(P\) contributes novel causal power \(C(K,E)\) then there will be no other property instances, acting collectively, distinct from a \(P\)-instance that make exactly the same causal contribution. I suggest that the following analysis of novel causal contribution captures this additional fact:

**Novel Causal Power Contribution:** A property \(P\) contributes a *novel* causal power \(C(K,E)\) just in case an instance of \(P\), were it to be instantiated in \(K\), would cause, in conjunction with property instances \(Q_1\ldots Q_n\), an instance of \(E\), no combination of property instances entailed to exist by \(K\) would, in conjunction with \(Q_1\ldots Q_n\), cause an instance of \(E\), and there is no microphysical property distinct from \(P\) that contributes \(C(K,E)\).

The qualification “in conjunction with property instances \(Q_1\ldots Q_n\)” is inserted due to the fact that a property instance rarely, if ever, causes an effect by itself, but normally will act as a partial cause of an effect. With this principle in hand, let us assume that Kim intends to capture the **Causal Efficacy** and **Causal Autonomy** by suggesting the following necessary condition on a property’s being ontologically emergent:

**Novel Causal Powers:** for all \(M\) such that \(M\) is an ontologically emergent property \(M\), and all \(m\) such that \(m\) is an instance of \(M\), \(m\) contributes a novel causal power (per **Novel Causal Power Contribution**).
3.3.3 Synchronic Physical Determination and Strong Supervenience

Finally, what about the **Synchronic Physical Determination** condition? Clearly, if a property instance is an instance of a functional property, it is determined by instances of another property or properties. But given that Kim rejects the idea that emergent properties could be functional properties, what notion does Kim appeal to account for **Physical Determination**? Kim thinks that supervenience will do the trick. Kim claims that any theory of Ontological Emergence must entail that “when the same basal conditions obtain, the same emergents must emerge.” (2003, p. 570) More specifically, he suggests that the emergentist should endorse Mereological Supervenience, a thesis that was introduced in Chapter One:

*Mereological Supervenience.* Systems with an identical total microstructural property have all other properties in common. Equivalently, all properties of a physical system supervene on, or are determined by, its total microstructural property. (1999, p. 7)

All properties of a physical system would of course include the emergent properties. So Kim’s way of cashing out the determination of emergent property instances by microphysical property instances is to require that emergent properties be supervenient upon microphysical properties.

But what kind of supervenience relationship holds between emergent properties and microphysical properties? It is clear from things Kim says elsewhere that he takes mereological supervenience to be a species of strong supervenience. And though there are different formulations of strong supervenience in the literature, the one I shall invoke is the following:

**Modal-Operator Strong Supervenience (MoSS):** A-properties strongly supervene on B-properties iff necessarily, if x has some property F in A, then
there is at least one property $G$ in $B$ such that $x$ has $G$, and necessarily everything that has $G$ has $F$.\(^{65}\)

Understand ‘$A$-properties’ and ‘$B$-properties’ here to refer to sets of properties (for instance, the $B$-properties might be all the microphysical properties and the $A$-properties all the chemical properties). But what will be the modal force linking emergent properties with their subvenient properties, i.e., how should we interpret the occurrences of ‘necessarily’ in MoSS when the supervenient properties under discussion are the emergent ones? As far as I can tell Kim is never clear on what the modal force should be, but I think it is safe to assume that we should interpret both instances of ‘necessarily’ to express nomological necessity and not metaphysical necessity. Doing so would at least make Kim’s analysis consistent with what others writing on the relationship between emergence and supervenience have said.\(^{66}\) Why

---

\(^{65}\) Modal Operator Strong Supervenience is to be distinguished from the other common formulation of strong supervenience, namely:

**Possible World Strong Supervenience (PoSS):** $A$-properties strongly supervene on $B$-properties iff for any individuals $x$ and $y$ in *any* possible world, if $x$ is $B$-indiscernible from $y$, then $x$ is $A$-indiscernible from $y$.

It seems clear that MoSS is stronger than PoSS. This is because it is consistent with PoSS that there be individuals instantiating $A$-properties that do not instantiate any $B$-properties (that is if you assume that two individuals both instantiating no $B$-properties qualify as being $B$-indiscernible, which seems fair enough an assumption), whereas MoSS does not allow for this possibility. There is thus a more intimate connection between property instantiations within individuals given the modal-operator versions of the thesis. If ontologically emergent properties are strongly supervenient upon microphysical properties, then the thesis that best captures this relationship is clearly MoSS and not PoSS. This is because each emergent property instance is determined by instances of microphysical properties—a given object cannot instantiate an emergent property without it (or its parts) instantiating microphysical properties. For more on the distinction between MoSS and PoSS (and other varieties of supervenience), see the excellent discussion in McLaughlin (1995).

\(^{66}\) James van Cleve builds his entire definition of an emergent property on supervenience, stating that, “[i]f $P$ is a property of $w$, then $P$ is emergent iff $P$ supervenes with nomological necessity, but not with logical necessity, on the properties of the parts of $w$”, the idea here being that the connection between emergent properties and their base properties is a contingent, albeit lawful, one (1990, p. 222). The sentiment is also echoed in the work of Tim O’Connor (1994, p. 98) and Brian McLaughlin (1997, p. 16), both of whom build upon Van Cleve’s work in constructing their analyses of emergent properties, and who both incorporate $P$’s being synchronically nomologically supervenient on lower-level microphysical properties into their analyses. It should be noted that this is early O’Connor
nomological necessity? I think it’s simply due to the belief that the emergence
relation is a lawlike relation (remember Broad’s contention that emergent properties
and their base properties are linked by trans-ordinal laws) and that lawlike
relationships express nomological necessity. Of course if one is of the opinion that
there is no distinction between nomological necessity and metaphysical necessity, then
the question of the modal force of the supervenience relationship is really a moot one.
The more important question is why strong supervenience? The main reason seems to
be that we need to identify a relation that holds between an instance of an emergent
property and instance(s) of its microphysical base property/properties that is strong
enough to ground our saying that every emergent property instantiation of some object
S is determined by instances of the microphysical properties of S (or S’s parts) Here
Kim points out that of the three most common types of supervenience—weak, strong,
and global—only strong and global supervenience are robust enough to qualify as
determination relations. But global supervenience does not capture what we want in
formulating mereological supervenience, namely, that the
microphysical/microstructural properties of any given individual S, and those

---

within the emergence discussion. Later O’Connor (e.g. 2000) has repudiated this idea. McLaughlin
suggests that in addition to an emergent property’s being nomologically supervenient on physical
properties, this supervenience relation must be a fundamental law, where a fundamental law is one that
is “not metaphysically necessitated by any other laws, even together with initial conditions” and is also
not deducible from any other laws and conditions (1997, p. 41).

Again, formulations of these supervenience theses will vary, but the two most common
formulations are probably as follows:

**Modal-Operator Weak Supervenience (MoWS):** A-properties weakly supervene on B-
properties iff necessarily, if some entity x has some property F in A then there is at least one
property G in B such that x has G, and everything that has G has F.

**Global Supervenience (GS):** A-properties globally supervene on B-properties iff all worlds
that are B-indiscernible are A-indiscernible.
properties alone, completely determine all the properties of S. Only strong
supervenience entails this.

We thus have extracted the following analysis from Kim of a property P’s
being an ontologically emergent property:

**Emergent Property Supervenience:** for all M such that M is an ontologically
er emergent property, M is strongly nomologically supervenient upon the set of
microphysical properties.  

**Non-Functionalizability:** for all M such that M is an ontologically emergent
property, M is not functionalizable.

**Novel Causal Powers:** for all M such that M is an ontologically emergent
property M contributes a novel causal power per Novel Causal Power
Contribution.

Note that Emergent Property Supervenience allegedly entails Synchronic Physical
Determination: And also note that Non-Functionalizability supposedly entails Non-
Reducibility and Emergent Property Unpredictability. So what we really have
here is an attempt at a more well-developed instance of Classic Ontological
Emergence. I will call the conjunction of these three claims K-COE, for Kimian
Classic Ontological Emergence.

3.4 Some Reasons to Think K-COE Is Too Restrictive

Obviously K-COE goes beyond MOE. In fact, every component of K-COE
goes somewhat beyond the corresponding component of MOE. Emergent Property

---

Kim does include in a footnote the following comment about Mereological Supervenience:
“Obviously extrinsic/relational/historical properties (e.g. being 50 miles to the south of Boston) must be
excluded, and the statement is to be understood to apply only to the intrinsic properties of systems.
There is also a tacit assumption that the intrinsic properties of a system determine its causal powers.”
(Kim 1999, p. 34). The lesson then is that it is only intrinsic properties of a composite object that can
qualify as emergent ones.
Supervenience entails Microphysical Determination, but places the further restriction that this determination must be synchronic. Non-Functionalizability supposedly entails Non-Reducibility, and includes the further restriction that an ontologically emergent property not be a functional property. Novel Causal Powers entails Causal Efficacy and Causal Autonomy, but entails the way in which an ontologically emergent property is causally efficacious and autonomous is via contribution of a novel causal power.

As far as the first of Kim’s necessary conditions goes, there seems to be no good reason to suppose that the determination relation holding between an emergent property and its base property must be a synchronic relation, and even less reason to suppose it must be one of strong supervenience. The main reasons that Kim advocates this condition seem to be, first, a desire to remain as faithful as possible to the writings of British Emergentism, and, second, to formulate ontological emergence as a physicalist theory. But I don’t see why we should feel compelled, in just developing a minimal conception of an ontologically emergent property, to do either of these things. Diachronic causal dependence should be an option equally open to someone who wants to construct a theory of ontological emergence; in fact, Timothy O’Connor (2000, 2001) has recently proffered a version of OE that does incorporate such an understanding of the Microphysical Determination Condition. Now, it may turn out that no account of OE that entails a relation of diachronic causal dependence between ontologically emergent properties and microphysical properties will be adequate, but we should not, without an argument to this effect, simply rule out the possibility by definition.
Similar points could be made about **Non-Functionalizability** and **Novel Causal Powers**. It may turn out that no version of OE that does not incorporate such requirements is tenable, but we should not force potential accounts of OE to satisfy these conditions. This is particularly true, I think, with the **Non-Functionalizability** condition. It is not clear at all that a property’s being functionalizable is sufficient for its being ontologically reducible. Clearly, at a minimum we have to say that P’s being ontologically reducible to Q requires that P not contribute anything causally beyond what Q contributes (or that P not be causally efficacious in a way that Q fails to be). Now, if causal contribution just amounts to novel causal power contribution, then maybe functionalizability is sufficient for ontological reduction. But it is not clear at all that that is all causal contribution amounts to. Rob Koons (unpublished), E.J. Lowe (2000), and Carl Gillett (2002b, 2002c) have all given reasons for thinking a property can make a causal contribution without contributing a novel causal power (Gillett’s reasons will be introduced in Chapter Five and discussed in Chapter Six). Until their arguments for this position are refuted, or until solid arguments are given for the conclusion that a property’s being causally efficacious can only consist in its contributing a novel causal power, then it seems best to exclude **Non-Functionalizability** and **Novel Causal Powers** from our analysis of what it is to be an ontologically emergent property.

### 3.5 Kim’s Challenge

All that having been said, there is still much to be learned from Kim. I think Kim’s greatest contribution to the debate comes in a challenge he presents to the proponent of Ontological Emergence. This challenge takes the form of an argument.
As I see it, the premises Kim uses in constructing his argument support not only the conclusion that there are no *actual* instances of emergent properties, but that it is not even *possible* that there be any instances of emergent properties (whether Kim intends his argument to be read in this way is not exactly clear, but given that his premises do support that conclusion, that is how I shall construct it). In this section I will lay out this argument against the possibility of emergent properties and show how it rests upon a couple of dubious principles. However, though I will suggest that Kim’s argument for the impossibility of emergent properties is unsound, there is still a significant challenge to the actual existence of ontologically emergent properties to be found in Kim’s work. This challenge can be expressed via a similar yet separate argument—this time for the conclusion that there are no actual instances of emergent properties. It is this second argument that I will call Kim’s Challenge.

Kim’s argument against the possibility of ontologically emergent properties (which I will label ‘Kim’s Main Argument’) appears in his 1999 piece “Making Sense of Emergence”, and is a variation on what Kim himself has previously presented under the moniker of the ‘Supervenience Argument’. The Supervenience Argument that has been widely discussed and dissected, but it is still worthwhile to review its application the question of ontologically emergent properties. To set us up for his main argument, Kim first argues for what he calls the **Principle of**

---

69 The argument I am discussing is spread out over pages 28-32 of Kim (1999). The Supervenience Argument is most succinctly and forcefully advanced in Kim 1998. Discussions of it include Crisp and Warfield (2001), Bontly (2002), Gillett and Rives (2001). However, to my knowledge no one has formulated the argument in a way that makes it clear that Kim’s objection is really an objection to the coherence of the notion of an emergent property. Also, no one has explicitly formulated Kim’s Challenge in terms of causal powers. Since the notion of a causal power is important to other things I am doing in this dissertation, I feel it is important to lay out Kim’s Challenge in terms of causal powers.
**Downward Causation**, a principle stating that, “to cause any property (except those at the very bottom level) to be instantiated, you must cause the basal conditions from which it arises (either as an emergent or a resultant).” (Kim 1999, p. 24) Now this principle is formulated very loosely, but I think we can easily restate it in a more precise fashion using the terminology I have been employing throughout this chapter:

**Principle of Downward Causation:** for all q such that q is an instance of a non-microphysical property Q and for all q* such that q* is an instance of property Q*, q* is caused or brought about by q only if q causes an instance (or instances) of some microphysical property or properties upon which Q* strongly supervenes.  

Kim’s builds his case for the **Principle of Downward Causation** in two steps, but I believe we can boil it down to one. Suppose that the instantiation of a supervenient property M brings about the instantiation of some other property M+. Assuming the following principle:

70 What if Q* is a microphysical property and is therefore such that there are non lower-level properties upon which it could supervene. Not to worry—the principle still holds as it is true that all properties strongly supervene upon themselves.

71 In what follow I am condensing two separate Kimian arguments into one, but I think it is easily seen why I am justified in doing so. In Kim (1999, pp. 22-24), Kim first argues for the conclusion that “upward causation entails same-level causation”, by which he means that, for example, to bring about an instance of some property M at level L+1 (a property that strongly nomologically supervenes upon a the set of L-properties), an instance of a member of the set of L-properties must be caused (one that is nomologically sufficient for M). He then argues that “same-level causation entails downward causation”, by which he means that, for example, to bring about an instance of some property M* at level L (a property that strongly nomologically supervenes upon a the set of (L-1)-properties), an instance of a member of the set of (L-1)-properties must be caused (one that is nomologically sufficient for M*). But both of these argument really make the same point, and can be generalized to the conclusion that for any property P such that P is strongly nomologically supervenient upon a set S of properties, an instance of P is only brought about via the causing of an instance of the member of the set of S. And if we assume that all properties of composite objects are strongly nomologically supervenient upon the set of microphysical properties, then ultimately the only way to bring about instances of any higher-level properties is by causing instances of subvenient microphysical properties to be instantiated. And that’s the gist of the **Principle of Downward Causation**.

72 For simplicity’s sake, and to stay consistent with Kim’s notation, I will here use capital letters to refer to both the property and its instance on this occasion, as does Kim. But again, it should be understood that it is the instance of the property that is involved in the causal relation, not the property itself.
Universal Supervenience: Necessarily, for any non-microphysical property Q, if an entity s has Q, there exists a microphysical base property P such that s has P, and, P is nomologically sufficient for Q, there is some subvenient microphysical property M* which is a member of M+’s supervenience base, and which is nomologically sufficient for M+. As Kim describes this situation,

M* alone suffices to guarantee M+’s occurrence on this occasion, and without M*, or an appropriate alternative base, M+ could not have occurred. This apparently puts M’s claim to have caused M+ in jeopardy. I believe that the only coherent description of the situation that respects M’s causal claim is this: M causes M+ by causing its base condition M*. The main lesson for our present concern is that if all the causally efficacious properties of composite objects (including the emergent ones) strongly nomologically supervene on microphysical properties, then, according to Kim, such higher-level properties’ causal efficacy must consist in effecting changes at the microphysical level. And with respect to causal powers, if an emergent property contributes novel causal powers to its bearer, these powers it contributes must be powers to cause instances of microphysical properties.

---

73 Universal Supervenience is of course a more wide sweeping principle than Emergent Property Supervenience, and as I am demonstrating it does play a part in Kim’s argument for the impossibility of emergent properties. And what might be tempted to ask what reasons there are for accepting it. But I think asking such a question leads us down a rabbit trail we need not investigate, for it is obvious that Universal Supervenience entails Emergent Property Supervenience. If one ends up rejecting Emergent Property Supervenience, then one will also reject Universal Supervenience. And, if one accepts Emergent Property Supervenience, then there seems no principled reason not to accept Universal Supervenience.

74 Kim 1999: 23 (I have substituted my variable notation for Kim’s in this quote)

75 For our present purposes, let us assume that Universal Supervenience does entail the Principle of Downward Causation. However, I will raise a question about whether this entailment holds in Chapter Five. It will make more sense to discuss this issue at that time rather than now.
Armed with the **Principle of Downward Causation**, Kim launches into his argument against ontologically emergent properties. He asks us to suppose that there is an instance of an emergent property M which contributes novel causal powers resulting in an instance of property M*, a property which has as its supervenience base property P*. Kim writes:

> If this is a case of downward emergent causation, M is a higher-level property, and as such it must have an emergent base, P. Now we are faced with P’s threat to preempt M’s status as a cause of P* (and hence of M*). For if causation is understood as nomological (law-based) sufficiency, P, as M’s emergence base, is nomologically sufficient for it, and M, as P*’s cause, is nomologically sufficient for P*. Hence, P is nomologically sufficient for P* and hence qualifies as its cause. The same conclusion follows if causation is understood in terms of counterfactuals—roughly, as a condition without which the effect would not have occurred. Moreover, it is not possible to view the situation as involving a causal chain from P to P* with M as an intermediate causal link. The reason is that the emergence relation from P to M cannot properly be viewed as causal. This appears to make the emergent property M otiose and dispensable as a cause of P*. (1999, p. 32)

Let’s try to break down Kim’s case. I will ignore the reference Kim makes to understanding causation in terms of counterfactuals and assume that he wants us to understand causation in terms of nomological sufficiency (everything that I will say here about understanding causal power contribution in terms of nomological sufficiency could also be said about understanding it in terms of counterfactuals).

I think we can most profitably view the argument as a *reductio* of the claim that ontologically emergent properties are possible. The reason we can view the argument as concluding that no such instance is even *possible* is that every premise of the argument is, if true, necessarily true, and thus the conclusion, if sound, is true in every possible world. Here is my suggestion for formalizing this argument.
Kim’s Main Argument

(1) Suppose it is possible that there be an instance of an ontologically emergent property M.

(2) If (1), then there is a possible world W where at some time t there exists an instance m of M which is causally efficacious in some circumstance K in bringing about of an instance m* of a distinct property M*.

(3) If (2), then m must bring about m* by causing an instance p* of M*’s supervenience base P*. (from Principle of Downward Causation)

(4) If (1) and (3), then M contributes novel causal power C (K, P*). (from K-COE)

(5) If M contributes novel causal power C(K, P*), then in circumstances K, m causes, in conjunction with some other property instances Q1…Qn, p*, and there is no combination of property instances entailed to exist by K that cause, in conjunction with Q1…Qn, p*. (from Novel Causal Power Contribution).

(6) There exists at t an instance p of a microphysical property P such that p is nomologically sufficient for m (entailed by Emergent Property Supervenience).

(7) In circumstances K, m causes, in conjunction with some other property instances Q1…Qn, p*, and there is no combination of property instances entailed to exist by K that cause, in conjunction with Q1…Qn, p*. [entailed by (1)-(5)].

(8) For all A and all B such that A and B are properties, an instance a of A causes an instance b of B iff a is nomologically sufficient for b, and a and b are non-simultaneous. [assumed analysis of causation]

(9) In circumstances K, m, in conjunction with Q1…Qn, is nomologically sufficient for p*, and there is no combination of property instances entailed to exist by K that are, in conjunction with Q1…Qn, nomologically p*. [entailed by (7)-(8)].

(10) P is a property that is such that an instance p of P, in conjunction with Q1…Qn, is nomologically sufficient for p*. [entailed by (6), (9), and transitivity of nomological sufficiency].

(9) and (10) contradict each other. Thus, the supposition in premise (1) is false—it is not possible for there to be an instance of an ontologically emergent property.
This seems to me to be a faithful representation of Kim’s argument. However, this formulation of Kim’s argument has one curious feature. Premise (3), the **Principle of Downward Causation**, really does no work in the argument. To see this, I ask the reader to perform a little mental exercise (OK, one that might require some paper). Suppose we remove premise (3), and simply allow that $m$ causes $m^*$ directly. Then in the remaining premises of the argument one would perform a uniform substitution of $M^*$ for $P^*$ and $m^*$ for $p^*$. If one does this, (9) and (10) would still contradict each other. So the **Principle of Downward Causation** turns out to be irrelevant to Kim’s argument as I’ve formulated it. Perhaps this is suggestive of the fact that I have actually misinterpreted Kim’s argument. In any case, even though the **Principle of Downward Causation** is not relevant to this argument, it is relevant to the subsequent argument I’ll be discussing shortly, namely, Kim’s Challenge. So this is why I have included it (this, along with the fact that Kim seems to claim that it should be included).

What should we conclude about **Kim’s Main Argument**? It’s formally valid, and if all of the premises are true then we would have to admit that ontologically emergent properties, as expressed in K-COE, are impossible. But by my lights all of the premises are certainly not true—either (5) or (8) is false. Here’s an example that indicates either (5) or (8) must be false.

Suppose I find myself, through a series of highly unfortunate events, on a hostile planet which is teeming with brain-eating creatures. These creatures have an additional sense which allows them to immediately detect the presence of brains. I am aware of this, and am obviously distraught over my situation. I reflect upon the fact that, given my current circumstances, it is highly doubtful I will ever live to see the
Cubs win the World Series. But, I still at that moment believe that the Cubs will win the World Series (when all hope in one thing is lost, it doesn’t mean one can’t have hope in another). Suppose A is the property of “believing the Cubs will eventually win the World Series”, A* is the property of “having a brain”, and B is the property of “sensing a brain in the area.” Assuming that it is a law in our world that only things having brains can have beliefs, my instantiating A at t is nomologically sufficient for my instantiating A* at t. Suppose at t also the presence of my brain causes one of my enemies to detect the presence of a brain, thus instantiating property B. If (8) is correct, then since my believing that the Cubs will eventually win the World Series at t (an instance of A) is nomologically sufficient for my having a brain (an instance of A*), and since my having a brain is nomologically sufficient (in conjunction with some other property instances) for the alien’s detecting one (an instance of B), then my believing that the Cubs will eventually win the World Series is a partial cause of the alien’s detecting my brain. This seems patently false.

But even if one could make oneself believe that A qualifies as a cause (if one were really, truly committed to a nomological sufficiency account of causation), surely one could not believe that A contributes the causal power to bring about B. No intuitive conception of causal power contribution would entail that. So in that case even if one believed (8) to be true, one would still have to judge (5) to be false.\textsuperscript{76}

\textsuperscript{76} This example alone helps highlight the reason to talk about causal powers instead of causation. It is much clearer when one speaks of causal powers that the instance A does not really contribute anything to the bringing about of the instance of B. Kim employs causal power talk to some extent in his work. But when it comes to giving the argument against emergent properties, he drops it in favor of talk about causation. Perhaps now we see the reason why.
So Kim’s Main Argument against the possibility of an ontologically emergent property is not sound. But that does not mean that we can find no challenge to their existence in Kim’s work. I think there is one. To see this, we first recognize that there is something intuitively attractive about (9). But the reason one might think that in this case p is a cause of p* has nothing to do with the transitivity of nomological sufficiency, but the belief in microphysical causal closure. We can express this principle as follows:

**Microphysical Causal Closure**: if a microphysical event (an instance of a microphysical property) has a cause at t, it has a sufficient microphysical cause at t.

Stripping away the premises about nomological sufficiency and replacing them with this premise dealing with Causal Closure gives a new argument that looks like this:

1. Suppose there exists at t an instance m of an ontologically emergent property M which is causally efficacious in some circumstance K in the bringing about of an instance m* of a distinct property M*.
2. If (1*), then m must bring about m* by causing an instance p* of M*'s supervenience base P*. **(Principle of Downward Causation)**
3. If (1*) and (2*), then M contributes novel causal power C (K, P*). **(K-COE)**
4. If M contributes novel causal power C(K, P*), then in circumstances K, m causes, in conjunction with some other property instances Q1…Qn, p*, and there is no combination of property instances entailed to exist by K that cause, in conjunction with Q1…Qn, p*. **(Novel Causal Power Contribution)**
5. In circumstances K, m causes, in conjunction with some other property instances Q1…Qn, p*, and there is no combination of property instances entailed to exist by K that cause, in conjunction with Q1…Qn, p*. [ entailed by (1*), (2*), (3*), and (5) ].
In circumstances K, there is at t an instance r of some microphysical property R such that r causes, in conjunction with Q1…Qn, p*. (Microphysical Causal Closure)

Now it is (7) and (10) that result in the contradiction. I will call this argument Kim’s Challenge.

Note that since Microphysical Causal Closure is only a thesis about our world, and not all possible worlds, an argument constructed around it can not disprove the possibility of emergent properties. So (1) has been replaced with (1)*. Note also that when the argument is set up like this, then the Principle of Downward Causation becomes relevant. However, Emergent Property Supervenience disappears as an explicit premise (although Universal Supervenience still plays a role in justifying the Principle of Downward Causation).

In Kim’s Challenge we now have an argument that does highlight a genuine concern for the proponent of OE. What the proponent of OE must do is deny one of the premises, namely, (2)*, (3)*, (5), or (10). But of course the premise she denies cannot be one entailed by her theory, thus, for example, denying (3)* would require rejecting K-COE as an adequate characterization of an ontologically emergent property. In the next chapter I want to introduce some theories of what it is for a property to be an ontologically emergent property that deviate from K-COE, and show how these theories offer resources to respond to Kim’s Challenge.
CHAPTER 4

ALTERNATE CONCEPTIONS OF ONTOLOGICAL EMERGENCE

4.1 Introduction

In the last chapter, I offered up a characterization of ontological emergence that I feel best embodies the spirit of the writings of the British Emergentists, a conception of ontological emergence that I dubbed Classic Ontological Emergence, or COE, I then reviewed Jaegwon Kim’s contemporary spin on COE, which I referred to as K-COE. According to K-COE, the following three theses are necessary conditions on any theory of ontological emergence:

**Emergent Property Supervenience:** for all M such that M is an ontologically emergent property, M is strongly nomologically supervenient upon the set of microphysical properties.

**Non-Functionalizability:** for all M such that M is an ontologically emergent property, M is not functionalizable.

**Novel Causal Powers:** for all M such that M is an ontologically emergent property M contributes a novel causal power per Novel Causal Power Contribution.

There is no doubt that this conception of OE was dominant throughout twentieth philosophy and has, for the most part, continued that dominance even throughout the revived interest in OE that has developed over the last fifteen to twenty years. But it is certainly not the only conception of OE, and in this chapter I turn to the task of identifying and explaining some alternative conceptions that have popped
up in the recent literature. In one way or another, each of the alternative conceptions I will review in this chapter is motivated by Kim’s Challenge. The upshot of Kim’s Challenge is that Novel Causal Powers is inconsistent with the conjunction of three highly plausible theses, and thus rational belief in ontologically emergent properties requires the rejection of one of the theses. These three theses (all of which have been introduced and discussed in previous chapters) are:

**Principle of Downward Causation:** for all q such that q is an instance of a non-microphysical property Q and for all q* such that q* is an instance of property Q*, q* is brought about q only if q causes an instance (or instances) of some microphysical property or properties upon which Q* strongly supervenes.

**Novel Causal Power Contribution:** A property P contributes a novel causal power C(K,E) just in case an instance of P, were it to be instantiated in K, would cause, in conjunction with property instances Q1…Qn, an instance of E, no combination of property instances entailed to exist by K would, in conjunction with Q1…Qn, cause an instance of E, and there is no microphysical property distinct from P that contributes C(K,E)

**Microphysical Causal Closure:** if a microphysical event (an instance of a microphysical property) has a cause that occurs at t, it has a sufficient microphysical cause at t.

Nearly everyone who is sympathetic to understanding OE along the lines of K-COE has argued (or simply assumed) that the existence of ontologically emergent properties requires a rejection of Microphysical Causal Closure. And since a rejection of this

---

77 This is not to say that everyone who has proposed an alternate understanding of OE has done so explicitly in response to Kim’s writings (though that is nearly the case), but rather to say that the considerations that motivate these different conceptions are all ones that are operative in Kim’s Challenge.

78 See for example Crane (2001), Kim (1999), Hasker (1999), and McLaughlin (1997). Lowe (2000) also assumes that conceiving of OE along the lines of K-COE implies that Microphysical Causal Closure is false, though he then goes on to suggest an alternate consumption of OE that does not entail the denial of Microphysical Causal Closure (a conception that will be revisited in Chapter Six). The notable exception to those who think K-COE goes along with the denial of Microphysical
principle is viewed by many to be a rejection of physicalism, many philosophers have scrambled to find a conception of OE that would be consistent with Microphysical Causal Closure and thus consistent with physicalism. As the reader can probably predict, philosophers who have proposed alternate conceptions of OE have constructed them in such a way that either the Principle of Downward Causation, Novel Causal Power Contribution, or Novel Causal Powers itself, can be plausibly denied without undermining the existence of ontologically emergent properties as conceived under her theory.

One might think then that a good way to divide up the alternate conceptions of an ontologically emergent property currently on offer is to divide them by which of the above three theses their proponents deny. And perhaps that would be a good way to carve up the landscape, but I don’t think it’s the best way. First of all, not all philosophers propose alternate conceptions of OE with the goal of salvaging Microphysical Causal Closure and thus with their eyes toward the denial of Principle of Downward Causation, Novel Causal Power Contribution, or Novel Causal Powers. Some philosophers have rejected K-COE because they do not like one of the conditions other than Novel Causal Powers. Second, there is a more straightforward and elegant way of sorting out theories of OE and that is to divide up different theories of OE based on different interpretations of the conditions I have

\footnote{Causal Closure is Sydney Shoemaker, who seems to believe that it is Novel Causal Power Contribution that should be denied, not Microphysical Causal Closure. See Shoemaker (2001). More on this later in the chapter.}

\footnote{O’Connor and Wong (forthcoming), for instance, reject K-COE because they think it is improper to view ontologically emergent properties as supervening on microphysical properties. The proper dependence relationship, by their lights, is a non-supervening causal one. They still think the proper response to Kim’s Challenge is to reject Microphysical Causal Closure.}
identified as minimally necessary for such a theory. These conditions, which together constitute what I call MOE, were given at the end of Chapter Two:

**Microphysical Determination:** For all m such that m is an instance of an ontologically emergent property, the occurrence of m is ultimately determined by an instance or instances of microphysical properties.

**Causal Efficacy:** For all M such that M is an ontologically emergent property, there is at least one instance of M that is causally efficacious.

**Causal Autonomy:** For all M such that M is an ontologically emergent property, M is causally autonomous.

Sorting out theories this way will allow us to construct a two-dimensional grid, with different candidates for the microphysical determination relation on one axis and different understandings of what it is for an ontologically emergent property to be causally efficacious and autonomous on the second axis. As a sneak preview for the reader, I will tell you that this grid is 4 x 2—four possible microphysical determination relations and two possible understandings of causal efficacy and autonomy—giving us eight “boxes” into which we may insert theories of OE. As it turns out, however, not all of these boxes represent genuine possibilities as some microphysical determination relations are incompatible with particular understandings of causal efficacy and autonomy.

In any case, **K-COE** will inhabit Box I on the grid (that’s “I” as in “the Roman numeral for one” not as in “the letter I”). **K-COE** is most naturally interpreted as entailing that the determination relation holding between ontologically emergent properties and microphysical properties is one of mere supervenience, a term of art that I am introducing here and which I will define below. The other two relations that have been proposed as candidates for the determination relation are ones that will be more familiar to the reader—realization and causation (I will be distinguishing two
different types of realization, thus giving us four total relations. As for interpretations of causal efficacy and autonomy, K-COE obviously entails that an ontologically emergent property contribute a novel causal power. But there is another interpretation in the offing, namely, that an ontologically emergent property is causally efficacious and autonomous by synchronically determining instantiations of microphysical causal powers.

So as the reader has surely by now ascertained, one primary goal of this chapter is to provide something like a taxonomy of current theories of OE. But before laying out the taxonomy, there is some metaphysical work to be done. As I’ve already mentioned, one of the microphysical determination relations I propose to discuss, mere supervenience, is not one that has been explicitly addressed in the literature, and so the metaphysics of this relation has obviously received no attention (by metaphysics of this relation I mean, more or less, the necessary and sufficient conditions for this relation’s obtaining). It is incumbent upon me to describe this relation in more detail. But it is also the true that realization, though bandied about with great frequency in the literature, has received little metaphysical attention. In fact, I think there is a good deal of equivocation on the term ‘realization’. So, in this chapter I will distinguish two distinct relations that have been labeled ‘realization relations’ in the literature (relations that are not necessarily coextensive), and spend some time investigating the metaphysics of both.\footnote{I note here that I will only briefly touch upon the last of the microphysical determination relations, namely, causation. This is not because I see it as a second-class citizen—I actually believe there is much promise in viewing the relationship between emergent properties and microphysical properties as a diachronic causal one—but simply because, first of all, I have nothing to add to what has already been said in the literature about what the metaphysics of the causation relation, and, secondly, I will not interact critically in this dissertation with any theories of OE that entail the}
chapter as I will be devoting a good portion of Chapter Six to the notion) is an alternative conception of causal efficacy and autonomy that has recently been proposed by Carl Gillett. According to Gillett, emergent properties are causally efficacious and autonomous by synchronically determining instantiations of microphysical causal powers.

I start then, by looking at the different options in play for the Microphysical Determination holding between an ontologically emergent property and its microphysical base properties.

4.2 Options for the Microphysical Determination Condition

4.2.1 Mere Supervenience

“Mere supervenience” is a term of art that I am introducing here—it is not one that has had any play in the literature. But I do think the relation as I will describe it is what is intended by most philosophers who invoke supervenience as the proper understanding of the determination relation between ontologically emergent properties and their microphysical base properties. So what is meant when it is said that emergent properties “merely supervene” on physical properties? Well, rather obviously, there are two key parts to the concept of mere supervenience—the “supervenience” part and the “mere” part. Let us begin with the notion of supervenience itself.

The “core idea of supervenience”, as Brian McLaughlin puts it, is dependent-variation, of a purely modal sort—if A-properties supervene on B-properties then if

---

microphysical determination relation to be one of causation (again, not because I think these views are unimportant, but simply because my project here is limited).
there is a variation in A-properties (i.e. if two things vary in the A-properties they instantiate or include\textsuperscript{81}), then there must be a corresponding difference in B-properties. (1995, p. 18)\textsuperscript{82} As a purely modal relation of dependent-variation, supervenience is by nature reflexive and transitive, but neither symmetrical nor asymmetrical. It is consistent with A-properties supervening on B-properties that B-properties also supervene on A-properties.\textsuperscript{83} The supervenience I have in mind in talking about the supervenience of ontologically emergent properties on microphysical properties is the type of supervenience that is at work in K-COE, strong nomological supervenience as expressed in MoSS.

Now let me explain the idea of mere supervenience, first intuitively, and then more precisely. For the set of A-properties to merely supervene on the set of B-properties is for the A-properties to supervene on B-properties, but for this supervenience relation to be neither a robustly explainable relation nor a causal relation.\textsuperscript{84} Roughly, a supervenience relation is robustly explainable just in case

\textsuperscript{81} I say “instantiate or include” because certain types of supervenience relations are formulated in terms of worlds, e.g., global supervenience which can be formulated as A-properties globally supervene on B-properties iff all worlds that are A-twins are B-twins. But in most cases the worlds in question do not instantiate the supervening or subvening properties (for instance, worlds don’t instantiate mental properties)—rather what we are really interested in is the distribution of properties across worlds, the way the worlds “include” these properties.

\textsuperscript{82} McLaughlin notes, rightfully so, that the relata of a supervenience relation are not essentially sets of properties—sets of events, or states of affairs, or laws could supervene on each other as well. But for my purposes I will always assume that the relata of the “mere supervenience” relation are sets of properties.

\textsuperscript{83} For instance if set A is the singleton \{being triangular\} and set B the singleton \{being trilateral\}, then A-properties supervene on B-properties, and B-properties also supervene on A-properties.

\textsuperscript{84} The term ‘robustly explainable’ comes from Terry Horgan (1993). My interpretation of what robustly explainable means is gleaned from Horgan, as well as others who have interpreted him such as Wilson (1999) and Kim (2003b)
knowledge of all the facts and laws governing the subvenient properties together with a conceptual grasp of the relevant A-properties is sufficient for knowing that an instance of that relation obtains. An example of a physically explainable relation is the relation that holds between a functional property and its realizer, as was described in Chapter Two. Suppose the functional property under consideration is that of being a paperweight. Given that this is what we might call a “forward-looking” functional property, in that it is defined simply by the kinds of events its instantiations cause and not by the kinds of events that cause its instantiations, it can be conceptualized, as follows: being a paperweight just is having one of a set of properties \( P_1, \ldots, P_n \) in \( B \) (which in this case will be a particular set of microstructural properties) such that instances of \( P_1, \ldots, P_n \) (in the right circumstances) cause instances of \( F_1, \ldots, F_m \).\(^8\)

Then, grasping this concept concurrently with knowing all of the causal laws involving the members of \( B \) will be sufficient for knowing that the property being a paperweight is instantiated.

Functional properties then are not merely supervenient upon their base properties. An example of a supervenience relation that is an example of mere supervenience in that it is not robustly explainable is the type of “brute bridge relation” linking reduced properties with reducing properties in classic Nagel reduction. Nagel (1961) conceived of reduction as the derivation of laws of the reduced theory from the laws of the reducing theory (combined with the laws of

\(^8\) Though it is not critical for the purposes of this example to precisely identify them, \( F_1, \ldots, F_n \) would probably be properties describing the velocity and trajectory of microphysical objects composing a piece of paper—the property of being a paperweight, in circumstances that would probably include the existence of medium to strong breezes or weak to medium tugs on the paper will cause the velocity and trajectory of certain particles to remain at or near zero. Of course, specifying exactly what \( F_1, \ldots, F_n \) should be would be terribly complicated.
logic). However, since each of these theories were couched in their own vocabularies, which often did not cross-classify, extra premises, the bridge laws, were required to complete the reduction. The ontological implications of Nagel reduction seem to be that each property referenced in the reduced theory is nomologically coextensive with a property referenced in the reducing theory. Thus, properties described in the reduced theory are strongly supervenient (where the relevant sense of the second occurrence of ‘necessarily’ in the definition is nomological) upon properties described in the reducing theory. However, it is consistent with Nagel reduction that these bridge laws are brute and thus unexplainable by the microphysical laws combined with the laws of logic.86

Now let us return to our specific question regarding the relation of ontologically emergent properties to microphysical properties. Claiming that the relation between ontologically emergent properties and microphysical properties is one of mere supervenience really amounts to saying that ontologically emergent properties supervene upon microphysical properties but that their instantiations are not explainable from the laws and facts regarding microphysical properties. But we have seen this idea before—in the last chapter under the guise of Emergent Property Unpredictability. We thus might be tempted just to say that the mere supervenience of ontologically emergent properties on microphysical properties just amounts to the conjunction of two theses we encountered in the last chapter, namely:

86 Kim (1999, p. 13) notes that Nagel reduction is consistent with many dualist theories that take the appearance of mental phenomena to be unpredictable from physical phenomena, including pre-established harmony, occasionalism, and Cartesian dualism.
Emergent Property Supervenience: for all M such that M is an ontologically emergent property, M is strongly nomologically supervenient upon the set of microphysical properties.

and

Emergent Property Unpredictability: for all M such that M is an ontologically emergent property M, instances of M are not theoretically predictable from instances of microphysical properties.

But this is not quite right. Mere supervenience, as I am defining it, requires not just that these two theses hold, but also that there be no causal relation holding between an ontologically emergent property and microphysical properties that grounds the supervenience relation. Mere supervenience is intended to be a synchronic relation, and causation is usually understood to be diachronic. But even if one could make sense of synchronous causation, I, for a more practical reason, want to exclude causal relations from counting as mere supervenience relations in that I want to deal with causal dependence as a separate category of determination.

But we’re still not quite there. If mere supervenience is to serve as the relation between ontologically emergent properties and microphysical properties, then it must qualify as a determination relation. But now recall our preferred formulation of strong supervenience from Chapter Three:

Modal-Operator Strong Supervenience (MoSS): A-properties strongly supervene on B-properties iff necessarily, if x has some property F in A, then there is at least one property G in B such that x has G, and necessarily everything that has G has F.

If the necessity here is interpreted as nomological necessity, does the fact that A-properties strongly supervene on B-properties guarantee that A-property instantiations are determined by B-property instantiations? I suggest to you that the answer is “no”. The problem is that determination should be an asymmetric relation, whereas strong
supervenience can be either symmetric or asymmetric. For instance, suppose the set of A-properties is \{being triangular\} and the set of B-properties is \{being trilateral\}. In such a case A-properties strongly supervene on B-properties, and vice versa, but it seems wrong to say that the B-property instantiations determine A-property instantiations (or vice versa). Here the two sets of properties covary simply because they both strongly supervene on a third set of properties, namely, the pattern of instantiation of fundamental (microphysical) properties and relations, and it is these properties that determine instantiations of triangularity and trilaterality. But even if there were no more fundamental properties to determine the instantiations of the A- and B-properties, still we would not be inclined to accept mutual determination. Suppose there is a world (maybe it’s our world, maybe not, it doesn’t really matter) in which it is true, of nomological necessity, that two microphysical properties (call them J and K) are, if instantiated, co-instantiated. So \{J\} nomologically strongly supervenes on \{K\} and vice versa. Were strong supervenience sufficient for determination, we would have to say that J determines K, and vice versa. But this seems intuitively false. So for any supervenience relation to qualify as a determination relation, we must somehow build in the notion that the relation is asymmetric. This is easily enough done. Now, we can finally formulate what we can call Mere Supervenience:

---

\(^{87}\) That there is confusion over the question of whether strong supervenience qualifies as a determination relation is no better evidenced than in the fact that the master of supervenience himself, Jaegwon Kim, seems to flip-flop on the issue with in the same paper. In one of his most recent essays (2003a) on supervenience, Kim writes that the “second way of explaining the core idea of supervenience is in terms of ‘determination’: if A-properties supervene on B-properties, any object’s B-properties will determine all of its A-properties.” (2003b, p. 559). But later on Kim, correctly by my lights, reverses course and says that supervenience relations “are not asymmetric; they do not prohibit
**Mere Supervenience:** a set of properties \( S_A \) is merely supervenient upon a second set of properties \( S_B \) if and only if \( S_A \) strongly nomologically supervenes on \( S_B \), but not vice versa, and for all \( E \) such that \( E \) is a member of \( S_A \), (i) instances of \( E \) are not theoretically predictable from instances of the members of \( S_B \) and (ii) no instance of \( E \) is caused (or partially caused) by an instance of some member of \( S_B \).

So, in **Mere Supervenience** we have one candidate for the determination relation holding between emergent properties and microphysical properties. But there are two others still to be discussed. The first is realization.

### 4.2.2 Realization

The next relation that has been advanced in the literature as being compatible with OE is realization. Talking about realization is somewhat tricky as there is nothing even close to a standard use of the term in the literature. As a matter of fact, the nature of the realization relation itself is most often not discussed at all; rather, what is provided is a characterization of what it is for a property to qualify as realized. About the only thing that could be identified as commonly entailed by all conceptions of the realization relation is that the realization relation is a synchronic relation that is a step up from mere supervenience, and is such that instances of realized properties are necessitated in virtue of instances of microphysical properties. In other words, whereas supervenience is a relation of dependent-variation, of a purely modal sort, and thus the holding of this relation does not entail any kind of robust connection between supervening property and its subvenient property (it could just be that the two are related by some accidental law, or perhaps not even related by law at all, but

---

*two families of properties supervening, in any of the three senses, on each other, whereas we expect the relation of determination or dependence to be asymmetric.* (2003b, p. 563).
merely co-instantiated in some interesting way), realization entails that the realized and realizing property have a very “tight” connection. As Kim writes,

The statement that P realizes F surely hints at a stronger relation than mere covariation; it includes the idea that F’s instantiation in a given case holds \textit{in virtue of} the fact that P is realized on that occasion, or perhaps, something even stronger, that F’s instantiation on a given occasion \textit{consists in} P’s instantiation on that occasion. (2003b, p. 578)

This “in virtue of” locution occurs over and over in characterization of realization, and it is what grounds the realization relation as one more robust than the relation of mere supervenience. It is the differing interpretations of this “in virtue of” locution that mark out the two different notions of realization I’ll be discussing.

Before distinguishing the two relations, however, I should perhaps explain why there has been interest in constructing a theory of OE around realization, even when it is fairly clear that the British Emergentists clearly understood the dependence relation between emergent properties and physical properties to one of mere supervenience. I think it is safe to say that the interest comes from physicalists who, gripped with the sense that physicalism entails the realization of all nonmicrophysical property instances, hold the belief that if a theory of OE is to be successful it \textit{must} incorporate realization as the determination relation.\footnote{For example, Poland (1994), Melnyk (2003) and Gillett (2002b) have all argued that physicalism entails the realization of all nonmicrophysical property instances. Of these three only Gillett has tried to make realization compatible with emergence, but there are others who have proffered theories of ontological emergence that are clearly compatible with realization, e.g., Derk Pereboom (2002) whose views will be discussed in Chapter Five.} I will not discuss here whether physicalism does in fact entail that all properties are realized by microphysical properties, for that is not really essential to the debate about OE itself. Learning whether or not OE is
compatible with realization is an interesting fact independently of whether physicalism requires realization.

4.2.2.1 Distinguishing Two Realization Relations: $R_{THEO}$ vs. $R_{CAUS}$

Now let us move on to the metaphysical issues surrounding realization. A good number of philosophers have taken the phrase “in virtue of” to express an explanatory connection between facts about the relata of the realization relation. So, for instance, Ernie LePore and Barry Loewer write that:

The usual conception is that e’s being P realizes e’s being F iff e is P and there is a strong connection of some sort between P and F. We propose to understand this connection as a necessary connection which is explanatory. The existence of an explanatory connection between two properties is stronger than the claim that $P \rightarrow F$ is physically necessary since not every physically necessary connection is explanatory. (1989, p. 180)

Terry Horgan tries to make more precise the idea of LePore and Loewer when he suggests that for a relation to count as one of realization (or what he calls superdupervenience) it must be “robustly explainable”, a notion introduced in the last section. In that section I equated the relation holding between emergent properties and microphysical properties being robustly explainable with the epistemic notion of this relation’s being theoretically predictable from the physical facts and laws. So the first relation that has been identified as a realization relation is the one that holds between two properties $X$ and $Y$ iff instances of $X$ necessitate instances of $Y$ and facts about instances of $Y$ are theoretically predictable from the facts about instances of $X$, meaning that were one to know all the facts and laws governing $X$ (excluding facts that involve $Y$), and were one to possess the concept of $Y$, one would be able to predict that $Y$ is instantiated. Let us call this type or realization relation ‘$R_{THEO}$’ (for
“theoretically predictable”). I will return shortly to the question of what metaphysical grounding $R_{THEO}$ might have, but first let me turn to the second notion of realization.

The second understanding of the realization relation trades upon what Carl Gillett has called the “ur-notion” of realization, one that hearkens back to the original use of the term by philosophers like Hilary Putnam and Jerry Fodor in the debates about functionalism.\footnote{Gillett (2002a, p. 316). For a “classic” paper that suggests this understanding of realization, see Putnam (1967).} As Gillett notes, the key element of this notion was that a realizing property “plays the causal role” of the realized property, i.e. that whenever an instance of a realized property causes an event $e$, it does so in virtue of the instance(s) of its realizer(s) causing $e$; the instance of the realized property “inherits its causal powers” from its realizer(s). A common example of this type of realization relation so understood is the one holding between a functional property and its realizer. Suppose $F$ is the functional property of having some property or other the instantiation of which is caused by instances of certain physical event types $C_1$…$C_n$ and causes instances of certain physical event types $E_1$…$E_n$. Suppose that in a given situation $P$ is instantiated, thereby resulting in $F$. $P$ causes event types $E_1$…$E_n$, and, if $F$ causes anything, it also causes event types $E_1$…$E_n$, and nothing beyond that.\footnote{Some people have suggested that functional properties have no causal powers at all. I will not address that possibility, as whether or not they have any causal powers does not really affect anything that comes after in this section. The important point is that if a functional property has any causal powers, then it has the same causal powers as its realizer.} The causal powers, then, of the realized functional property do not go beyond that of the realizing property. Sydney Shoemaker expresses the idea thusly—“property $X$ realizes property $Y$ just in case the conditional powers bestowed by $Y$ are a subset of
the conditional powers bestowed by $X$…” (2001, p. 78) Let us call this understanding of realization ‘$R_{CAUS}$’ (for the relation that expresses causal power subsumption).

Now the relation that holds between a functional property and its base property, the example I just used to illuminate $R_{CAUS}$, is also an example of $R_{THEO}$, as was shown in the previous section on “mere supervenience”, the relation between a functional property and its base property is one that guarantees the theoretical predictability of the functional property instantiations. But in general it is not the case that a relation that qualifies as an instance $R_{CAUS}$ will qualify as an instance of $R_{THEO}$, or vice versa. To see this, suppose there is a possible world $W$ in which an instance of the property $being in pain$ stands in $R_{CAUS}$ to an instance of some physical property $P$, in other words, as it so happens, the causal powers of $being in pain$ in this world are just a subset of the causal powers of $P$. Does this fact entail that instantiations of $being in pain$ are theoretically predictable from instantiations of $P$? Not if the concept of $being in pain$ is not that of playing some causal role, but rather that of imparting some kind of qualitative feel or sensation. It may be that it is only contingently the case that the causal powers of $being in pain$ are a subset of the causal powers of $P$. In such a case there is no a priori connection between the physical facts and laws to the instantiation of $being in pain$, so even though an instance of $being in pain$ stands in $R_{CAUS}$ to an instance of $P$, it will not stand in $R_{THEO}$ to that instance. Conversely, instantiations of property $Q$’s being necessitated by and theoretically predictable from instantiations of some property $P$ (and thus bearing $R_{THEO}$ to instantiations of $P$) does not entail that an instance of $Q$ bears $R_{CAUS}$ to an instance of $P$, mainly because an instance of $Q$’s bearing $R_{THEO}$ to an instance of $P$ does not require that $Q$ possess any causal powers at all, whereas an instance of $Q$ bearing $R_{CAUS}$ to an instance of $P$ does.
Suppose $Q$ is a second-order property over $P$ (so $Q$ is just the property of having $P$), or perhaps a disjunctive property like being $P$ or $R$. Instances of such properties bear $R_{\text{THEO}}$ to instances of $P$, but, intuitively, do not bear $R_{\text{CAUS}}$ to instances of $P$, for the instances of $Q$ possess no causal powers at all.$^{91}$

So we have two distinct relations, $R_{\text{THEO}}$ and $R_{\text{CAUS}}$, each referred to under the guise of ‘realization’. But what is the upshot of this distinction? Why take the time to make this distinction? Well, first of all, it makes sense of some of the confusion that is present in the realization debate. When authors offer up different characterizations of realization (some of which I shall discuss presently) one needs to understand that the authors may not necessarily be disagreeing, but may actually be describing different relations. But as important for our purposes is the truism that distinguishing the two different realization relations in the way I have has a big impact on our understanding of the relationship between realization and emergence. Let me explain why.

Recall that the conception of ontological emergence that has dominated the recent literature is the one expressed by Jaegwon Kim, which we are calling $\text{K-COE}$. According to $\text{K-COE}$ a property’s being ontologically emergent entails that it is not a functional property. The main argument against a realized property’s possibly qualifying as an ontologically emergent property has been that since a realized property “inherits its causal powers” from its realizer, no realized property can

$^{91}$ If you are not persuaded by this example, then it is probably due to the fact that you believe that there are no properties that fail to have causal powers—that being a property just is possessing (or being constituted by) some set of causal powers. If that is your position, then you may be led to conclude that $R_{\text{THEO}}$ being instantiated does entail $R_{\text{CAUS}}$ being instantiated. But my point that they are not co-extensive relations is still proven by the less-contentious argument that the reverse entailment does not hold.
contribute a novel causal power. As Jaegwon Kim puts it in what he has labeled the Causal Inheritance Principle,

**Causal Inheritance Principle:** If a property F is realized on a given occasion in virtue of the instantiation of one of its realizers, P, the causal powers of this instance of F are identical with the casual powers of P.” (Kim 1993, pp. 326-7)

And since a property’s being ontologically emergent requires that it contribute a *novel* causal power, no realized property can be ontologically emergent. But we see now the soundness of this argument depends on conceiving of realization along the lines of $R_{CAUS}$, and not $R_{THEO}$. If one takes ‘realization’ to denote $R_{THEO}$ and not $R_{CAUS}$, then the above argument is not sound.\(^9\)

Now, as this dissertation concerns the metaphysics of ontological emergence, it does not seem proper to leave $R_{THEO}$ formulated as I have—as a relation that is grounded largely in an epistemic notion of predictability. There have been numerous attempts to provide a metaphysical grounding for this relation, i.e., to analyze with purely metaphysical notions how $R_{THEO}$ is to be characterized. In the next section I take up this challenge. I will first review some of the analyses currently on offer, showing how they are inadequate. Then, I will suggest a formulation that I believe to be adequate.

### 4.2.2.2 Formulating $R_{THEO}$

Generally speaking, what philosophers have tried to do in formulating realization is to identify a realized property as a property whose essence wholly

---

\(^9\) Someone who exploits this distinction, although somewhat unknowingly, is Derk Pereboom (2002), who argues that the **Causal Inheritance Principle** is false and that X’s being realized is not inconsistent with X’s contributing novel causal powers. Ultimately, what Pereboom is suggesting is that realization be understood along the lines of $R_{THEO}$ and not $R_{CAUS}$. 

107
involves or consists in the instantiation of some other type of property. If a realized property can be characterized in such a fashion, then it is thought that the instantiation of the latter property necessitates and explains the instantiation of the realized property.

One of the first philosophers to pay significant attention to the metaphysics of the realization relation was Jeffrey Poland. Poland suggested that an instantiation of \( N \) is realized by instantiations of other properties \( R_1..R_n \) iff \( N \) “has a nature or essence which can be instantiated by the specific configuration of physical objects and attributes that results when the members of \( R \) are instantiated. The relevance of the members of \( R \) consists in their contributing to the constitution of \( N \) in this sense.” (1994, p. 17) But Poland leaves this notion of constitution more or less at the intuitive level. More recently, Jaegwon Kim has suggested that we characterize the realization relation (\( R_{\text{THEO}} \)) as the relation that holds between a second-order property over a given domain and the property in that domain that “satisfies” the second-order property. In doing so, Kim is attempting to give an analysis of \( R_{\text{THEO}} \). First of all, Kim characterizes a second-order property so:

\[
F \text{ is a second-order property over } D = df. F \text{ is the property of having some property } (P) \text{ in } D \text{ such that } C(P), \text{ where } C \text{ specifies a condition on members of } D. \text{ (2003b, p. 578)}
\]

Kim then adds that \( P \) is a realizer of \( F \) if and only if \( P \) satisfies \( C \). But this characterization of \( R_{\text{THEO}} \) is too narrow. It will exclude many properties that intuitively qualify as realized properties. For instance, the property of \textit{being a paperweight} should count as realized according to \( R_{\text{THEO}} \) if any property does. It is a property that strongly supervenes on microphysical properties and, intuitively, it is robustly explainable from the microphysical—anyone that has a complete knowledge
of the microphysical facts and laws and understands the concept of *being a paperweight* (an object that causes certain such paper molecules to remain stationary under certain conditions) will be able to theoretically predict when *being a paperweight* is instantiated. But *being a paperweight* is not a second-order property as defined above. To see this, note that the type of property that would intuitively qualify as a realizer of *being a paperweight* is a structural property, a property which is a member of the class of properties all characterized by their having such-and-such parts standing in such-and-such relations. Suppose P is one of these structural properties. It is by having P that condition C (possessing a certain causal role) is satisfied. But then the second-order property F over the domain of structural properties would be the *property of having property P*. But that’s not the same as the property of *being a paperweight*. We can’t identify *being a paperweight* with the *property of having property P* because then every object that was a paperweight would have to have P, i.e. would have to have the same microstructural property. But that’s clearly false—paperweights can have all types of different structures.

Interestingly, right after giving his characterization of realization in terms of second-order properties, Kim writes: “second order properties can have multiple realizers.” (2003b, p. 579) But as I’ve just shown this is false according to his analysis of realization, and therein lies the rub. If a realized property is just the *having of some other property P*, then it can have only one realizer—P.

But perhaps what Kim meant to say is that being a second-order property over D is having *some property or another* in D such that the property in D had meets condition C. If so, there is still a problem, for this is ambiguous between two readings:
(a) F is a second-order property over D = df. F is the property of having P1 or F is the property of having P2 or F is the property of having Pn, where for any Px such that Px is a member of P1…Pn, Px is such that C(Px), where specifies a condition on members of D.

(b) F is a second-order property over D = df. F is the property of having (P1 v P2 v….v….Pn), where for any Px such that Px is a member of P1…Pn, Px is such that C(Px), where specifies a condition on members of D.

I don’t think we can take option (a) seriously—it would imply that in one situation F is identical to the property of having P1 while in another it is identical to the property of having P2. Unless we embrace some kind of relative identity or somehow justify the identification of the property of having P1 and the property of having P2 (which I don’t see how that could be done), then (a) cannot be correct. But (b) is a genuine possibility. It would require accepting the existence of disjunctive properties, which may not be an issue in and of itself, but it does lead to a rather unfortunate explosion of realizers. For once we admit the existence of the disjunctive property P1 v P2, it satisfies C just as well as P1 does. So if P1 realizes C, so does P1 v P2. And so does P1 v P2 v P3. And so does P1 v the property of being a monkey’s uncle. But perhaps this would not faze Kim; however, it seems to me sufficient enough grounds to look for another characterization of realization.

Where should we turn? Kim’s characterization is too narrow, so we need a characterization that will insure properties like being a paperweight are realized.

Andrew Melnyk has set forth the following characterization of R_{THEO}, that, barring a couple of significant, yet fixable, technical problems, is just about right. Let’s call this

On-the-Right-Track Formulation of R_{THEO}:

Token x realizes token y iff (i) y is a token of some functional type, F, such that, necessarily, F is tokened iff there is a token of some or other type that meets condition, C; (ii) x is a token of some type that in fact meets C; and (iii)
the token of $F$ whose existence is logically guaranteed by the holding of condition (ii) is numerically identical with $y$. (2003, p. 21)

Melnyk intends **On-the-Right-Track-Formulation of $R_{THEO}$** to encompass not just properties, but objects and events as well, thus he speaks of “token $x$” and “token $y$”, but for our purposes we can just understand “token $x$” to mean “property instance $x$”. Also, Melnyk does not restrict condition $C$ to being a condition that specifies a causal role, as is normally associated with functional properties. On this understanding of functional type, any property that is necessarily such that it is instantiated iff there is some other condition that is met qualifies as a functional type. So, biological types turn out to be functional types on Melnyk’s view, regardless of whether they can be defined according to a functional role (as long as they, as Melnyk says, are nomologically necessitated by “lower-order” types). (Melynk 2003, p. 38) And it also seems that determinable properties would turn out to be functional types—the property of *being colored* is such that it is tokened iff there is a token of some other type that meets condition $C$ (iff there is a token of a color that meets the condition of being instantiated), and *being red* realizes *being colored*. Now one may happen to think it a bit odd that Melnyk refers to all these as functional types—but that is merely a terminological issue. If being a “functional type” conjures up images of causal roles for you, then you could think of Melnyk’s functional types simply as “realized types”.

**On-the-Right-Track-Formulation of $R_{THEO}$** is advantageous, and an improvement upon Kim’s in that it includes properties like *being a paperweight* as realized. But it still suffers from a few problems, though ones I think can be (relatively) easily remedied. The first problem it suffers from is the same problem we ascribed to **MoSS (Modal-Operator Strong Supervenience)** that disqualified that
relation from being a determination relation—it is not asymmetric. A token of the property *being triangular* realizes a token of the property *being trilateral*, and vice versa. But intuitively these property tokens do not realize each other, in that neither determines that the other is instantiated, that role being reserved for the instantiations of “lower-level” microphysical properties.

To see why *being triangular* qualifies as a realizer of *being trilateral* given On-the-Right-Track-Formulation of $R_{\text{THEO}}$, let $F = \text{*being triangular*}$, $y = \text{a token of *being triangular*}$, $x = \text{a token of *being trilateral*}$, and $C = \text{being a token of *being trilateral*}$. Given this, $y$ is a token of some functional type $F$, which Melnyk defines as a “higher-order type such that, necessarily, it is tokened iff there is a token of some or other lower-order type that plays some particular role or—more generally—meets some particular condition.” (2003, p. 20) The lower-order type in question here would be some structural property—*being triangular* is tokened iff a structural property that plays the role of ‘constituting a triangle’ is instantiated. So $F$ (\text{*being triangular*}) is a functional type. And $F$ is such that it is tokened iff there is a token of some or other type that meets condition, $C$. *Being triangular* is tokened iff *being trilateral* is. So condition (i) is satisfied. Condition (ii) is also obviously satisfied, for a token of *being triangular* trivially meets the condition of being a token of *being triangular*. And (iii) is also satisfied.\(^9\)

This failure of On-the-Right-Track-Formulation of $R_{\text{THEO}}$ can be remedied by including a condition that insures realization is asymmetric. This could be done

\(^9\) Note that *being triangular* will also be realized by whatever lower-order structural type it is on this occasion that insures an instance of *being triangular* is present. So Melnyk’s characterization gets it right in that respect, it just allows another property to qualify as a realizer that intuitively should not.
simply by adding condition insuring that token y does not realize token x. Here is how an amended analysis of $R_{THEO}$ might look (with my amendments in italics, “token” changed to “property instance”, and “tokened” changed to instantiated). Let’s call this

**Very-Nearly-the-Right-Formulation-of-$R_{THEO}$:**

Property instance x realizes property instance y iff (i) y is a property instance of some functional type, F, such that, necessarily, F is instantiated iff there is an instance of some or other type that meets condition, C; (ii) x is a property instance of some type that in fact meets C; and (iii) the instance of F whose existence is logically guaranteed by the holding of condition (ii) is numerically identical with y, and (iv) either x is not an instance of some functional type, G, such that, necessarily, G is instantiated iff there is an instance of some or other type that meets condition, D, or y is not an instance of some type that meets D.

But **Very-Nearly-the-Right-Formulation-of-$R_{THEO}$** is still not quite right as an analysis of $R_{THEO}$ (as you may have guessed by the way I named it). This is because this definition does not exclude realization from being a diachronic, causal relation. Suppose that x at t1 causes y at t2. Also suppose y is an instance of some functional property F. This means (according to Melnyk’s definition of a functional property) that F is such that, necessarily, F is instantiated iff there is an instance of some or other lower-level property that meets condition C*. Furthermore, suppose x is not an instance of a functional type (perhaps x is just an instance of some fundamental, microphysical property). Finally, let condition C = “being a cause of F or being a lower-level property that meets condition C*.” Given all this, (i) is satisfied because the second disjunct of C is just the definition of what it is to be F, (ii) is satisfied because x is an instance of a cause of F, and thus also meets condition C, (iii) is obviously satisfied and (iv) is satisfied because x is not a token of a functional property.
This problem could be remedied by placing a restriction that $x$ and $y$ must exist at the same time. But that would only be a band-aid on a greater problem that this example exposes. The problem is that unless some kind of restriction is placed on condition $C$, then any property instance $x$ whatsoever will qualify as a realizer of any instance of a functional type in worlds in which $x$ exists. For if $C = \text{“being an instance of a property } Q \text{ or being an instance of a lower-level property that meets condition } C^*\text{”}$, where necessarily, $F$ is instantiated iff there is an instance of some or other lower-level property that meets condition $C^*$, then if $q$ is an instance of $Q$ in a world $W$ at which $y$ is a token of $F$, $q$ realizes $y$, for all the same reasons that $x$ was a realizer of $y$ in the example of the previous paragraph. And since $Q$ could be any property that is capable of being instantiated, nearly any property could qualify as a realizer. But this can’t be right.

What this shows is that some restrictions need to be placed on $C$ to make \textbf{Very-Nearly-the-Right-Formulation-of-R_{THEO}} a truly adequate account of $R_{THEO}$. These restrictions will include (but probably not be limited to) $C$’s being non-disjunctive, and being such that a disjunctive property type could satisfy it. The restrictions will also have to include that the type that satisfies condition $C$ be an intrinsic property, otherwise, for example, an instance of the property \textit{being such that property instance } $y$ \textit{is realized} (a property that everything has in worlds in which $y$ is realized) will also qualify as a realizer of $y$, which it of course should not. I will not attempt to give an exhaustive list of such conditions here (though my intuitions tell me that the ones I’ve just listed may be the only ones required, I fear that is probably not the case). However, the point is that such restrictions on $C$ could be provided, and
once the proper restrictions are placed on C, then **Very-Nearly-the-Right-Formulation-of-R\_THEO**, I believe, would be an adequate account of R\_THEO.

The main lesson to be learned here is that two property instances x and y (of properties P and Q respectively) bear the R\_THEO relation to each other iff the very essence of P is such that it is instantiated when and only when an instance of Q is instantiated. If that is the case, and one truly grasps the essence of P and knows all the facts about Q-instantiations, then one will be able to theoretically predict that P is instantiated. **Very-Nearly-the-Right-Formulation-of-R\_THEO**, with C suitably restricted, provides the metaphysical basis for this epistemic notion.

4.2.2.3 Formulating R\_CAUS

We now turn to analyzing R\_CAUS. Coming up with a formulation of R\_CAUS that will satisfy everyone in the debate is an even more difficult task than formulating R\_THEO, for even though all those who have talked about the relations will ascent to the general truth that a realized property “inherits all of its causal powers” from another property or properties, there is significant disagreement about which properties it inherits its causal powers from. In other words, the disagreement centers around whether R\_CAUS is a relation that holds between just two entities, or between numerous entities. Sydney Shoemaker, for example, explains R\_CAUS as follows: “property X realizes property Y just in case the conditional powers bestowed by Y are a subset of the conditional powers bestowed by X...”

But as Carl Gillett notes, the implicit

---

94 Shoemaker (2001): 78. By conditional powers here Shoemaker means causal powers. The full quote from Shoemaker is actually “property X realizes property Y just in case the conditional powers bestowed by Y are a subset of the conditional powers bestowed by X (and X is not a
assumption behind this characterization of $R_{CAUS}$ is that a realized property and its realizer are instantiated by the same individual. Gillett contends that this assumption flies in the face of what scientific investigation has revealed, namely, property instances of a composite object are realized by the properties of that object’s *parts*, not by other properties of the object itself. For instance, an instance of the property of *being hard* exhibited in a diamond bears $R_{CAUS}$ to instantiations of the properties (and relations) of the diamond’s parts—it is the microphysical properties and relations of the diamond’s parts that not only explains the appearance of *being hard*, but also endows that property with its causal powers. (2002a, p. 319) This leads Gillett to proffer the following formulation of $R_{CAUS}$, which he calls the Dimensioned View of Realization:

**Dimensioned View of Realization (DVR):** Property/relation instance(s) $F_1$-$F_n$ realize an instance of a property $G$, in an individual $s$, iff $s$ has powers that are individuative of an instance of $G$ in virtue of the powers contributed by $F_1$-$F_n$ to $s$ or $s$’s constituents, but not vice versa. (2002a, p. 322)

DVR is a bit disappointing in that it still includes the ‘in virtue of’ locution that we were hoping an analysis of realization would explain. But Gillett does say this about a property’s being realized according to DVR: a property instance $G$’s being realized according to DVR implies that “we can account for all the causal powers of $G$ simply using the contributions of powers by the microphysical properties of these individuals, or their constituents.” (2002a, p. 319) In other words, Gillett takes it that a property instance being realized entails that it does not contribute any novel causal powers—there is nothing that this instance of $G$ causes that is not caused by the instances of the microphysical properties that realize it. So DVR implies, as does

conjunctive property having $Y$ as a conjunct).” I’ve left out the parenthetical in the main text just for purposes of simplification.
Shoemaker’s explanation of realization, that realized properties do not contribute novel causal powers.

So I think we can conclude at least that a necessary condition on \( R_{CAUS} \) is that any property that qualifies as being realized vis-à-vis \( R_{CAUS} \) not contribute any novel causal powers according to **Novel Causal Power Contribution**. And for my purposes here in this dissertation, that rather rudimentary insight into \( R_{CAUS} \) will be sufficient. A more thorough analysis of \( R_{CAUS} \) will have to wait for some other time.

This concludes my discussion of the two realization relations that have been put to work in the contemporary debate. As we shall see shortly, distinguishing these two relations allows us to distinguish a version of OE that until now has not been recognized as a genuine possibility. This will be reviewed in the final section of this chapter.

### 4.2.3 Causation

The last relation of microphysical determination that has been advanced as the one holding between microphysical properties and ontologically emergent properties is causation. I have nothing significant to say here about the causation relation—those who employ the idea that emergent properties are causally dependent upon microphysical properties intend it to be a rather straightforward idea. Whatever theory one has of causation, whatever relation two property instances \( P \) and \( P^* \) must bear to each other (or to certain events, facts, etc) in order to say that the instantiation of \( P \) causes an instantiation of \( P^* \), it is this same relation that holds between an emergent property and the microphysical property (or properties) upon which it is dependent. The most prominent defender of the claim that emergent properties are causally
dependent upon microphysical properties is Timothy O’Connor, and in the final section of this chapter I will have a bit more to say about O’Connor’s specific theory of OE.

4.3 An Alternate Interpretation of Causal Efficacy and Autonomy

Up until this point the only interpretation of a property’s being causally efficacious I have discussed is that it contribute a causal power, and the only understanding of a property’s being causally autonomous I have discussed is that it contribute a novel causal power. But there is another condition that has been proposed in the literature as being at least sufficient for a property’s being causally efficacious, one that has been championed most prominently by Carl Gillett. By my lights, such a condition is also a sufficient condition on a property’s being causally autonomous.95 I will here just briefly outline the basic contours of this view—Chapter Six of the dissertation is devoted to a more detailed analysis and evaluation of the prospects of building a theory of OE around such an interpretation of causal efficacy and autonomy.

According to Gillett, a property is causally efficacious if it synchronically determines (or partially determines) that a distinct property contributes a causal power. Let us call this allegedly sufficient condition on a property’s being causally efficacious

95 Gillett does not discuss whether the condition is sufficient for a property’s being causally autonomous as he does not explicitly make the distinction between causal efficacy and causal autonomy as I do. It is a bit difficult to say, but it appears that Gillett might take causal efficacy as he understands the notion to include causal autonomy as I understand the notion. If so, then he does in fact agree with me here.
Synchronic Causal Power Determination, or SCPD. An advocate of SCPD says that a property Q’s being causally efficacious does not require that an instance of Q produce an effect, does not require that an instance of Q provide the causal “oomph” so to speak, but merely that the presence of Q synchronically determines that a some other property instance is an “oomph” giving property instance, i.e. that it contributes a causal power by which it causes an effect.

Allow me to illustrate with an example from Gillett. According to Gillett, certain realized properties are causally efficacious in that they synchronically (and non-causally according to Gillett) determine the causal powers contributed by at least one of the fundamental microphysical properties that realize it. (2002c, p. 115) The hardness of a diamond may be one such property. Suppose a diamond has the property H of being hard, and this causes it to have the causal power C, which is the causal power to scratch glass in circumstances K. H is a realized property, realized by the properties of and relations between the microphysical entities that compose the diamond. Thus, H contributes the causal powers it does in virtue of the causal powers contributed by the microphysical parts of the diamond, and thus H fails to contribute a novel causal power. Now suppose P is a property of one of the microphysical entities composing the diamond (call this microphysical entity ‘a1’). Gillett reasons that an instance of P may possess certain causal powers that it contributes only upon realizing

---

96 Gillett actually refers to this type of determination as non-causal determination. But I will not use this moniker because I don’t want to take a stand on whether such a relation is causal or not. Since it is synchronous, there are prima facie reasons that it is not causal, but as we shall see in Chapter Five, at least one philosopher, E.J. Lowe, posits a kind of determination very much like what Gillett is suggesting, yet refers to it as a type of causal determination. It will be beneficial to simply group these two views together under the heading Synchronic Causal Power Determination.

97 I am only speculating here, however, as Gillett himself never offers an example of a property he thinks would qualify as ontologically emergent according to his view.
$H$. Suppose, for instance, upon realizing H and upon realizing H alone, an instance of 
P causes (or contributes to causing) the instantiation of a distinct microphysical event 
E (perhaps the movement of some microphysical particle that would not occur unless 
P also was one of the properties that realize the property of being hard). Gillett claims 
that in such a situation the instance of H is exerting a non-causal determinative 
influence on the powers of this instance of P. It is because this instance of P 
(partially) realizes this instance of H that it contributes the causal power (call it C*) to 
cause (or partially cause) Q. As Gillett says, “H is a necessary member of the 
properties which are only jointly sufficient for determining the contribution of C* to 
a1. There is thus prima facie reason to believe, in this situation…that the realized 
property H is a causally efficacious property, since it partially determines the 
contribution of a power to an individual.” (2002b, p. 95) The instance of P is the 
sufficient cause of the instance of Q, but H non-causally determines that a sufficient 
cause of Q (namely this instance of P) has the causal power that it does. This is 
enough, claims Gillett, for us to say that the instance of H is causally efficacious with 
respect to e.

Would a property that qualifies as causally efficacious via this condition also 
qualify as causally autonomous. It certainly seems so in that were the situation Gillett 
describes to obtain the causal efficacy a property like H would have would be non-
derivative. The instance of H certainly owes its existence to the property instances 
that realized it, but this alone of course should not undermine its causal autonomy. 
Every property instance owes its existence to another property instance in some sense 
(whether it be that it is caused to exist by another or is realized by another), but that 
does not imply that no property is causally autonomous. So there is no prima facie
reason to suppose H would not be causally autonomous. I will have more to say about this and many other questions regarding SCPD in Chapter Six. For now, I trust this example illuminates well enough the idea behind SCPD, well enough at least that I can proceed to the final section of this chapter, wherein I present a taxonomy of possible theories of OE.

4.4 A Taxonomy of Possible Theories of Ontological Emergence

I have identified four different candidates for the microphysical determination relation that holds between emergent properties and their microphysical base properties, and two different sufficient conditions for causal efficacy and autonomy. This gives us the matrix of possibilities for theories of OE expressed in Table 1, with dependence relations running down the side, and options for causal efficacy running across the top.

TABLE 1:
MATRIX OF POSSIBILITIES FOR THEORIES OF OE

<table>
<thead>
<tr>
<th></th>
<th>Novel Causal Powers</th>
<th>SCPD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mere Supervenience</td>
<td>I</td>
<td>V</td>
</tr>
<tr>
<td>( R_{THEO} )</td>
<td>II</td>
<td>VI</td>
</tr>
<tr>
<td>( R_{CAUS} )</td>
<td>III</td>
<td>VII</td>
</tr>
<tr>
<td>Causation</td>
<td>IV</td>
<td>VIII</td>
</tr>
</tbody>
</table>

4.4.1 Theories Incorporating SCPD (Boxes V, VI, VIII, and VIII)

Options V, VI, VIII, and VIII all involve the notion of SCPD, and, as I have mentioned, I will be discussing this relation as well as one particular versions of OE that has been constructed around SCPD in Chapter Six, a theory that falls in to box VII. Unfortunately, space does not permit to discuss other possible theories of OE.
incorporating SCPD in this dissertation. However, another individual who has suggested a theory along these lines that may be worth exploring is E.J. Lowe. Lowe suggests that one way of preserving the causal efficacy of the mental would be to suppose that in certain cases a mental property instance (what he calls a mental event) brings it about that certain microphysical events cause each other. Lowe is careful to note that what he is suggesting here is not that a mental event cause a microphysical event (say P1), which in turn causes a subsequent and distinct microphysical event (say, P2), but rather that the mental event actually causes the fact that P1 causes P2 to be true. This would be an instance of what Lowe calls fact causation, and not event causation. Now one way of interpreting Lowe here (or, if not an interpretation of Lowe, perhaps this would simply be a subtle variation on his view) is as suggesting that the mental event synchronically determines that an instance of causation (an instance of P1 causing P2) is present, and this way of looking at the situation at least resembles the idea of SCPD. In fact, in Chapter Six, I suggest that one way of interpreting Gillett’s claim that emergent properties synchronically determine the causal powers of microphysical properties (and perhaps the only way of interpreting this claim that gives his account of OE a chance of succeeding) is precisely as a suggestion that instances of special science properties are causally efficacious by determining that instances of causation between microphysical entities occur.

But I shall set all that aside until Chapter Six. For now, what about the boxes appearing in the middle column? Let me finish off this chapter by saying just a little bit about each, beginning with the box that is the most straightforward to address.
4.4.2 Box III: $R_{CAUS}$ Plus Novel Causal Powers

Box III is not a live option in that any theory falling in the box would be self-contradictory. Recall that a necessary condition on a relation’s being an instance of $R_{CAUS}$ is that any property that qualifies as being realized vis-à-vis $R_{CAUS}$ not contribute any novel causal powers according to Novel Causal Power Contribution. So obviously it is not possible for a property to be causally autonomous by contributing a novel causal power according to Novel Causal Power Contribution and still bear $R_{CAUS}$ to microphysical properties.

4.4.3 Box II: $R_{THEO}$ Plus Novel Causal Powers

Theories that appear in Box II will not suffer from the problem that befalls the theories of Box III, and here is where distinguishing the two different realization relations sharpens our understanding of the relationship between realization and emergence. The compatibility of realization and ontological emergence has been consistently dismissed in the literature, but in virtually every case in which the compatibility of these two views is addressed, the type of realization that is being imagined is $R_{CAUS}$ and not $R_{THEO}$. In the next chapter we will look at an account of special-science properties that implies that such properties are causally autonomous and realized along the lines of $R_{THEO}$, namely Derk Pereboom’s “Robust Non-Reductive Physicalism” and ask the questions, first, whether this “box II” theory is internally consistent and, second, whether there are good reasons for believing it.
4.4.4 Box IV: Causation plus Novel Causal Powers

The philosopher who provides the most detailed exposition and defense of the claim that emergent properties are causally dependent upon microphysical properties is Timothy O’Connor. O’Connor has authored several items on this subject, but it is his most recent work, co-authored with Hong Yu Wong, on which I shall focus. O’Connor and Wong give their reasoning for taking up this position as follows:

Emergent properties are basic properties, token-distinct in character and propensity from any microphysical structural properties of their bearers. If their appearance in certain systems is to be explained at all, they must be explained in terms of a causal, not purely formal, relationship to underlying, preceding structures. (O’Connor and Wong, forthcoming)

The authors theorize that upon the instantiation of a certain complex microstructural property, an emergent property E will be caused (diachronically). E will possess causal powers that no microphysical or microstructural property has and will subsequently cause the instantiation of other microphysical properties, or perhaps other emergent properties. As it turns out, on the specifics of O’Connor and Wong’s account of emergent properties, emergent properties fail to strongly supervene, or even globally supervene, on microphysical properties. Their view also entails the denial of Microphysical Causal Closure. In this sense, it is one of the most robust, and most controversial, characterizations of ontological emergence. Though in this dissertation I will not be discussing the specifics of the O’Connor/Wong account in any detail, allow me here o make a couple of observations about the possibility that the microphysical determination relation holding between microphysical properties and ontologically emergent properties is one of causal dependence.

As I mentioned in Chapter Two, the British Emergentists dismissed the possibility that emergent properties were causally dependent on microphysical
properties, but did so without giving much of a reason. However, that does not mean there are no reasons. The main objection to O’Connor’s view is one that he and co-author Hong Yu Wong have rather eloquently formulated themselves, and that they refer to as the “Latent Microphysical Disposition Objection.” In actuality, this turns out to be an objection that threatens not just a causal emergence theory like that of O’Connor but all theories of ontological emergence, all of which entail that ontologically emergent properties are necessitated by microphysical properties.

O’Connor and Wong’s expression of the objection is as follows:

Emergent hypotheses assert an unnecessary middle term. They suppose that complex microphysical configurations jointly cause the occurrence of an emergent property, which in turn has a causal effect on the evolution of the physical state of a system. Yet suppose we were actually to observe such discontinuous patterns of effects within complex systems. It would be simpler to suppose in consequence that the microphysical entities have otherwise latent dispositions towards effects which are continually manifested in (nearly) all contexts, and the observed difference is a result of the manifestation of these latent dispositions. All of what happens, we could continue to suppose, is a product of the elementary constituents of matter, though the range of microphysical dispositions would be richer than one might have supposed through observation of relatively simple particle interactions alone. (O’Connor and Wong, forthcoming)

The objection here is really one based on simplicity considerations. On O’Connor’s view, a group of microphysical properties P1…Pn cause the instantiation of a “higher-level” emergent property M, which then “reaches back down” and causes the instantiation of a different microphysical property P2. The objector is basically saying this: would it not be simpler to attribute to each of P1…Pn “latent microphysical dispositions”, or using the terminology of this dissertation “latent microphysical causal powers”, that are only activated, so to speak, when P1…Pn are instantiated in some complex pattern of relation to each other? In other words, in explaining an effect E of an alleged emergent property M, it is always simpler to attribute (perhaps heretofore
unknown) series of causal powers of a previously recognized type (i.e. a microphysical causal powers) to P1…Pn rather than a causal power C(K*, E) of a brand new type (an emergent type) to a different kind of property E, an emergent property.

But this objection generalizes. This objection, if it has any force, also has force against theories of OE that take the microphysical determination relation to be one of mere supervenience or of realization. What should be said in response to this objection? O’Connor and Wong’s response is simply to deny that attributing a new microphysical causal power C(K,E) to P1 is a simpler solution than positing the existence of several “new” latent microphysical causal powers. They write “our view is simpler and so to be preferred: just one basic disposition—to give rise to an emergent property under suitable circumstances—that is had by each fundamental constituent (i.e. property)” and thus not a different causal power had by each microphysical property. But this response is treading on thin ice, for I imagine most people’s intuitions are going to point in the other direction, i.e., toward the view that all else being equal, it is always simpler to posit more instances of a certain type we already know to exist rather than an instance of a brand new type that, according to the theories we currently have, we have no reason to believe exist.

This is why I think the response that O’Connor and Wong should give is that all other things are not equal. Their response should be that there are independent reasons for believing in the existence of emergent properties beyond considerations of causal power contribution, or simplicity, or the like. What might such reasons be? Well, as I mentioned in Chapter One, though we might be persuaded into accepting that a predicate like ‘having electronegativity of 3.44’ simply refers to some complex
microphysical property or set of microphysical dispositions, there seems to be strong reason for believing that a predicate like ‘believing that the Cubs will eventually win the World Series’ refers to something much different—to a feature of me that is not identical to nor reducible to any microphysical property or disposition. O’Connor and Wong themselves, in a separate section of their paper than the one in which they discuss their response to the Latent Microphysical Disposition objection, imply something very similar to what I have just said:

A person’s experiences and other conscious mental states—states which manifestly influence our behavior—exhibit features quite unlike those of physical objects, whether as revealed in ordinary sense perception or as uncovered in the physical or biological sciences. And the maximally direct nature of our first-person awareness of these conscious states precludes the a posteriori ascription to them of underlying physical micro-structure hidden to introspection. (O’Connor and Wong, forthcoming)

And this seems to me the exact sort of response that not only O’Connor and Wong but anyone who argues for the existence of emergent properties should give to the Latent Microphysical Disposition objection. Yes, it might be the case that all other things being equal, it is simpler to posit latent microphysical causal powers of microphysical entities than to suppose that genuinely emergent properties exist. But evidence from our first person experience and what we believe about our own conscious mental states suggests that there are certain properties whose nature is not exhausted by their causal powers, whose nature includes a kind of raw feel or experience that no microphysical property has. Now, I’m not saying that this is a sound argument for the existence of emergent properties, but it is probably the best one out there. The main point I hope to be making here is that none of the theories of OE being put forth house arguments for the existence of emergent properties, but merely try to account for how such properties
could exist. Why we should believe they exist is a question largely independent from the metaphysical theorizing behind the different accounts of OE.

That being said I believe anyone who believes in the existence of ontologically emergent properties does so (or should do so) on grounds independent of those reasons O’Connor and Wong give for thinking that their particular theory of OE is true. And the person who believes in the existence of ontologically emergent properties already has the resources for responding to the Latent Microphysical Disposition objection.

4.4.5  Box I: Mere Supervenience plus Novel Causal Powers

This brings us to Box I. Box I is the box in which the conception of an ontologically emergent property expressed by the British Emergentists (COE) and the one delineated by Kim (K-COE) most naturally fit. Most who have defended such a conception have settled on the fact that the way to avoid Kim’s Challenge is to deny Microphysical Causal Closure. This includes the likes of Roger Sperry (1980, 1991) and Karl Popper and John Eccles (1977), all of whom had been discussing theories of emergence long before the recent explosion of interest in emergence began, as well as more recent entrants to the discussion like O’Connor (1994) and Hasker (1999).

What are the prospects for theories of OE that entail the denial of Microphysical Causal Closure? I don’t think the prospects are as bleak as some people make them out to be. Though there have been arguments advanced for the

---

98 O’Connor (1994) and Hasker (1999) both contain fairly rigorous and enlightening discussions of the work of Sperry and Popper, and I highly recommend these to the reader who might be interested in learning more about this particular type of emergentist theory. Also, Brian McLaughlin (1997) and Jaegwon Kim (1999 and 2003b) have quite a bit to say about the relationship between COE, or K-COE, and the denial of Microphysical Causal Closure.
principle, I don’t find any of them to be anything close to a demonstration that the principle is true. It still seems to me that accepting the principle comes down to a matter of expediency or inference to the best explanation—given that nearly every microphysical event we have investigated has a microphysical cause, it’s likely that 

**Microphysical Causal Closure** is true. But if one has strong reasons to think, on independent grounds, that there are better reasons to believe in non-reducible, causally autonomous special-science properties, and that such a position requires the denial of **Microphysical Causal Closure**, considerations of expediency or inference to the best explanation do not look as strong. Why could it not be, for instance, that though every microphysical event has at least a partial microphysical cause, perhaps some microphysical events are only caused jointly by an instance of a microphysical property and an emergent special-science property (the most likely candidate here being mental properties). No amount of information about the microphysical facts of the world could distinguish this possibility from a situation under which **Microphysical Causal Closure** holds. And while one may charge that it is only the obscurantist who would seriously consider such a possibility, such a charge is sheerly rhetorical. There are some well-developed theories of how such a situation might obtain. The main point is that it is not at all obvious that **Microphysical Causal Closure** holds, and theories that deny the thesis are thus still very much in play.

---

99 For example, Papineau (1993) and Loewer (2001).

100 See for instance Lowe (2000). He suggests that our world may be such that though every microphysical event P has a completely sufficient immediate microphysical cause, there may be some immediate causes which are the conjunction of two microphysical events (say P1 and P2) that, given all the study of microphysics has to tell us, appear independent, but in actuality linked by some mental event. This mental event links the occurrence P2 with a microphysical causal ancestor of P1. P2 then may have a microphysical cause and a mental cause, but the mental cause would be “invisible” from the point of view of physical science. I realize the details of this example may be hard to follow, but the
4.5 Conclusion

This concludes this chapter on the metaphysical issues that arise from different understandings of microphysical determination, causal efficacy, and causal autonomy. In the remaining two chapters, I will examine in more detail specific accounts of ontologically emergent properties that have been suggested in the recent literature. Though there are many such accounts to choose from, I will focus on accounts that have been constructed with the goal of being consistent with physicalism. This is not because I think physicalism is true, nor because I necessarily believe these are the best accounts of OE nor the ones that have the best chance to succeed, but simply because I find the project of making ontological emergence compatible with physicalism to be an intriguing one, and one that would perhaps be of most interest to contemporary philosophers.

---

101 However, theories that fit into Box I have been developed that do not entail the denial of **Microphysical Causal Closure**. For instance, Sydney Shoemaker (2001) has proposed an understanding of an ontologically emergent property according to which ontologically emergent properties are actually a type of higher-level structural property. On Shoemaker’s view, while closure still holds at the microphysical level, it is still possible that a property be emergent by contributing causal powers that no other structural property contribute (even though the causal effects of such a property may be co-extensive with the causal effects of its constituent microphysical properties). However, Shoemaker’s view is beset with many difficulties, not the least of which is that on his own theory of properties, the causal theory of properties, emergent properties as Shoemaker describes them turn out to be identical with non-emergent properties—thus one would have to reject the causal theory of properties to embrace Shoemaker’s insights on emergence. I’m actually not sure about this—there may be another way to develop an account of emergence along the lines of Shoemaker. Those interested can see Shrader (2004).
CHAPTER 5

PEREBOOM’S ROBUST NON-REDUCTIVE PHYSICALISM

5.1 Introduction

In this chapter I examine in detail Derk Pereboom’s “Robust Non-Reductive Physicalism” (hereafter, R-NRP), an account of special-science properties and their relation to microphysical properties consistent both with MOE and with the conjunction of Microphysical Causal Closure and Universal Supervenience, two doctrines thought to be entailed by physicalism. R-NRP also incorporates the idea that instances of special-science properties are determined by instances of microphysical properties in that they bear $R_{\text{THEO}}$ to these latter property instances. All of these features of R-NRP make it a species of MOE that is much more attractive to physicalists than theories contained in Box I or Box IV of our taxonomy, all of which will generally entail the denial of Microphysical Causal Closure or Universal Supervenience, or both.

This is one reason I have singled out Pereboom’s view for discussion. The other is Pereboom’s commitment to exploring the metaphysical issues that confront Non-Reductive Physicalism. Like so many philosophers today, Pereboom sits squarely in the non-reductionist camp, claiming that at least some special-science predicates refer to

---

102 Pereboom (2002). Page references in the text and footnotes are all to Pereboom’s 2002 article, unless otherwise noted.
properties ontologically irreducible to the fundamental microphysical properties. But unlike many non-reductive physicalists, Pereboom seems to appreciate that it is one thing to say that there are irreducible special science properties, and another thing to actually present an in depth account of such properties that makes good metaphysical sense. Where Pereboom parts company with the proponent of other types of NRP (e.g., advocates of qualified micro-determinism or token physicalism), is in his contention that these irreducible properties, beyond merely playing a role in interesting regularities and novel explanations, contribute novel causal powers and are thereby causally efficacious and causally autonomous. According to Pereboom, the causal powers that special science-properties (e.g., mental properties) contribute are, in general, powers to bring about other special-science properties, not microphysical properties. However, Pereboom thinks he can square the idea of a special science property (in Pereboom’s case, he specifically addresses the issue of mental properties) contributing a novel causal power with such a property’s being realized (or as he puts it “constituted by”) microphysical properties and with the principle of Microphysical Causal Closure’s being true. Pereboom preferred way of escaping the conclusion of Kim’s Challenge is to deny the Principle of Downward Causation. Whether such a denial is plausible remains to be seen, and will be one issue addressed in my critical discussion of his view at the end of this chapter.

So R-NRP can be placed in Box II of our taxonomy. It is particularly interesting (and I believe unique) in that it is a thoroughly physicalist species of MOE that entails the

---

103 Pereboom only discusses his theory as it applies to the predicates of psychology. He never explicitly denies that what he says about psychological predicates, properties, and causal powers could be applied to other domains, but he never explicitly endorses it either.
falsity of a claim most people in the current debate accept rather uncritically, namely the **Principle of Downward Causation**. But another highly interesting aspect of R-NRP is the metaphysical machinery that drives the view. Pereboom relies heavily on the concept of a token causal power, which he employs as a central notion in his argument for the nonreducibility of mental causal powers. Another interesting metaphysical notion contained in R-NRP is that of property instance (or event) constitution. According to Pereboom, certain mental property instances are *wholly constituted* by microphysical property instances. Though Pereboom is not very forthcoming vis-à-vis the nature of this constitution relation, he does suggest it is a species of realization, and the way he describes mental properties suggests, or so I shall argue, that the relation between a mental property and its realizer is that of a determinable property to one of its determinates. That Pereboom understands constitution in this way is intriguing in that other philosophers who have invoked the idea that a mental property is related to its physical realizer as determinable to determinate have generally done so in support of theories of qualified micro-determinism.\(^{104}\) Qualified micro-determinism, you may recall from Chapter One, is the view that special-science properties, though causally efficacious, are not causally autonomous. That R-NRP entails that realized, determinable properties can be causally autonomous makes Pereboom’s theory, as far as I can tell, a unique candidate in the plethora of potential solutions to the Problem of the Special Sciences.

It is, however, a bit ironic that I have chosen R-NRP as a featured position of this dissertation in that Pereboom *explicitly denies* that his theory is an emergentist one. He

\(^{104}\) Yablo (1992) and Wilson (1999) come to mind. Ehring (1996) criticizes the notion that mental properties stand to their physical realizers in the relation of determinable to determinate.
actually refers to emergentism as a “metaphysically extravagant” view. (2002, p. 504) But in denying that R-NRP is an example of emergence, he is conceiving of emergence along the lines of COE (most likely K-COE), and not MOE. Pereboom claims that Emergentism entails both Emergent Property Unpredictability and the denial of Microphysical Causal Closure. But I hope it is abundantly clear to the reader at this point why MOE entails neither of these principles.

So in Pereboom’s R-NRP we have a unique, intriguing, and metaphysically rich example of MOE. But are there good reasons to believe the view is true? The main goal of this chapter is to examine that question and offer up my reasons for thinking that the answer is, most probably, “no”. Despite this assessment I do not think the view is obviously false, and I certainly find it to be one of the most promising physicalist proposals for remedying the Problem of the Special Sciences. My reasons for thinking R-NRP is false ultimately flow from my acceptance of a couple of admittedly

105 Here is some more detail on the reasons Pereboom gives for thinking R-NRP is not a species of Emergentism. Pereboom claims that emergentism entails that there are certain emergent properties that can “effect changes in the laws that govern the microphysical level independently of any emergent properties” (p. 505). It is hard to know exactly what Pereboom means by this claim, since he offers little more by way of explanation. It looks like he may be attempting to give an “example” of such law-changing behavior when he writes the following: “Suppose that M1 were such an emergent property, M1 could then cause P2 in such a way that P2 is no longer governed by the ordinary microphysical laws, but instead by laws that take into account the special characteristics of the emergent properties (or no laws at all).” (p. 505). I think this can best be read, as I mention in the text, as suggesting that Emergentism entails the denial of physical causal closure. I don’t know where he gets the idea that such a claim is entailed by ontological emergence—I cannot find this claim made by any of the historical writers nor their contemporary expositors. But perhaps by “effecting changes in (microphysical) laws” he simply means that some microphysical events are brought about by nonmicrophysical events, thus “trumping” the supposed microphysical laws. If this is the right way to read Pereboom, then he is suggesting that Emergentism entails the denial of Microphysical Causal Closure. The second reason Pereboom gives for R-NRP’s not being an emergentist view is that emergentism entails the existence of emergent properties, instances of which are not “straightforwardly calculated and theoretically predictable from the facts about [their] basal conditions” (p. 504). In other words, Pereboom is advancing the idea that any theory of ontological emergence must entail Emergent Property Unpredictability. But as I argued in Chapter Three, Emergent Property Unpredictability is not a requirement upon MOE—it is an additional constraint that goes beyond what is needed to distinguish MOE from its rivals, and should be dropped from the theory.
controversial metaphysical principles, principles that I have no conclusive arguments for, but which nevertheless I believe to be highly plausible. But Pereboom, as we shall see, denies these principles. In the end, whether one accepts or rejects Pereboom’s argument for R-NRP will ultimately depend on one’s intuitions regarding these principles.

But before we embark upon a critical analysis of R-NRP, there is much stage setting to be done. Here is how I shall proceed. In the next section, I will lay out the basics of Pereboom’s account and show how it, as an instance of MOE, differs from both standard NRP and qualified micro-determinism. In the next two sections I will look deeper into two of the metaphysical notions central to Pereboom’s account—that of a “token causal power” and that of “property instance constitution.” Finally, in the last section, I will examine Pereboom’s main argument for R-NRP, arguing that its success depends upon the denial of two fairly plausible (by my lights anyway) principles about property instances and causal powers.

5.2 An Overview of Robust Non-Reductive Physicalism

Pereboom is driven to R-NRP for two reasons. The first is to account for the apparently unique types of psychological explanations that can be had only by reference to psychological properties—beliefs, desires, experiences, and the like. But of course anyone driven to a nonreductive physicalism is probably so driven partially, if not mainly, by that consideration. The other is his belief that what accounts for these irreducible explanations is that instances of psychological properties are causally efficacious—they actually bring about events in the world. Pereboom introduces his position as follows:

Robust nonreductive materialism, as I conceive it, is a view about specifically psychological explanations, states, and causal powers, although it easily generalizes to other levels of explanation. In this view, an event such as Mary’s
buying ice cream (M2) will have a psychological explanation in terms of a complex of mental states – beliefs and desires she has (M1). Each of M1 and M2 will be wholly constituted of microphysical events (P1 and P2 respectively) and there will be a microphysical explanation of P2 in terms of P1. The explanation of M2 by M1 will not reduce to the explanation of P2 by P1…Underlying the irreducibility of this explanation is the fact that M1 is not type-identical with P1, and that M2 is not type-identical with P2. More fundamentally yet, the psychological explanation appeals to the irreducibly mental causal powers of M1 to account for M2, while the microphysical explanation appeals to microphysical causal powers of P1 to account for P2. (p. 500, emphasis added)

The first distinguishing feature of R-NRP is that mental events (mental property instances) are wholly constituted of microphysical events. I will discuss in more detail presently the metaphysics of the constitution relation, but no matter how the precise metaphysical details fall out it seems clear enough that if a special science property instance is wholly constituted of a microphysical property instance, then the special science property instance is ultimately determined by the microphysical property instance, and thus the Microphysical Determination condition of MOE is met. We can also note that the picture Pereboom offers of the relationship between the mental and the physical is very much like the description I attributed in Chapter One to Terry Horgan, which was as follows:

\[
\begin{align*}
\text{M1} & \quad \text{causally explains} \quad \text{M2} \\
\text{P1} & \quad \text{causally explains (and causes)} \quad \text{P2}
\end{align*}
\]

**Figure 2**
Pereboom agrees that this diagram captures part of the relationship of the mental and physical, but not the entirety of the relationship. Pereboom’s belief is that not only does M1 causally explain M2, but that by contributing irreducibly mental causal powers, M1 actually causes M2. It is this aspect of R-NRP that makes Pereboom’s account stronger than versions of NRP advocated by Horgan and Antony (whose position was also discussed in Chapter One). Though standard NRP entails that mental properties are causally relevant in that they figure in laws and causal explanations distinct from the laws and causal explanations of microphysics, it leaves no room for these higher-level properties to possess genuine causal efficacy or causal powers, that is, to add real causal oomph to the world (that honor is of course reserved for the fundamental microphysical properties).

That some mental properties are causally efficacious is of course sufficient to distinguish R-NRP from standard NRP, but not necessarily from another view discussed in Chapter One, qualified micro-determinism. What is needed for R-NRP to be distinct from this view is that some mental properties are also causally autonomous. But as clear from the following quote from Pereboom, R-NRP entails just that, for it entails that some mental properties contribute novel causal powers:

Accordingly, the causal powers of M1 will not be type-identical with those of P1, and those of M2 will not be type-identical with those of P2. But neither will a corresponding token-identity thesis for these causal powers hold. For if it did, then the causal powers to which the psychological explanation refers would in the last analysis in fact be microphysical. (p. 501)
So M1 contributes at least one causal power that is not even token identical to any causal powers contributed by microphysical properties. But if one recalls our definition of

**Novel Causal Power Contribution**, namely,

**Novel Causal Power Contribution:** A property P contributes a *novel* causal power \( C(K, E) \) just in case an instance of P, were it to be instantiated in K, would cause, in conjunction with property instances Q1…Qn, an instance of E, no combination of property instances entailed to exist by K would, in conjunction with Q1…Qn, cause an instance of E, and there is no microphysical property distinct from P that contributes \( C(K, E) \)

then it is true that \( C(K, M2) \) is a novel causal power. And since novel causal power contribution is sufficient for causal autonomy, M1 is a causally autonomous property.

Note however that M1’s causal autonomy does not impinge upon the physical. Mental events cause other mental events, not physical events. So **Microphysical Causal Closure** is not violated. Thus, so far we have a view that appears to be consistent with the most standard formulations of physicalism.

That in a nutshell is Pereboom’s R-NRP. Though the basics of the view are easy enough to comprehend, we are still left with many questions. The first two questions have to do with the new concepts of a “token causal power” and “property instance constitution” that Pereboom introduces to the discussion. Those concepts will be examined in the following two sections. But the most pressing question surrounding R-NRP that I think we must address up front is this—why should we think, given the fact that mental property instances are asymmetrically necessitated by physical property instances, that they do any causal work at all? Isn’t it more natural to suppose that mental property instances simply “ride piggyback” on their microphysical necessitates, and that mental events, like successive images of billiard balls in a mirror that appear to interact with each other causally, really have no causally efficacy at all, but are merely reflections
of the thing in this world that do have real causal efficacy, namely, microphysical events.\textsuperscript{106}

Let me pause a minute to address this objection here, for if the objection is successful, then Pereboom’s position is a non-starter and there is no point in discussing the metaphysics of his view any further. But what precisely is the objection? It seems to amount to this—not only is it wrongheaded to suppose that mental property $M_1$ contributes a \textit{novel} causal power $C(K, M_2)$, it’s wrongheaded to suppose that $C(K, M_2)$ is a causal power \textit{at all}. The reason is that an instance of $M_2$ (call it $m_2$) has a complete explanation in the fact that it is necessitated by an instance of $P_2$ (call it $p_2$), and to suggest that there is any further causal explanation for the $m_2$ would be inappropriate. More specifically, it would violate what Jaegwon Kim has called “the principle of causal/explanatory exclusion”, according to which “no event can be given more than one complete and independent explanation.” (1993, p. 239) The argument against $M_1$’s contributing $C(K, M_2)$ would thus go as follows: As $m_2$ is fully explained by $p_2$, and as any explanation of $m_2$ that references an instance of $M_1$ (call it $m_1$) would be a complete explanation independent of the explanation of $m_2$ in terms of $p_2$, $m_1$ does not causally explain $m_2$. And if $m_1$ does not causally explain $m_2$, then $m_1$ does not cause $m_2$, and thus does not contribute $M_1$ does not contribute $C(K, M_2)$. It follows that the central tenet of R-NRP, that mental properties do contribute such causal powers, is false.

Now Pereboom never addresses this problem directly. He addresses what might be seen as a related problem, which is this: given that $p_2$ necessitates $m_2$ and that $p_1$

\textsuperscript{106} The imagery used here, that mental property instances (events) are like reflections in a mirror, is borrowed from Jaegwon Kim, who in turn borrows it from Jonathon Edwards. Edwards uses the imagery as part of an argument for the conclusion that there can be no temporally persisting objects, since God is actually recreating the world at every moment. See Kim (2003a, p.154).
causes m2, there are grounds for saying that p1 causally explains m2, and thus we have
two competing causal explanations for m2—how is this to be resolved? Pereboom
resolves it by arguing that M1 and P1 contribute distinct causal powers to their
bearers—the former contributes C(K, M2) whereas the latter contributes C(K, P2)—and
this is enough for us to be able to say that the two causal explanation for m2 do not
compete with each other, and that the mental is causally efficacious and autonomous.
That is a substantive claim and I will address it in the last section of this chapter as part of
my critical assessment of R-NRP. But we are still left wondering about the argument
introduced in the preceding paragraph—where, if anywhere, does it go wrong?

I think the argument goes wrong in assuming that the explanation p2 provides of
m2 and the explanation m1 provides of m2 are competing explanations. The former is a
synchronic, non-causal explanation, whereas the latter is a diachronic, causal explanation.
And it is certainly in no way obvious that these are competing explanations.\[^{107}\] Maybe p2
and m1 are just multiple, non-competing explanations of m2—after all, there are
numerous examples of events to which we are more than comfortable attributing multiple
types of explanation.\[^{108}\]

So I think this type of objection to the causal efficacy of special-science
properties is ultimately unsuccessful. Of course, there is another Kimian type objection
against the idea that M1 contributes C(K, M2) and that is that it violates the **Principle of
Downward Causation**. That principle is as follows:

\[^{107}\] This point has also been made in Crisp and Warfield (2001, p. 310).

\[^{108}\] Crisp and Warfield (2001, p. 310) cite the example of the property having the measles. It
seems correct to say that one instantiates this property because (causal explanation) one has been infected
with the measles virus. But, in addition, it is equally correct to say that one instantiates this property
because one instantiates certain microphysical base properties upon which having the measles supervenes.
When one gives an actual example like this, it seems even clearer that causal explanations and non-causal
explanations are not competing explanations.
Principle of Downward Causation: for all q such that q is an instance of a non-microphysical property Q and for all q* such that q* is an instance of property Q*, q* is caused or brought about by q only if q causes an instance (or instances) of some microphysical property or properties upon which Q* strongly supervenes.

If this principle is correct, m1 cannot cause m2 directly, but can only do so by causing p2. But p1 is already a sufficient cause of p2 (by Microphysical Causal Closure). So we have another problem—it seems that causal overdetermination of p2 looms.

Pereboom responds to this argument by rejecting the Principle of Downward Causation. In Chapter Two, I granted Kim this principle, but that was just for rhetorical purposes—I think there is good reason to think that the principle is false, or at least unmotivated. Recall that in Chapter Two I suggested that Kim’s argument was that Universal Supervenience entailed the Principle of Downward Causation. But actually it does no such thing. Recall Universal Supervenience:

Universal Supervenience: Necessarily, for any non-microphysical property Q, if anything has Q, there exists a microphysical base property P such that it has P, and, P is nomologically sufficient for Q.

All that is entailed by Universal Supervenience is that for all Q such that Q is a non-microphysical property and for all q* such that q* is an instance of some property Q* caused by Q, q* is brought about only by the causing of an instance (or instances) of some microphysical property or properties upon which Q* strongly supervenes. It does not entail that an instance q of Q must be the cause of q*; it is consistent with Universal Supervenience that some other property (perhaps a realizer of Q) be the cause of q*. So in the case we have been discussing, Universal Supervenience does not entail that m1 cause m2 by causing p2, it only entails that in order for m2 to be instantiated some base

\[109\] He never explicitly denies it, as he never even talks about it. But everything he says supports the conclusion that he denies it, and the tenets of R-NRP entail that it is false.
property nomologically sufficient for m2 must be caused by some property or other.\textsuperscript{110} Denying the Principle of Downward Causation then does not seem like much of a stretch, and this opens the door for a defense of MOE along the lines of Pereboom’s R-NRP.

So much for the necessary preliminaries. It’s time to examine the metaphysics of Pereboom’s view. I’ll start by taking a look at the idea of property instance constitution. Then I will turn to Pereboom’s claim that mental properties contribute novel token causal powers. Finally, I will look for Pereboom’s argument for the nonreducibility of mental causal powers, an argument that really amounts to his whole case for R-NRP.

5.3 Pereboom’s Constitution Relation

In this section I want to make the case for understanding the constitution relation described by Pereboom (which as we shall see is clearly an instance of R\textsubscript{THEO}) as a species of the determinable-determinate relation. But before doing so I should address what is probably a burning question in the minds of some readers, that is, why not just read Pereboom’s constitution talk literally by interpreting him as endorsing something along the lines of D.M. Armstrong’s ontology, according to which a universal (say property M), a particular (substance S), and the instantiation relation wholly constitute a state of affairs (S-having-M), which in turn can wholly constitute other states of affairs.\textsuperscript{111} But there are a couple of reasons I believe we should not interpret Pereboom as

\textsuperscript{110} I take this point from Crisp and Warfield (2001, p. 311).

\textsuperscript{111} See Armstrong (1989, pp. 88-96) and Armstrong (1997, pp. 113-138). Armstrong distinguishes between atomic states of affairs, which are constituted by a substance, property, and the instantiation relation, and molecular states of affairs, which are constituted by other states of affairs. Actually, a molecular state of affairs just turns out to be conjunction of other states of affairs, but they still have a useful role to play in Armstrong’s ontology.
advocating such an ontology. For one, he never speaks of states of affairs per se, but rather of events or property tokens, which may ultimately be the same thing as states of affairs, but seem to me to connote a different type of metaphysics. Second, I would think that if Pereboom intended his talk of constitution to be interpreted along Armstrongian lines, then he would be very diligent in stating so, but as a matter of fact he does not. And the final reason to suspect that Pereboom does not endorse an Armstrongian ontology is that, for Armstrong, all constituted states of affairs (what he calls molecular states of affairs) are conjunctive states of affairs, e.g., a constituted states of affairs might be the state of affairs (S-having-F and S-having-G). But when Pereboom speaks of a mental property instance as being wholly constituted by a (micro)physical property instance, he clearly does not mean, as we shall see, that the mental property instance is a conjunctive instance, or conjunctive state of affairs. For all of these reasons, I think we can eliminate the possibility that Pereboom’s ontological assertions should be interpreted a la Armstrong.

So how should we interpret Pereboom here? He makes it quite clear that he wants us to understand the relation of “constitution” holding between a mental property instance and a microphysical property instance as a type of realization relation and also to understand that mental properties are multiply realizable. Perhaps the first natural assumption then is that Pereboom understands mental properties to be functional properties, which, as anyone who reads much of the literature on reduction and emergence knows, are the premier example of realized properties. But Pereboom is adamant that R-NRP entails that mental properties are not functional properties. Rather, Pereboom describes mental properties as intrinsic properties each of whose instances are
identical to a certain type of structure realized by different physical configurations. He likens an instance of a mental property to a ball piston engine:

Characteristic of this engine is its having parts with particular shapes and rigidities, and these parts must be arranged in a particular way. These features are manifestly not functional relations that such an engine stands in; rather, they constitute intrinsic characteristics of this type of engine. At the same time, these characteristics are multiply realizable. The parts of the engine can be made of material of different sorts – as long as the material can yield, for example, the required shapes and rigidities. The ball piston engine, then, has a nonfunctionalist intrinsic structure that instantiates its causal powers, but it nevertheless admits of distinct realizations. (p. 511)

Of course this analogy is imperfect in that a ball piston engine is a substance, whereas a mental property instance is an event or state of affairs, but the idea being communicated is nevertheless fairly clear. For Pereboom, being a mental property just is being a certain type of structure. As Pereboom goes on to explain:

I suggest that we might indeed identify a mental state-type with a structure, but a structure more abstract than any specific neural structure, and one that can potentially be realized by a silicon-based system…My proposal is that there are structure-types that cannot be classified as specifically neural, but which must rather be categorized as mental, and which would be intrinsic properties of mental states. (p. 512)

Now Pereboom notes that one challenge to his theory is that such structures may not exist, e.g., that there may be no identifiable structural type common to both, say, a man and a dog (and perhaps a computer) when said creature is having the belief that there is danger nearby. But that seems largely an empirical matter that I will not be concerned with here.¹¹² What I am interested in is the metaphysics behind Pereboom’s theory of

¹¹² The reader may be concocting on his own another possible challenge to Pereboom’s theory, namely, that the proposed identification of mental properties, e.g. pain, with certain types of structures is highly implausible. How could pain be a type of structure? Perhaps even if one thought Pereboom had it right that mental properties are intrinsic properties, surely what is intrinsic about pain is its hurtfulness, not any type of structure. I am sympathetic with these concerns, but I will not address them here. I am concerned primarily with the metaphysics of Pereboom’s proposal and whether it can succeed as a theory of ontological emergence. Even if what Pereboom says about mental properties is false, his insights, it seems to me, could be applied to other special-science properties. For instance, is the property denoted by ‘having electronegativity of 3.44’ ontologically emergent? It is much more plausible that the property
mental properties. However, Pereboom doesn’t elaborate anymore on the metaphysics of mental properties, so we are left to do that work on our own.

It might appear at first glance that we should understand Pereboom’s claim to be that for any given mental property M, M is identical to some structural property P. But this cannot be the case. Consider once again the description of a structural property given by Armstrong: if F is a structural property had by an object x composed of proper parts \( x_1 \ldots x_n \) which instantiate intrinsic properties \( P_1 \ldots P_n \) respectively, then instantiating F is just having nonoverlapping parts \( x_1 \ldots x_n \) such that \( P_1 x_1, P_2 x_2, \ldots P_n x_n \) and \( R(x_1 \ldots x_n) \) (where R is some multi-place relation). (1997, p. 32) What is relevant here is that exemplifying a given structural property F entails having certain proper parts exemplifying certain properties. But if this is so, then structural properties are not multiply realizable—change any of the “base” properties or “base” parts of a structural property and one finds oneself with a new structural property. Since Pereboom wants mental properties to be multiply realizable, no mental property can be identical to a structural property.

What kind of properties are Pereboom’s mental properties then? They are properties asymmetrically necessitated by structural properties. And this makes sense even when we consider the analogy of the ball piston engine. The property being a ball piston engine (call it BP) is not identical with any structural property—if it were then a ball piston engine could only have one configuration, which is clearly false. Rather, BP is asymmetrically necessitated by a variety of structural properties, and BP itself is identical to the property of being of structure type Q, where Q specifies the relevant denoted by this predicate could be identical to a property like being structure S. So perhaps in the end Pereboom’s theory, if workable, will more plausibly apply to non-mental properties. My goal here is to see if it is workable at all.
structure that all objects instantiating any of the myriad realizers of BP have. Similarly, a property such as *believing that there is danger nearby* (call it M) cannot be a structural property, but it can be a property realized by a variety of structural properties, and M itself is identical to the property of being of structure type Q*, where Q* specifies the relevant structure that all objects instantiating any of the myriad realizers of BP have.

But the type of necessitation here is not mere nomological necessitation. It is not a matter of contingent law that when a structural property P is instantiated that M, the property of being a structure of type Q* is instantiated. This is a case of metaphysical necessitation—it is not true in any possible world that P is instantiated while M fails to be instantiated. What could be the reason for this? Well, one reason might be that M is a functional property—that is, having M is, by definition, having a property with causal role C, and P is a property that fills causal role C. But Pereboom has already told us that the properties that characterized R-NRP are not functional properties. Another possible explanation of P’s metaphysically necessitating M is that M is a property whose causal powers are a proper subset of P’s, and that the causal theory of properties is true.

According to the causal theory, “properties that have causal features non-dervatively...have them essentially, and are individuated in terms of them.” (Shoemaker 1997, p. 65) The conjunction of these two propositions would entail that it is metaphysically impossible for P to be instantiated without M being instantiated. But this also is not an option for Pereboom, who maintains that some mental properties contribute novel causal powers not contributed by any of their microphysical necessitators.

What options are left then for characterizing the constitution relation (which I will heretofore refer to as R_{\text{CON}})? As far as I can tell, the only remaining relation of
metaphysical necessitation that could apply here is the relation that holds between a
determinate property and its determinable (let us call this relation \( R_{DET} \)). \( R_{DET} \) seems to
meet both of the desiderata for \( R_{CON} \) prescribed by Pereboom—that \( R_{CON} \) be a relation of
asymmetric necessitation (what he calls realization) and that \( R_{CON} \) allow that mental
properties be multiply realizable. A determinate property (like \textit{being red}) does
asymmetrically necessitate its determinable (\textit{being colored}), and of course \textit{being colored}
is a multiply realizable property. But not only does \( R_{DET} \) match Pereboom’s desiderata
for \( R_{CON} \), it seems that many of the necessary conditions for a relation’s qualifying as an
instance of \( R_{DET} \) are satisfied by \( R_{CON} \). What necessary conditions am I thinking of?
Well, in his own discussion of the relevance of \( R_{DET} \) to the problem of the special
sciences and mental causation, Douglas Ehring lays out several such necessary
conditions, the most relevant of which are:

(i) \textbf{Asymmetric necessitation}—If \( P_1, \ldots, P_n \) is a complete set of same-
level determinates of \( Q \), then, necessarily, if \( x \) has \( P_j \) such that \( P_j \) is one
of \( P_1, \ldots, P_n \), then \( x \) has \( Q \); however, it is possible for \( x \) to have \( Q \)
without having \( P_j \).

(ii) \textbf{Downward necessitation}—If \( P_1, \ldots, P_n \) is a complete set of same-level
determinates of \( Q \), then if \( x \) has \( Q \), then \( x \) has \( P_1 \) or \( P_2 \), \ldots, or \( P_n \).

(iii) \textbf{Exclusion}—If \( P_1, \ldots, P_n \) is a complete set of same-level determinates
of \( Q \), then, if \( x \) has \( Q \), then \( x \) has only one of the properties in the set \( P_1, \ldots, P_n \).

(iv) \textbf{Ordering Principle}—If \( P_1, \ldots, P_n \) is a complete set of same-level
determinates of \( Q \), then for any set of objects \( o_1, \ldots, o_n \) such that \( o_1 \) has
\( P_1 \) ….. and \( o_n \) has \( P_n \), these objects are orderable with respect to \( Q \) (or
of some determinable of \( Q \)) in virtue of differences among these
determinates. \footnote{Ehring (1996, pp. 470-71). All four of these conditions are given explicitly just as quoted by
Ehring except for number (i). Ehring does list (i) as a necessary condition, but describes it only as such:
"asymmetric necessitation—determinates necessitate their determinables but determinables don’t
necessitate any particular determinate. I have made the principle more precise to match the precision of his
other necessary conditions. There are other necessary conditions Ehring gives that I don’t list here, what he
calls \textit{Non-Conjunctive Determinates} and \textit{Non-Disjunctive Determinables}. But these are conditions that
These all seem to be highly reasonable requirements on an instance of $R_{DET}$. (i) and (ii) are obviously such requirements. (iii) just tells us that an object instantiating a determinable can instantiate no more than one of its determinates, e.g. if $x$ is triangular, it can be equilateral or scalene, but not both. I’ve already explained why $R_{DET}$ is an example of asymmetric necessitation. It is also clear that (ii) is satisfied by $R_{DET}$, for in every case in which the property like being a certain structure $Q^*$ is instantiated, an instance of one of its realizing structural properties must be instantiated. Condition (iii), Exclusion, communicates the idea that if an object instantiates a determinable property it will instantiate only one of that determinable’s determinates (e.g., a triangular object can be equilateral or scalene, but not both). This is true because any given object can only instantiate one structural property. Finally, it seems relatively clear that the microphysical necessitators of Pereboom’s mental properties satisfy (iv), the Ordering Principle condition. Just as you can order the determinates of being colored by the differences among the determinates (e.g. red is more like orange than it is like yellow—they are orderable by color), so can you order entities that exemplify structure $Q^*$ based on the differences among the objects that instantiate the microstructural bases of $Q^*$. Consider the example of being an automobile engine (since I do not know much

---

115 One might think that (iii) fails to be true for one of the premier examples of $R_{DET}$, namely, the relation between color properties and the property being colored. Surely an object can instantiate more than one color—can not a ball be half blue and half red? But though this example does undermine Exclusion as formulated, it does not undermine the spirit of Exclusion. It simply shows that Exclusion probably needs to be reformulated to express the idea that for any object that instantiates a determinable like being colored, that object cannot be such that any given spatial location is completely characterized by two determinates (like being blue and being red).
about ball piston engines, I’ll use this more convenient example). The many determinates of the property *being an automobile engine* are orderable by structure—one can easily imagine one mechanic saying to another, “The Toyota engine is much more similar to a Honda engine than to a Ford engine. You know, the parts on the first two are just arranged in pretty much the same way, but the Ford’s parts are arranged much differently.” What is being communicated here is that the determinate property (here, a structural property) exemplified by the Toyota’s engine is more similar to the determinate property exemplified by the Honda’s than that exemplified by the Ford’s. Structures are orderable by their structural properties.

So my conclusion is that the $R_{\text{CON}}$ relation described by Pereboom is an instance of $R_{\text{DET}}$. Now as I noted in the introduction, if the determination relation constitutive of $R$-$\text{NRP}$ is indeed the determinable/determinate relation, this makes $R$-$\text{NRP}$ a unique and interesting position because the dominant view in recent literature has taken it to be the case that X’s being a determinable property is sufficient to guarantee X’s causal efficacy, but not its causal autonomy. For instance, Steven Yablo has argued that determinables and determinates are not causal rivals and that when it comes to the question of what

\[\text{116 But } R_{\text{DET}} \text{ is also consistent with } R_{\text{THEO}}. \text{ Recall Very-Nearly-the-Right-Formulation-of-} R_{\text{THEO}}:\]

Property instance x realizes property instance y iff (i) y is a property instance of some functional type, F, such that, necessarily, F is instantiated iff there is an instance of some or other type that meets condition, C; (ii) x is a property instance of some type that in fact meets C; and (iii) the instance of F whose existence is logically guaranteed by the holding of condition (ii) is numerically identical with y, and (iv) either x is not an instance of some functional type, G, such that, necessarily, G is instantiated iff there is an instance of some or other type that meets condition, D, or y is not an instance of some type that meets D.

If property instance x is a determinable property, then it is an instance of some functional type, F, i.e. it is instantiated only when a type meeting another condition C is instantiated (in this case of *being colored*, for instance, condition C would be “being a color”). So (i) is satisfied. (ii) and (ii) would also be clearly satisfied, and (iv) is satisfied in that y is not an instance of any type that meets condition D—suppose x is an instance of being red, then the condition upon x’s functional type G would be something like “being a red color”. But *being colored* is not a type that meets condition D.
causes the bull to charge on a given occasion, the cape’s being red or the cape’s being scarlet, being red and being scarlet both cause the charging. But all Yablo argues for is the causal efficacy of determinable properties, not casual autonomy. Instances of determinables “share in the honors”, as Yablo puts it, of causing the same effects instances of their determinates do, but no determinable contributes a novel causal power. (Yablo 1992, p. 272) But Pereboom seemingly disagrees. We will shortly get to Pereboom’s argument for his view, but first we need to address one more important issue that arises when we acknowledge the existence, as Pereboom does, of token causal powers.

5.4 Token Causal Powers and Revised Novel Causal Power Contribution

A token causal power is just an instance of a causal power. As I suggested in Chapter One, if a causal power is a property of a property (my tentative conclusion), then a token causal power is just an instance of one of these properties. Pereboom actually speaks as of token causal powers are constituents of the property tokens that have them, and that, just as mental property tokens are constituted by physical property tokens, so token mental causal powers are constituted by token physical causal powers.

Now once we recognize the distinction between causal power types and causal power tokens, a technical problem looms vis-à-vis Novel Causal Power Contribution, one that I have put off addressing to this point as, up until now, it would have introduced complications that were not relevant at the time. As it is, I have been suggesting all along that a property’s contributing a novel causal power, per Novel Causal Power Contribution, is sufficient for that property’s being causally autonomous. But an example should illustrate that this is not the case. Suppose there are two properties, both exemplified by some substance S: P, a microphysical property, and M, a special-science
property. Suppose that M contributes causal power \( C(K, M^*) \) and P contributes causal power \( C(K^*, P^*) \). Let us label the token of \( C(K, M^*) \) on this occasion \( c_1 \) and the token of \( C(K, P^*) \) as \( c_2 \). Suppose now that \( c_1 = c_2 \); though P and M contribute different causal power types, the causal power tokens each contribute are identical. In such a case, though M contributes a causal power that no microphysical property (including \( P_1 \)) contributes, it still seems that it would be wrong to say that M is causally autonomous, for the token causal power it contributes is identical to a token microphysical causal power.

Is such a situation possible? I have no argument that it is possible, but it certainly seems possible. And the scenario under which it seems most clearly possible is with a determinable and its determinates. Suppose, for example, that P is the property of being scarlet and M is the property of being red. Suppose also that M contributes causal power \( C(K, Q^*) \), the power to bring about an instance of a bull’s charging (as we all know that red capes are wont to do), whereas P contributes a distinct causal power \( C(K, Q) \) which is the causal power to bring about an instance of a bull’s charging very swiftly (as the bull is highly agitated by the scarlet cape). If, as it seems appropriate to think in this case, the instance of red just is the instance of scarlet, it also seems appropriate (once one accepts the existence of token causal powers as properties of or constituents of causally efficacious properties) that \( c_1 \), the token of \( C(K, Q) \), just is \( c_2 \), the token of \( C(K, Q^*) \), and that it is the possession of this one token causal power that is responsible for bringing about both the bull’s charging and the bull’s charging swiftly.

---

117 I use “\( K \)” and “\( K^* \)” here to reflect the fact that these properties are instantiated in different circumstances—M is instantiated in circumstances that include \( P \) (whereas \( P \) is not), and vice versa.

118 Although, as we shall see, a big part of Pereboom’s case for R-NRP is that, for some properties at least, such a situation is not possible. More on that to come.
If it is in general true that an instance of a determinable property is identical to an
instance of its determinate, then of course every instance of a mental property as
described by Pereboom would be identical to an instance of a microphysical structural
property.\(^{119}\) If that were the case, then M’s contributing a type of causal power that P
does not (nor any other microphysical property does not) would still not be sufficient for
M’s being causally autonomous, for the token of this power would be identical to the
token of the microphysical power. This is the reason Pereboom’s assertion that mental
properties contribute novel causal power tokens, not just types, makes his view an
instance of MOE. Pereboom’s mental properties contribute novel causal powers even
according to this revised principle that reflects the fact that a property’s contributing a
novel causal power should be sufficient for its being causally autonomous (I have
emphasized the additional phraseology emphasized that has been added to Novel Causal
Power Contribution to produce this new principle):

**Revised Novel Causal Power Contribution**: A property P contributes a novel
causal power C(K,E) just in case an instance of P, were it to be instantiated in K,
would cause, in conjunction with property instances Q1…Qn, an instance of E, no
combination of property instances entailed to exist by K would, in conjunction
with Q1…Qn, cause an instance of E, and there is no microphysical property
distinct from P that contributes C(K,E), and for all c such that c is a token of
causal power C(K,E), it is not the case that there exists some c* such that c* is a
token of a causal power contributed by some microphysical property such that
c=c*.

Given **Revised Novel Causal Power Contribution**, we are now prepared to give a final,
complete formulation of R-NRP.

---

\(^{119}\) When I here use the phrase ‘instance of a determinable property’ and ‘instance of its
determinate’, I am referring to property tokens, that is, property instances as were described in Section
1.4.2. Of course, there is another sense of ‘instance of a determinable property’ in which an instance of a
determinable just is its determinate, e.g., red is an instance of color, or being red is an instance of being
colored. In this sense of the term ‘instance’, a property could be an instance of another property. But that
is not how I am using the term. I am using the term ‘instance’ to refer to the instantiation of a property by
an object, that is, an event or a state of affairs, or a trope. Thanks to Mike Loux for helping me see this
distinction and pointing out the importance of clarifying my terminology.
(R-NRP): There is a set of properties $S$ such that for every member $M$ of $S$ and every instance $m$ of $M$:

i. $M$ is a special-science property irreducible to any microphysical property

ii. there exists some $P$ such that $P$ is a microphysical property and $m$ is “irreducibly constituted” by an instance of $P$, meaning that $P$ is a determinate of $M$ ($M$ bears $R_{\text{DET}}$ to $P$).

iii. There are at least some instances of $m$ that are causally efficacious with respect to the occurrence of some other property instance $n$ of $N$, and in such cases, $m$ contributes novel causal power $C(K, N)$ per Revised Novel Causal Power Contribution.

The question that confronts us now is whether there are good reasons to think that an instance of a determinable property is distinct from any instance of its determinate, and, secondly, to think that a determinable property could contribute a novel causal power not contributed by its determinate.

5.5 Analyzing Pereboom’s Case for R-NRP

Pereboom’s argument for the conclusion that a mental property instance $M$ “irreducibly constituted” by a physical property instance $P$ can still contribute novel causal powers proceeds as follows.\(^{120}\) It is possible, claims Pereboom, that $M$ be constituted by a different token base, $P^*$, differing only in that “a few neural pathways

\(^{120}\) For simplicity sake and to aid the reader visually, I will in this section use capital letters to refer to property instances, not properties, departing from my usual notation. I do this because capital letters are simply easier to read.
are used that are token distinct from those actually engaged.” (2002, p. 502) In other words, Pereboom believes there is a possible world at which, for instance, the very token mental instance that is your concentrating on this paper is realized by a different token physical state (not a different type state, mind you, just a different token state). Given this possibility, M is not identical to P. Furthermore, Pereboom argues, if M is not identical to P, then the token causal powers of M are not identical to the token causal powers of P. This is because if the token constituent of M had been P* rather than P, then the token causal powers of the token constituent would also have been different. But the causal powers of M would not have been different. So the token causal powers of M do not equal the token causal powers of its realizer.

Here then, presented in a more formalized way, is Pereboom’s argument. Where P, P* are distinct property instances of the same physical property type, and M a property instance of a mental property type, all of a substance S, and C_P, C_P*, and C_M are token causal powers contributed to S by P, P*, and M respectively, the argument proceeds as follows:121

(1) Suppose at time t in W:
   (a) P wholly constitutes M
   (b) P contributes C_P to S
   (c) M contributes C_M to S

(2) If (1), then C_P wholly constitutes C_M at time t in W

(3) There is a possible world W* where at time t,
   (a) P is absent

---

121 C_P, C_P*, and C_M are would more properly be designated C_P(K, Q), C_P*(K*, Q), and C_M(K**, N) as the type of causal power contributed by P and P* is different than that contributed by M. P causes a physical event and M a mental event.
(b) $P^*$ wholly constitutes $M$.

(c) $P^*$ contributes $C_{P^*}$ to $S$

(d) $M$ contributes $C_M$ to $S$

(e) $\sim (C_{P} = C_{P^*})$

(4) If (3), then $C_{P^*}$ wholly constitutes $C_M$ at time $t$ in $W^*$

(5) Therefore, $C_M$ is wholly constituted by $C_P$ in one world and fails to be constituted by $C_P$ in another world.

(6) Therefore, $C_M$ is not identical to $C_P$.

(2) and (4) are true according to Pereboom’s understanding of token causal power constitution, which I have no qualms with. The move from (5) to (6) is impeccable, so all of the action is clearly at premise (3)

I think Pereboom faces three main challenges in defending the truth of this premise. We can reveal the first challenge by looking at (3d). Why think that $M$ contributes the same token causal power in $W^*$ that it does in $W$? Could not $M$ contribute different token causal powers across different worlds? The only reasonable explanation for (3d), it seems, is that $M$ contributes $C_M$ essentially. Now there is of course much debate about whether the causal powers of properties are essential to them, whether a property $P$ contributes the same causal powers to objects across all metaphysically possible worlds. But this issue cross-cuts the issue at hand. One could hold a thoroughly dispositionalist view of properties and still object to Pereboom’s assertion that $M$ in $W$ and $M$ in $W^*$ contribute the same token causal power—all the dispositionalist thesis entails of $M$ is that it exhibit the same type of causal power across possible worlds. So one might be attracted to dispositionalism about property types, but also accept something like the following when it comes to property tokens:
If M is constituted by P in W and P* in W*, then in W, M contributes the same token causal power that P does, and, in W*, M contributes the same token causal power that P* does.

This principle can be viewed as a corollary to a more general principle, one that we might call Essential Causal Power Constitution.

**Essential Causal Power Constitution:** If C_X is a token causal power constituted by a distinct token causal power C_Y, then C_X is essentially constituted by C_Y.

This can be contrasted with the principle that Pereboom apparently (though never explicitly) endorses, which I shall call Essential Causal Power Contribution.

**Essential Causal Power Constitution:** If a property instance X contributes causal power token C_X, then X essentially contributes C_X.

Why do I believe that Pereboom endorses Essential Causal Power Contribution? Well, I think it is reasonable to infer he accepts it based on how he constructs (or better, how he does not construct) his argument. Pereboom could have argued for C_M’s non-identity with C_P in a much simpler way, without introducing different property tokens at all, simply by assuming that P contributes different causal power tokens in different worlds, whereas M does not. The only reasonable assumption for his not employing this line of argumentation is that he does in fact endorse Essential Causal Power Contribution.

That this principle is false is the first significant challenge to Pereboom’s argument. Is it false? Well though there may not be conclusive reasons to believe it is false, as we shall see shortly, it appears that it conflicts with another principle Pereboom endorses.

But first, let us turn to (3b). It’s not at all obvious that we should grant that M, being constituted by P in the actual world, could be constituted by a different token P* in another world. In his most recent article, Pereboom does not argue for this claim—he simply asserts it. However, in a previous piece (1991) penned with Hilary Kornblith, Pereboom and Kornblith elucidate the intuitions behind (3b). He asks us to imagine W*
as a world in which everything is the same as the actual world except for the fact that instead of eating from baby cereal box CB on your first birthday, you ate from type-identical baby cereal box CB* (and all the other minor differences this would entail). In the actual world, eating from CB resulted in your brain now being constituted by certain molecules and also resulted in the production of physical property instance P1 (of your brain) that constituted a mental property instance M1, your desire for ice-cream (with its concomitant causal power C_{ICE}). However, in W*, eating from CB* results in your brain now being constituted by different molecules than the ones it is constituted of in the actual world, and furthermore it results in the production of physical property instance P1* (of your brain) that also constitutes a desire for ice cream (and the same concomitant power C_{ICE}). Pereboom and Kornblith go on to say that “you clearly would have had the very same token desire for ice-cream with its attendant token psychological causal power C_{ICE}.” (1991, p. 132).

We should take some time to dwell on this line of reasoning, for it is absolutely crucial to the success of Pereboom’s view. The first question to ask here is this: why should mental property instances not be individuated by the microphysical property instances that constitute them? In other words, why not think a principle such as the following is true:

**Essential Property Instance Constitution:** if a property instance x is constituted by a property instance y, then x is essentially constituted by y.

In the earlier (1991) article, Pereboom seems to reject the intuitions behind **Essential Property Instance Constitution** when he writes that, “token psychological states, and their causal powers, are individuated by the psychological causal networks in which they play a role. They are not individuated by their physical constitution.” (1991, p. 132) But mental states being individuated by their roles in a causal network suggests that mental
properties are functional properties, something that Pereboom has now obviously repudiated. So this response is of no use to him now. Another reason to reject **Essential Causal Power Constitution** and **Essential Property Instance Constitution** suggested by Pereboom is that there is a strong intuition that, when it comes to material objects, analogous principles are false. Pereboom brings out this intuition when he writes about the renowned ship of Theseus:

> The ship of Theseus is not identical with its current token microphysical realization base, for it would have been the same token ship had the token microphysical realization been slightly different, and it will be the same ship when this microphysical realization changes. The ship is in this sense *token multiply realizable*. (2002, p. 502)

But this seems to be a case of comparing apples and oranges—that is, comparing property instance constitution to material object constitution. Barring some compelling reason, it doesn’t seem we should believe the intuitions we have about the latter apply to the former. We say “the statue could have been made out of different clay”. Does this provide justification for saying that: “this causal power could have been constituted of different causal powers”? I can’t see how it should. We say the first because of our intuitions about the persistence conditions of material objects. But do we have any intuitions about the persistence conditions of token causal powers? I don’t think we do; certainly not intuitions nearly as strong as we have concerning material objects. In any case, R-NRP depends upon the truth of (3b), but I do not think we have yet found a convincing reason to believe it.

But I think Pereboom encounters his greatest challenge in defending (3a). (3a), though critical to the defense of the nonidentity of physical causal powers and mental causal powers, seems very difficult to justify. In fact, there is a dilemma Pereboom faces
that follows from his joint commitment to **Essential Property Instance Constitution** and the truth of (3a). The dilemma is as follows:

Property instances are either events as described Jaegwon Kim, or they are not. Suppose that property instances are events of the Kimian sort, so that property instance $P$ is a structured complex whose three constituents are an object, the property exemplified by that object, and the time of the exemplification. On a Kimian view of events, if $E$ and $E^*$ are events, then $E = E^*$ iff all of the corresponding constituents of $E$ and $E^*$ are identical. In other words, if $E$ is an event that is $S$-instantiating-$P$-at-$t$ and $E^*$ (where $P$ stands for the property of which $P$ is an instance) is an event that is $S^*$-instantiating-$P^*$-at-$t^*$, then $E = E^*$ iff $S = S^*$ and $P = P^*$ and $t = t^*$. This feature of Kimian events is actually helpful to Pereboom in arguing for the distinctness of a token mental property and its constituent token physical property, for Kimian events are identical only if their constituent properties are identical, and this is not the case with a determinable and its determinate. However, if property instances are Kimian events, then (3a) is true only if the object instantiating $P$ in $W$ is distinct from that instantiating $P$ in $W^*$. This is because it is obvious that the time of exemplification and the property of exemplification is the same in $W$ and $W^*$, so the only thing that could distinguish $P$ from $P^*$ would be that the constituent substances are different. But this can’t be what Pereboom intends. Clearly you are the substance that in $W$ instantiates $P$. If making room for genuinely causally autonomous mental properties requires our positing that your having ingested different baby cereal at age one would have resulted in your nonexistence today, then so much the worse for causally autonomous mental properties.

For the second horn of the dilemma, suppose property instances are not events of the Kimian sort. What could they be? Well, they could be tropes, or they could be
Armstrongian states of affairs, or perhaps some other imaginative category. But the question as to exactly what they are is actually not as relevant as the question of what could possibly *ground* the fact that P in W is distinct from P* in W*. Pereboom seems to believe that what grounds this fact is that P and P* is that either they or the objects in which they occur are *made out of different stuff*. That is what the cereal box example seems to suggest—your ingesting a different box of cereal would have resulted in differences in your brain which would have resulted in P* rather than P. It seems then that Pereboom believes property instances of a given property are individuated by the stuff that either *they* or *the objects in which they occur* are made of.

But which of these options is it? Consider the second option—that what makes P distinct from P* is that P and P* are instances of objects constituted of different stuff (your brain in W versus your brain in W*). This is not a viable option for Pereboom. Recall that we concluded earlier that the physical properties that determine Pereboom’s mental properties are structural properties. Recall what a structural property is—a structural property F is the property of having nonoverlapping parts x₁ … xₙ such that P₁ₓ₁ , P₂ₓ₂ , … Pₙₓₙ and R (x₁ … xₙ) (where R is some multi-place relation). But if what grounds P’s being distinct from P* is that your brain in W has different parts than your brain in W*, then P and P* are indeed distinct, but they are not instances of the same property (changing the “constituent parts” of a structural property is sufficient for obtaining a new property type). But this contradicts Pereboom’s assumption that P and P* are instances of the same property. So P’s nonidentity with P* can’t be grounded in the fact that the brain’s parts in W are distinct from those in W*. As a matter of fact, for Pereboom to suppose that P and P* are instances of the same property, he has to say that
But perhaps what grounds the difference between P and P* is that they themselves are constituted of different stuff. But of what are P and P* constituted? I can’t see it being anything besides other property instances. And thus the rub. In his defense of (3b) and (3d), Pereboom utilized the notion that a property instance could be constituted of different property instances across possible worlds, thus rejecting Essential Property Instance Constitution. But a consistent rejection of this principle would lead us to the conclusion that P’s being constituted by different property instances than P* is not sufficient for P’s nonidentity with P*.

So let’s review the dilemma. If property instances are Kimian events, then the nonidentity of P in W and P* and W* obtains only if the subject instantiating P in W is distinct from the subject instantiating P in W*. But that is highly undesirable. If property instances are not Kimian events, then the nonidentity of P and P* is grounded either in the fact that S’s brain in W has different parts than S’s brain in W* or in the fact that P and P* are constituted of different property instances. But the former is not an option given that P and P* are instances of structural properties and the latter should be rejected by Pereboom if he is to be consistent in his rejection of Essential Property Instance Constitution. The overall conclusion is that we have no reason to believe it is even possible that there could be different instances of the exact same structural property across possible worlds.

122 All ontologies that I am aware of that acknowledge property instance constitution entail that property instances are constituted of other property instances. Consider D.C. Williams (1953) trope theory, of which everything is ultimately constituted of tropes, or Armstrong’s ontology, according to which the world is a state of affairs constituted by other states of affairs.
5.6 Conclusion

R-NRP is an intriguing, metaphysically rich, and unique theory that seeks to answer the Problem of the Special Sciences. As it is one that is consistent both with MOE and most of the doctrines commonly associated with physicalism, it should be of particular interest to those involved in the debate over emergence and reduction. Ultimately I think Pereboom’s defense of R-NRP is unsuccessful, but I do not think this shows conclusively that his position is false. What is needed are better reasons for thinking that both that a token mental property instance (construed as a determinable structure with physical structural properties as its determinate) could be constituted by different physical property instances across possible worlds, and that a token mental causal power could be constituted by different physical causal powers across possible worlds. Pereboom’s argument for those conclusions is not up to the task.

Nonetheless, there is much about Pereboom’s position that will be attractive to many. It is physicalist, but non-reductionist. And it is non-reductionist in a way that, if correct, preserves the causal efficacy and autonomy of mental properties. But for the reasons I have given, I can not bring myself to endorse the view. Perhaps new arguments from Pereboom or others could support its truth. Until then, however, I think we are best to conclude that Pereboom’s theory is just another attempt at a physicalist solution to the Problem of the Special Sciences that comes up a bit short.
CHAPTER 6

GILLETT’S STRONG EMERGENCE

6.1 Introduction

In this chapter I take up the task of clarifying, sharpening, and critiquing an account of MOE (Minimal Ontological Emergence) that has been set forth by Carl Gillett. No one in the last few years has written more on emergence and some of its related concepts—realization, causal efficacy, the nature of the physical—than has Gillett. Not only has Gillett been prolific, but he confronts head-on many of the metaphysical issues that surface from a careful reflection on these notions. For these reasons alone Gillett’s work is worthy of discussion in a dissertation covering the metaphysics of ontological emergence, and the reader has already been apprised of many of his ideas. But Gillett also offers his own account of an ontologically emergent property consistent with MOE, one that is interesting in that, so far as I can tell, it occupies a box unto its own in the taxonomy I presented in Chapter Five. Gillett’s characterization of an ontologically emergent property, or what he calls a “strongly emergent property”, combines the idea that emergent properties are realized by microphysical properties in that they bear $R_{CAUS}$ to microphysical properties with the thesis that emergent properties are still causally efficacious.\(^{123}\) However, realized

---

\(^{123}\) Gillett actually reserves the term ‘ontologically emergent’ to refer to properties that are unrealized and that could potentially be instantiated by nonphysical individuals, and groups together philosophers like O’Connor (1993), Hasker (1999) and Crane (2001) as advocating the existence of ontologically emergent properties. See Gillett (2002c), p. 107. Of course, an ontologically emergent
properties are not causally efficacious by contributing a novel causal power (which is disallowed by the nature of $R_{\text{CAUS}}$), but rather by synchronically, non-causally determining that microphysical properties have certain causal powers. Like Pereboom, Gillett’s tactic for escaping Kim’s Challenge is to deny the Principle of Downward Causation. According to Gillett, the only “downward determination” in which a strongly emergent property is involved is not causal determination, but that of synchronically determining microphysical causal powers. My goal in this chapter is to examine Gillett’s account, paying particular attention to this notion of synchronic causal power determination. Along the way I try to sharpen some of the concepts that appear in Gillett’s account and toward the end of the chapter I raise some concerns I have with Gillett’s account. None of these concerns, I suspect, show that Gillett’s account is irredeemably flawed; rather, my purpose here is simply to advance the discussion and note some areas where Gillett’s account needs to be sharpened and some of its central tenets better explained and defended.

This chapter is divided into three main sections. In the first section I exposit Gillett’s account, trying to reproduce it as faithfully as possible with minimal interpretation and commentary. I will, however, try to make the account more clear to the reader by introducing an example of a candidate emergent property and showing what Gillett’s account would have to say about this property (this is something, as I mentioned in Chapter Four, that Gillett himself does not provide us with). In the second section I try to illuminate Gillett’s account by suggesting some sharper analyses of concepts that property as described by Gillett could still qualify as ontologically emergent on my use of the term as well, since neither a property’s being realized nor its being instantiated by physical substances is required for it to be ontologically emergent in my sense of the term.

124 The reader is directed to section 3.5 for a statement of the Principle of Downward Causation and the argument that constitutes Kim’s Challenge.
he employs. In the final section I review several concerns I have with Gillett’s account, and then offer up some thoughts on what the future might hold for an account in the spirit of Gillett’s Strong Emergence.

6.2 How Gillett Develops His Account

Gillett’s account of MOE is built around the idea that certain realized properties may be causally efficacious not by contributing novel causal powers, but by synchronically, or, as he states, non-causally, determining the causal powers contributed by the microphysical properties that realize them. Ultimately it is Gillett’s dual commitment to the truth of physicalism and his belief that physicalism implies that all special-science properties are realized that drives him to this unique position on the nature of emergent properties. For Gillett, the operative notion of realization is an instance of what in Chapter Four I referred to as $R_{CAUS}$. That all causally efficacious non-microphysical properties bear $R_{CAUS}$ to microphysical properties is what Gillett believes is required by physicalism, and that all emergent properties are causally efficacious by synchronically determining (or partially determining) that a distinct property contributes a causal power is what makes Gillett’s account of MOE consistent with Causal Efficacy and Causal Autonomy.

In Chapter Four I gave a brief overview of the idea of Synchronic Causal Power Determination (SCPD). I want now to recount Gillett’s explanation of this phenomenon. Along the way I will attempt to make Gillett’s account more transparent by showing how it would apply to a property many who espouse Ontological Emergence would concur is
an emergent property, that of *being in pain*. Gillett first asks us to note that when a special science property is realized, it is realized in virtue of one of its instances being realized (it is instances that are truly realized, properties are only said to be realized in virtue of having instances that are realized). Furthermore, Gillett notes that in most cases in which an instance of a special science property is realized, it is actually realized by a collection of microphysical property instance, and not a single microphysical property instance. Our definition of realization then must capture this fact. With this in mind, Gillett defines realization with the principle he refers to as the Dimensioned View of Realization, which was introduced earlier:

**Dimensioned View of Realization (DVR):** Property/relation instance(s) $F_1-F_n$ realize an instance of a property $G$, in an individual $s$, iff $s$ has powers that are individuative of an instance of $G$ in virtue of the powers contributed by $F_1-F_n$ to $s$ or $s$’s constituents, but not vice versa. (Gillett 2002c, p. 99)

Gillett does give an example to illuminate his account of realization, that of a diamond (which he refers to using the variable ‘$s$’) instantiating the property $G$ of *being hard*. Gillett contends that the instance of hardness is realized by a set of properties $P_1…P_n$ of entities $s_1…s_n$ that compose $s$. Furthermore, included among the causal powers contributed by the diamond’s hardness will be causal power $C$, the power to cause scratches in glass. Gillett says that the sciences have revealed to us that it is the particular properties and relations of $s_1…s_n$ with their concomitant causal powers that account for

---

125 Again, I do not know that Gillett would agree that this property is a strongly emergent property (nor whether he thinks *any* properties actually qualify as strongly emergent), as he gives no concrete examples of such a property in his work.

126 What are the powers “individuative of an instance of $G$”? Gillett never says exactly, but I think he must mean the conditional powers contributed by $G$. This issue of what kinds of powers Gillett is referring to in his use of the term ‘causal power’ will be discussed at length later in this chapter.
the diamond’s being hard and its having causal power C. \textsuperscript{127} Thus $P_1 \ldots P_n$ are best understood as jointly realizing G and jointly determining that s has the powers individuative of G. \textsuperscript{128}

How would DVR apply to our featured example of being in pain? Suppose that G is the property of being in pain and that s is the composite physical object exemplifying this property—we can suppose for the sake of argument that it is a brain. \textsuperscript{129} According to DVR, an instance of being in pain will be realized by a set of properties $P_1 \ldots P_n$ of entities $s_1 \ldots s_n$ that compose the brain. Of course G will contribute many causal powers, but among its causal powers will be, let us suppose, the causal power $C^*$, which is the power to cause a certain muscle to contract in certain circumstances. Gillett would have us believe that if being in pain is realized, then it is the particular properties and relations of $s_1 \ldots s_n$, with their concomitant causal powers, that account for the brain’s being in pain and its having causal power $C^*$.

Next, Gillett notes that some properties are what he calls “heterogeneous in their contributions of causal powers” in that they “contribute some powers only under certain

\textsuperscript{127} Throughout his work Gillett actually alternates back in forth between the idea that substances possess causal powers and the idea that properties possess causal powers. He also sometimes speaks as if properties contribute causal powers. Of course, Gillett’s use of these terms is no anomaly—that’s why I tried to lend some clarity to causal power talk in Chapter One. In any case, Gillett would concede that another way of saying that “s has powers that are individuative of an instance of G” would be to say that “G contributes powers individuative of G to s.” I will leave DVR phrased as Gillett has phrased it, but as the notion of causal power contribution is the one that has been more rigorously developed in this dissertation; the reader can feel free to translate the locution into the more familiar language of “causal power contribution.” In other words, one can read the claim that a substance has a causal power as just another way of stating the idea that the substance instantiates a property which has or contributes that causal power.

\textsuperscript{128} Gillett (2002a): 318-19. If the reader is interested, in this piece Gillett gives more of the scientific details behind a diamond’s being hard and exactly how the properties of its component parts determine its hardness.

\textsuperscript{129} Of course, for all we know, the thing that instantiates being in pain may actually be a proper part of the brain or maybe even something bigger than the brain, but we can be fairly sure, if physicalism is true, that it is a composite entity. Otherwise we would have a very strange physicalism indeed.
conditions, thus taking such properties to have what Shoemaker (1980) terms ‘conditional’ powers.” (Gillett 2002c, p. 113) He cites the following example made famous by Shoemaker:

To better understand the scenario being proposed, let us examine the notion of a conditional power more carefully using Shoemaker’s original example of the property of being knife-shaped. When this property is instantiated in an individual with the properties of being made of steel and being knife-sized, then this property contributes causal powers resulting in an individual that cuts flesh. But when instantiated in an individual with the properties of being made of wax, or being of microscopic size, then the property of being knife shaped contributes causal powers that do not result in an individual that cuts flesh. Shoemaker thus concluded that many properties have “conditional” powers, causal powers they contribute to individuals only conditionally upon the instantiation of other properties. (Gillett 2002c, p. 113)

With the idea that some causal powers are contributed only conditionally upon the instantiation of other properties in hand, Gillett then explains synchronic causal power determination as follows. Suppose a composite substance s instantiates a property H that is realized by property instances P₁…Pₙ. Again, let’s assume that s is the brain, H the property of being in pain, and P₁…Pₙ microphysical properties of the brain’s constituents. Further, suppose P₁ is a property of one of the microphysical entities (call it ‘a₁’) composing the brain. Assuming that P₁ is heterogeneous in its causal power contribution, an instance of P₁ may contribute a certain causal power C* only upon realizing H. As Gillett says,

Thus we are assuming that C* is slightly different from the causal powers P₁ contributes when not realizing H. In such a situation, P₁ is thus individuated by one conditional power which instances of P₁ contribute only when realizing H…The property H partially determines the contribution of a causal power to an individual…Assuming that P₁ is instantiated in a₁, then a₁ having C* is accounted for only by ascribing the realized property H to the individual s that a₁ constitutes. (Gillett 2002b, p. 100)
Here is how I think this might work in the case of mental causation. An instance of the property *being in pain* does not actually cause anything that is not also jointly caused by the property instances of the parts of the brain acting together. So an instance of being in pain causes, say, the firing of a certain nerve ending N1 only if instances of microphysical properties also cause the firing of N1. But the idea here is that there will be at least one causal power that P1, a microphysical property of a part of the brain, contributes only upon the brain’s being in pain. It is hard to know what this might be. But perhaps it is a causal power that results in the firing of N1. By contributing that causal power, P1 is a partial cause of the firing of N1, but it would not have contributed that power had H (*being in pain*) not been realized in the brain. H synchronically determines that P1 contribute this causal power, and this role of synchronically determining the causal power contributed by a separate property is sufficient, reasons Gillett, for saying that H is causally efficacious.

Now it is when this notion of synchronic causal power determination is combined with a property’s being realized that one obtains what Gillett calls a “strongly emergent property.” I will label Gillett’s characterization as **Strong Emergence** (Gillett 2002c, p. 115):

A property instance X is strongly emergent iff:

(i) X is realized by other properties/relations and
(ii) X is causally efficacious, i.e. X partially non-causally determines the causal powers contributed by at least one of the fundamental microphysical properties/relations realizing X.

Gillett’s formulation of emergence is certainly novel and interesting. But there are some concepts and terminology that are left a bit vague and/or unclear. Let me turn to an examination of some of these now.
6.3 Sharpening Gillett’s Account

6.3.1 Gillett’s Causal Powers—Conditional or Manifest?

Before I move on to an examination and analysis of the details of Gillett’s account, I want to address the issue of what type of causal powers Gillett is speaking of when he says that some causal powers are contributed heterogeneously. Gillett suggests that it is conditional powers that are contributed heterogeneously, claiming that the knife example from the quote above shows that Shoemaker himself was committed to the existence of “conditional” powers, causal powers contributed to individuals, as Gillett says, “only conditionally upon the instantiation of other properties.” Now Gillett is correct in saying that examples like these motivate Shoemaker to conclude that properties have conditional powers, but it seems to me he has misinterpreted what Shoemaker believes a conditional power is. According to Shoemaker, “something has a conditional power if it is such that it would have a certain causal power if it had such and such other properties, where the possession of those properties is not itself sufficient to bestow that power.” (Shoemaker 1997, p. 63) Recall that when I discussed this issue in Chapter One, I labeled the “certain causal power” that a property would contribute under the right conditions (if such and such other properties were instantiated) as a causal power simpliciter, or a manifest causal power, and I described manifest causal power contribution as follows:

**Manifest Causal Power Contribution:** A property P contributes a manifest causal power M(E) just in case there is an instance of P that causes (or contributes to causing) an instance of E.

I also suggested that we interpret Shoemaker as claiming that a conditional causal power of a property P is an intrinsic property or constituent of P, whereas a manifest power (or causal power simpliciter) is an extrinsic property of P that P contributes only upon one of
its instances being involved in an actual case of causation. It is actually manifest powers that are contributed conditionally, or “heterogeneously” as Gillett says, by properties (although I suppose it could be truly stated that a conditional power is “activated” conditionally—what is really conditional then is the instance of causation that results from the causal power).

At least this is how it seems to me. But perhaps Gillett understands conditional powers differently. Recall the understanding of conditional causal power contribution I suggested in Chapter One:

**Conditional Causal Power Contribution**: A property P contributes a *conditional* causal power C(K,E) just in case an instance of P, were it to occur in circumstances K, would cause (or contribute to causing) an instance of E.

I further suggested that we conceive of K as a complex state of affairs that includes all of the states of affairs relevant to the causal production of E that do not involve P. K thus will also include states of affairs involving all other properties whose presence are relevant to the production of E (in that they also contribute causal powers relevant to E’s production), states of affairs involving properties whose absence is crucial to the production of E as well as states of affairs describing the laws of nature. I proposed that K be understood in this way to preserve both the intrinsicality of conditional causal powers and their essential possession. But perhaps Gillett conceives of a conditional causal power differently—as another type of extrinsic and contingent feature of a property. On such a conception we could suppose that a property P instantiated in substance s contributes a conditional causal power C(K,E), where K specifies a more limited range of circumstances than I had originally suggested. Suppose, for instance, that K only includes states of affairs describing the laws of nature, but fails to include any states of affairs about the distribution of properties within a given world. If K were to be
specified in this fashion, it would be possible that \( P \) contribute \( C(K,E) \) to \( s \) only upon \( s \)'s instantiating other properties. So suppose \( C(K,E) \) is the causal power to cut flesh when in circumstances \( K \) (i.e., when the laws of nature are a particular way). Then if \( P \) is the property of \textit{being knife-shaped}, it contributes \( C(K,E) \) only conditionally upon \( s \)'s instantiation of the properties \textit{being knife-sized, not being made of butter}, etc. So \( s \)'s contributing \( C(K,E) \) is thus a contingent state of affairs, contingent upon the distribution of properties (beyond \( P \)) in the world.

It’s hard to tell if this is how Gillett wants us to understand conditional causal powers and conditional causal power contribution, for he never gives us much of an explanation of what he takes a conditional causal power to be. I have no huge qualms with this understanding of conditional causal power contribution other than it makes conditional causal powers extrinsic, contingent properties, which to me seems counterintuitive. What could be more part of the essence of a property than its causal profile? Still, the reader may be wondering why the fuss about this issue. Why does it matter exactly what type of powers (conditional or manifest) we understand to be contributed heterogeneously by properties? The reason it matters is that the powers that are contributed heterogeneously by microphysical properties are the ones that wind up being synchronically determined on Gillett’s account. But, by my lights, if the microphysical causal powers that a realized property synchronically determines are in fact \textit{conditional} powers, then Gillett’s account faces a serious internal difficulty (one that I will reveal in the next section). Alternatively, if it turns out that the microphysical causal powers that a realized property synchronically determines are \textit{manifest} powers, the internal difficulty can be avoided. However, if the powers are manifest powers, then Gillett then faces a separate challenge, namely, that it seems that there is nothing for such
powers to do. As I reasoned in Chapter One, manifest powers seem extraneous and we should thus adopt a deflationary view of manifest powers. Thus the way we interpret Gillett’s use of the term ‘causal powers’ in condition (ii) of *Strong Emergence* determines which challenge Gillett’s account is up against. But before discussing either of these problems, we need to take a look at two other concepts that play a central role in Gillett’s account, both of which I think need some sharpening.

6.3.2 A Technical Issue with DVR

The first thing to do is to address a technical issue concerning *DVR*. As it stands, *DVR* does not rule out realization’s being a diachronic, causal relation, for *DVR* seems to be consistent with F1-Fn existing at some time $t_0$ and causing property $G$ to be instantiated at some later time $t_1$. Suppose that the diamond that was mentioned earlier is such that its parts $s_1…s_n$ are characterized respectively by property instances F1-Fn at time $t_0$ and these property instances contribute to their respective parts causal powers that result in the production of some property $G$ of $s$ at time $t_1$. $G$ might be some very particular “value” of hardness, perhaps a value the diamond did not have before $t_1$. Of course, along with this value of hardness will come certain causal powers, perhaps causal powers the diamond did not have before (it can now scratch bulletproof glass, whereas it could not before). The described scenario certainly seems to be a possibility—microphysical property instances (F1-Fn) of the diamond cause (or at least contribute to causing) the diamond ($s$) to instantiate at some macroscopic property ($G$) with its concomitant causal powers at a later time. But if so, then would not it be true to say that $s$ has powers that are individuative of an instance of $G$ in virtue of the powers contributed by F1-Fn to $s$ or $s$’s constituents, but not vice versa? It seems so.
This problem with DVR results from the fact that Gillett never spells out in more detail exactly what is meant by the locution after the biconditional, particularly the words ‘in virtue of’. Elsewhere, it seems clear that he thinks realization is synchronic; for instance, he writes that the kind of determination operative in DVR is “not temporal in nature, since the upward determination involved in realization is instantaneous” (Gillett 2002c, p. 105) but his definition does not specify this. However, I think his definition can be remedied. I suggest the following:

**DVR**: Property/relation instance(s) F1-Fn realize an instance of a property G, in an individual s, iff s’s possession of the powers that are individuative of an instance of G is synchronically determined by the powers contributed by F1-Fn to s or s’s constituents, but not vice versa.

This modification rules out the possibility of realization being a causal relation. It also rids Gillett’s account of realization of the troublesome “in virtue of” locution and since determination (as I am using the term) already incorporates the notion of asymmetric dependence (a notion Gillett clearly sees is an important aspect of realization), it seems an appropriate term to utilize here.

### 6.3.3 Non-Causal Determination

The next question to address is how we should interpret Gillett when he says that a property instance X synchronically, or non-causally, determines the causal powers contributed by at least one of the fundamental microphysical properties/relations realizing X. Earlier on in the chapter I glossed over this question but now it is time to look at it more seriously. In the paper where Gillett offers his most rigorous defense of NCD, he fails to provide a concrete example of non-causal determination, which makes interpreting him somewhat difficult. The best he gives is the following quote:
The property $H$ partially determines the contribution of a causal power to an individual, since $P_1$ only contributes $C^*$ to individuals when realizing $H$. Assuming that $P_1$ is instantiated in $a_1$, then $a_1$ having $C^*$ is accounted for only by ascribing the realized property $H$ to the individual $s$ that $a_1$ constitutes...$H$ is a necessary member of the properties which are only jointly sufficient for determining the contribution of $C^*$ to $a_1$. (Gillett 2002b, p. 100)

Let’s look at this last quote and see if we can ascertain from it a more precise notion of causal power determination. I will assume that Gillett intends us to read the conditional in the first sentence counterfactually. If we were to go on that quote alone, we might wish to formulate partial determination as follows:

**Synchronic Causal Power Determination 1 (SCPD1):** for all properties $X$ and $Y$ and all causal powers $C$, $X$ partially determines $Y$’s contributing $C$ to an individual iff were $X$ not to be instantiated, $Y$ would not contribute $C$.

But as formulated this won’t do. Consider once again our brain in pain. **SCPD1** would allow properties totally irrelevant to the brain to partially determine the causal powers of $a_1$ (one of the brain’s components). For instance, on this definition the property of *being a material universe* would partially determine $a_1$’s contributing $C^*$ since in no world where that property fails to be instantiated does $a_1$ contribute $C^*$ (since $a_1$ does not exist in such a world). One might think one could remedy this problem by characterizing the SCPD relation in such a way that its relata must be property instances of the same object (or at least of parts of the same object). Doing so would give us the following:

**Synchronic Causal Power Determination 2 (SCPD2):** for all individuals $s$, causal powers $C$, and properties $X$ and $Y$ such that $X$ and $Y$ are both instantiated either by $s$ or by one of $s$’s parts and all causal powers $C$, $X$ partially determines $Y$’s contributing $C$ to $s$ iff were $X$ not to be instantiated, $Y$ would not contribute $C$.

But this still is not quite there. Now, the property *being such that a material universe exists* will qualify as a property that partially determines $a_1$’s contributing $C^*$. Again, this does not seem correct. Intuitively, what is wrong here is that these properties are just not the types of properties that should be causally efficacious. But of course we cannot
remedy SCPD2 by quantifying only over causally efficacious properties, for we are looking for the principle that is supposed to illuminate for us a sufficient condition on a property’s being causally efficacious! But maybe we don’t need to quantify over causally efficacious properties—intuitively the causally efficacious properties are restricted to the intrinsic properties of an object. If we restrict the relata of this relation to intrinsic properties, we come up with the following:

**Synchronic Causal Power Determination 3 (SCPD3):** for all individuals s, causal powers C, and properties X and Y such that X and Y are both intrinsic properties of either s or one of s’s parts, X partially determines Y’s contributing C to s iff were X not to be instantiated, Y would not contribute C.

I think this is a fair and natural interpretation of Gillett’s notion of synchronic causal power determination. Let me turn now to some concerns I have with *Strong Emergence* as Gillett has formulated it.

### 6.4 Raising Some Concerns with Strong Emergence

I now want to look at Gillett’s account more critically and raise four separate areas of concern. I don’t claim that any of these are devastating objections to *Strong Emergence*, but simply point toward areas where perhaps some further work can be done in making the account more clear and more worthy of acceptance.

**Concern #1:** *It seems that realized properties are not the best candidates to be the determiners of the causal powers of microphysical properties. It seems other microphysical properties are better candidates*

This first worry I have concerns the motivations Gillett provides for assigning the role of “causal power determiner” to a realized property rather than the group of microphysical properties (or perhaps a proper subset of that group) that realize it.
Consider once again an object exemplifying realized property H that is realized by microphysical properties \( P_1 \ldots P_n \) of the object’s parts \( a_1 \ldots a_n \). And suppose for the sake of argument that a manifest causal power \( C^* \) of \( P_1 \) is synchronically determined according to SCPD3. The objection would go like this: It is unnecessary to suppose that it is \( H \) doing the non-causal determining when we could equally well suppose that it is the combination of \( P_1 \ldots P_n \) (or perhaps \( P_2 \ldots P_n \) if one finds it problematic that a property could be one of a group of properties that synchronically non-causally determines its own casual power contribution) doing the non-causal determining. After all, it would seem that just as it is true that were \( H \) not to be instantiated, \( P_1 \) would not contribute \( C^* \), so if \( P_1 \ldots P_n \) were not instantiated, \( P_1 \) would not contribute \( C^* \). Plus, we already have independent reasons for believing in microphysical properties. \( H \) seems unnecessary here—it is realized and so does not contribute any novel causal powers, and it is at best redundant in the role of causal power determiner. Thus, we should simply reject the existence of properties like \( H \), i.e., strongly emergent properties.

Gillett is clearly cognizant of this objection. In one article, after introducing the concept of synchronic causal power determination, he continues:

An obvious question is commonly raised at this point. Why is the realized property \( H \) the best candidate for partially determining the power contributed by \( P_1 \), rather than the combination of all the microphysical individuals, properties and relations that realize/constitute \( H \) and \( s \)? (2002b, p. 103)

Indeed this is exactly my concern. Gillett’s response is as follows:

I think there are real difficulties surrounding the metaphysics, and even existence, of such structural properties, but we can see that the point I am making still goes through even if we accept their existence. For...structural properties, if they exist, are themselves realized properties. The combination of microphysical properties/relations is not identical to any of these particular microphysical

\[130\] I myself find no principled objection to the possibility of \( P_1 \) partially determining (along with other properties) what causal powers it contributes, and so I will continue to refer to this possibility.
properties/relations, but is instead realized by them. The structural property is thus not an alternative to some realized property instance, but is a realized property instance itself. (2002b, p. 103)

I think this response fails on two levels. First, and most importantly, it simply misses the objection. The objection is not that some structural property Q is a better candidate for the role of “causal power determiner” than is H. The objection is that the combination of properties P₁…Pₙ is a better candidate—and those properties are not the same as some structural property (though they may, as Gillett suggests, realize one, or, as Armstrong would prefer to say, be “parts of” one). Secondly, even if structural property Q is a realized property, it is still a different property than H. And it is importantly different in this way—though H is multiply realizable, Q is not. While it is certainly true that certain structures of composite objects are multiply realizable, structural properties are not multiply realizable. For instance, the property of being knife-shaped is a multiply-realizable property that a knife (and many other things) has, but it is not a structural property (at least as we are defining them). Rather, there are many structural properties that realize being knife-shaped. A structural property is a very specific property—the property of having such and such parts with such and such properties standing in such and such relations. Change one of these parts, or one of these properties, or one of these relations, and one now is confronted with a different structural property. Structural properties are to their “constituent” properties as composite objects are to their parts if mereological essentialism is true, namely, essentially constituted by them. Thus a structural property cannot be multiply realized

Can we find in Gillett’s writings another reason for crediting H as the causal power determiner of P₁? Perhaps this—Gillett says that the causal efficacy of some emergent properties (like mental properties) is supported both by our first-person
experience and the scientific investigations conducted within the special sciences. While I’m tempted to agree with him on this, it is not the same as saying that our first-person experience or the investigations of the special-sciences supports the idea that properties are causally efficacious via SCPD3. It seems like any first-person experience we have of causal efficacy of mental properties (like beliefs) is in their directly bringing about certain other events, not in their being causally efficacious in the way described by Gillett, something imagined only by philosophers. And it seems equally dubious to suppose that the sciences uncover non-causal determination relations. That just seems like a philosophical “discovery.”

Gillett’s proper response to Concern #1, it seems to me, should be that the premise upon which this objection is based is not likely to be true. What is this premise? Well, Concern #1 is predicated, it seems to me, upon belief in the following premise:

Premise: Were it to be true that if H were not instantiated, P1 would not contribute C*, then it would also be true if P1…Pn were not instantiated, P1 would not contribute C*.

This premise is dubious in that H could be a multiply realized property. Suppose it were the case that H is realizable both by P1…Pn and by a distinct combination of microphysical properties P0…Pm, a combination that still includes P1. And suppose that in both case (both where H is realized by P1…Pn and where H is realized by P0…Pm) it is still true that P1 contributes C*. Then it is not the case that were P1…Pn not to be instantiated, P1 would not contribute C*. However, it is still be the case that were H not to be instantiated, then P1 would not contribute C. This possibility is what makes Premise false, and, it seems to me, is all that Gillett needs to undermine Concern #1.

If we can learn anything from Concern #1, it is perhaps that emergent properties on Gillett’s account need not only be realized, but multiply realizable. For if they are not
multiply realizable, then one of the main motivations for supposing that such properties exist, namely, that they serve as non-causal determiners of the causal powers of their microphysical realizers, is seriously undermined.

*Concern #2: It appears that it may be impossible for a property to qualify as strongly emergent on Gillett’s account.*

I fear this is true if we interpret the reference to causal powers in (ii) of *Strong Emergence* as being to conditional powers. Let me explain why. Recall DVR*:

**DVR***: Property/relation instance(s) F1-Fn realize an instance of a property G, in an individual s, iff s’s possession of the powers that are individuative of an instance of G is synchronically determined by the powers contributed by F1-Fn to s or s’s constituents, but not vice versa.

And then recall **Strong Emergence**:

A property instance X is strongly emergent iff:

(i) X is realized by other properties/relations and
(ii) X is causally efficacious, i.e. X partially non-causally determines the causal powers contributed by at least one of the fundamental microphysical properties/relations realizing X.

If every reference to causal powers in these principles is to conditional causal powers, then (i) is satisfied only if (ii) is not satisfied, and thus no property could ever qualify as strongly emergent. Suppose that H is a property realized by microphysical properties P1…Pn, and thus that (i) is satisfied. By **DVR** it’s true that the conditional causal powers contributed by H are synchronically determined by the conditional powers contributed by P1…Pn, but not vice versa. So H qualifies as realized only if it does not synchronically non-causally determine any of the conditional causal powers contributed by P1…Pn. But of course (ii) is satisfied only if H *does* synchronically non-causally determine the causal powers contributed by at least one of P1…Pn. So (i) is satisfied only if (ii) is not satisfied. This makes strongly emergent properties impossible.
But naturally this conclusion only follows if the reference to powers in (ii) is to conditional powers. The solution I suggest for Gillett is to interpret the power referenced in (ii) as manifest powers, for this will alleviate his account of the internal difficulty. Of course, then Gillett’s account confronts the next concern.

Concern #3: There are good reasons to think that there are in fact no manifest causal powers. So if making sense of Gillett’s account requires accepting the existence of manifest causal powers, things do not bode well for Strong Emergence.

Perhaps things do not bode well if we insist on taking Gillett’s reference to causal powers in condition (ii) as denoting a real entity. But though making sense of Gillett account requires that we interpret the reference to causal powers in condition (ii) as a reference to manifest powers, this does not commit Gillett to the existence of manifest powers. He could, and I believe should given the reasons I have given in Chapter One, adopt a deflationary view of manifest causal powers, according to which saying that a property P contributes a manifest causal power to produce E is just another way of saying that an instance of P causes an instance of P (that is, actually causes, not just potentially causes). If we adopt this deflationary view of manifest powers, then condition (ii) of Gillett’s Strong Emergence could, for purposes of clarity, be revised to eliminate reference to causal powers and replace it with reference to instances of causation. The revised account would look like this:

**Revised Strong Emergence**

A property instance X is strongly emergent:

(i) X is realized by other properties/relations and
(ii) X is causally efficacious, i.e. X partially non-causally determines that an instance of at least one of the fundamental microphysical properties/relations realizing X causes an instance of some other property.
So, rather than determining that a causal power is contributed, X determines that an instance of causation obtains. Again, just as X was not causally responsible for the contribution of a causal power by one of its realizers (say, P1), but merely synchronically determined it, so X is not causally responsible for the instance of causation holding between P1 and some subsequent property instance E, but merely synchronically determines it. Thus, it would not be correct to say that X causes E.

What would it mean for X to determine that an instance of causation occurs? Well, I suppose we could construct a principle analogous to SCPD3. Recall that principle:

**Synchronic Causal Power Determination 3 (SCPD3):** for all individuals s, causal powers C, and properties X and Y such that X and Y are both intrinsic properties of either s or one of s’s parts, X partially determines Y’s contributing C to s iff were X not to be instantiated, Y would not contribute C.

A principle regarding synchronic causal instance determination might look like this:

**Synchronic Causal Instance Determination:** for all individuals s, and property instances X, Y, and E such that X, Y, and E are all distinct, X and Y are both instances of s, and Y causes or contributes to causing E, X partially determines Y’s causing or contributing to causing E iff were X not to be instantiated, Y would not cause or contribute to causing E.

This principle thus reveals to us an alternative way to interpret condition (ii) of **Revised Strong Emergence.** And though it eliminates all references to causal powers, I think it is still in the spirit of Gillett’s original idea, just modified to accommodate our finding that it must be manifest powers and not conditional powers that are synchronically determined by realized, emergent properties.

*Concern #4: Even if we interpret the causal powers referred to in condition (ii) of **Strong Emergence** to be manifest causal powers, and recast Gillett’s account*
Accordingly via Revised Strong Emergence, Gillett’s account still appears to be too weak. Too many properties qualify as strongly emergent on his view.

Actually, whether the account is too weak depends on exactly which manifest causal powers are being quantified over in condition (ii) of Strong Emergence, and what kind of property instances E are being quantified over in condition (ii) of Revised Strong Emergence. Is it all the possible manifest causal powers or just some subset of them? If it is all such powers then it indeed appears that Strong Emergence will be too weak. Allow me to illustrate.

Though Gillett does not provide us with examples of strongly emergent properties, if we know anything about ontological emergence it is that mass properties should not qualify as being ontologically emergent. So, for instance, the property of *having mass of 1 kg* should not qualify as being emergent—there is no clearer example of an additive or resultant property than this one. Yet if the quantification in condition (ii) is over all manifest causal powers, it does appear that *having mass of 1 kg* qualifies as being strongly emergent. To see this, suppose that a1 is a microphysical component of a baseball, which instantiates some mass property P1 (we’ll just call it *having mass M*, since the actual mass value is not important). As a1 is one of the microphysical components of the baseball, P1 will be one of a collection of properties (which includes the masses of all the other parts of the baseball) that realizes an instance of H in the baseball, where H = *having mass of 1 kg*.\(^{131}\) Suppose the baseball is placed on a scale. Now, in such circumstances, it seems clear that one of the manifest causal powers that P1 will contribute is the power to partially cause an instance of E, where E is the property of

---

\(^{131}\) We can assume that this is one of the “heavy balls” that were used back at the beginning of the last century (if the balls used back then were indeed heavy, it would surely explain why home runs were so hard to hit!).

183
registering a mass of 1 kg instantiated by the scale. But of course P1 will contribute this causal power only when realizing H—were H not to be instantiated, then P1 would not contribute the manifest power to partially cause an instance of E. So H qualifies as strongly emergent.

Clearly this won’t do. It seems to me that the only reasonable reply to this concern is to restrict the caused instances (instances caused by the microphysical properties realizing the realized property) quantified over in Synchronic Causal Instance Determination to exclude instances of macroscopic properties (like registering a mass of 1 kg) for which realized properties are also causally responsible. Perhaps this will amount to restricting these caused instances to instances of other microphysical properties.

6.5 Conclusion

Gillett’s account of MOE is a very interesting one, and will be of particular interests to physicalists in that it is consistent both with every causally efficacious property’s being realized by microphysical properties and with Microphysical Causal Closure. However, I think Gillett’s account suffers from some deficiencies that need to be remedied. For one, it appears that if the causal powers of instances of microphysical properties that are determined by realized properties are in fact conditional causal powers, then Strong Emergence implies that emergent properties are impossible. However, it appears that Gillett does intend us to interpret these causal powers as conditional powers. So Gillett must either provide us with a conception of a conditional power that will resolve the internal difficulties of his account, or embrace the idea that the power contributed heterogeneously by microphysical properties are in fact manifest powers. If this is the route he takes, then Gillett needs to explain why his account is not
too weak, as it appears on the most natural interpretation of Revised Strong Emergence that a property like having a mass of 1 kg will qualify as strongly emergent, which certainly should not be the case. Gillett needs to somehow restrict the types of manifest causal powers that genuinely emergent properties synchronically determine. If Gillett is up to these tasks and can produce adequate responses to my concerns, then I do think we have a very intriguing account of MOE that would be a departure from the ordinary conception of ontological emergence, and thus worthy of serious consideration by many philosophers, who have, it seems, grown quite tired with the ordinary.

6.6 Some Final Thoughts on Ontological Emergence

In this dissertation I have looked more deeply at the concept of an ontologically emergent property, and theories that include the idea that there are such things, which I referred to as theories of Ontological Emergence, or OE. Some of my main conclusions have been that the classic conception of an ontologically emergent property, what I referred to as COE, and many of the contemporary variations on this concept (including Jaegwon Kim’s) are too strong, for they include necessary conditions like Emergent Property Unpredictability and Novel Causal Powers, that are not required to distinguish OE from rival theories. What I did conclude is that the following is sufficient for distinguishing OE from its rivals:

**Non-Reduction:** For all M such that M is an ontologically emergent property, there is no microphysical property P to which M is reducible.

**Microphysical Determination:** For all m such that m is an instance of an ontologically emergent property, the occurrence of m is ultimately determined by an instance or instances of microphysical properties.

**Causal Efficacy:** For all M such that M is an ontologically emergent property, there is at least one instance of M that is causally efficacious.
**Causal Autonomy:** For all M such that M is an ontologically emergent property, M is causally autonomous.

I called the conjunction of these conditions **Minimal Ontological Emergence**, or MOE.

I have also presented a taxonomy of possible accounts of OE. As I see it there are four likely candidates for the dependence relation that holds between ontologically emergent properties and microphysical properties, and two plausible explanations for what it is for a property to be causally efficacious and autonomous. Of the eight possible “boxes” this gives it, seven are live options, it seems to me, although only four of the boxes house accounts currently on the offing. Finally, I looked in some detail at two accounts consistent with MOE and with physicalism, Derk Pereboom’s **Robust Non Reductive Physicalism** and Gillett’s **Strong Emergence**. I raised concerns with each, and while I believe each encounters some serious challenges, I think it is possible that either account might work with more explanation and justification of its major claims.

Where does this leave us? Hopefully with a better understanding of the concept of ontological emergence and a better understanding of what additional theses or principles must be true if an account consistent with MOE is to succeed. I did not set out to defend any particular account of MOE, nor even the idea that there are ontologically emergent properties, but merely to sort out the playing field in a way better than what was currently available.


