Explaining Stability And Change In Natural Systems

Stephen Esser
University of Pennsylvania, sesser76@gmail.com

Follow this and additional works at: https://repository.upenn.edu/edissertations

Part of the Philosophy of Science Commons

Recommended Citation
https://repository.upenn.edu/edissertations/3111

This paper is posted at ScholarlyCommons. https://repository.upenn.edu/edissertations/3111
For more information, please contact repository@pobox.upenn.edu.
Explaining Stability And Change In Natural Systems

Abstract
An aim of science is to increase our understanding of the natural world. A primary means for doing so is by providing explanations, which often proceed by tracing the causes of phenomena. How can a causal explanation lead to understanding? While explanations can take many forms, I argue that to succeed they must embody a conception of causation shared with their audience. The challenge then, is to describe this conception and detail its role in explanation. While there is good evidence that scientists employ more than one causal concept, I argue that the concept of productive causation (centered on the notion of bringing about change via a connection) has a primary role in natural science explanations. After critiquing other philosophical accounts, I develop a new theory of productive causation and show how it provides an underpinning for successful explanations. The heart of the theory is a network of persisting processes that possess dispositions toward change-producing mutual interactions. I argue that in a good explanation, the scientific entities, properties and activities invoked will correspond to the theory's depiction of causal structure. One important dimension of the theory describes how repeated patterns of interaction can give rise to a hierarchy of composite processes. This allows the theory to account for stabilized entities at various spatio-temporal scales. In turn, this enables the approach to be applicable throughout the natural sciences. After starting with simple examples, I show how the theory deals with more challenging cases from physics to biology. I conclude that the approach illuminates how explanations of various forms across diverse disciplines can lead to scientific understanding.

Degree Type
Dissertation

Degree Name
Doctor of Philosophy (PhD)

Graduate Group
Philosophy

First Advisor
Michael Weisberg

Subject Categories
Philosophy of Science

This dissertation is available at ScholarlyCommons: https://repository.upenn.edu/edissertations/3111
EXPLAINING STABILITY AND CHANGE IN NATURAL SYSTEMS

Stephen Esser

A DISSERTATION

in

Philosophy

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2018

Supervisor of Dissertation

__________________________

Michael Weisberg

Professor of Philosophy

Graduate Group Chairperson

__________________________

Errol Lord, Associate Professor of Philosophy

Dissertation Committee

Karen Detlefsen, Professor of Philosophy

Quayshawn Spencer, Associate Professor of Philosophy
ACKNOWLEDGMENT

I am extremely grateful to Michael Weisberg for his insight and guidance. The care, expertise and energy he brings to everything he does never ceases to amaze me, and his unflagging encouragement at every juncture enabled me to move forward. I am indebted to Karen Detlefsen for invaluable mentorship, advice and feedback over the entire course of my graduate career. This importantly includes my uncertain early days as a student returning to academia after a very long absence. I also thank Quayshawn Spencer for incredibly astute comments and suggestions, and for making me consider other philosophical perspectives that might be brought to bear on my work. I was extremely fortunate to receive instruction from or otherwise professionally interact with nearly all of Penn’s philosophy faculty during my course of study. Also, the support from my fellow graduate students was simply astounding. You all have my admiration and gratitude.

I thank Justin Bernstein, Chetan Cetty, Devin Curry, Louise Daoust, Lucas Dunlap, Louis Esser, Nabeel Hamid, Melissa Jacquart, Karen Kovaka, Mary Morgan, Pierce Randall, Carlos Santana, and Dan Singer for helpful comments.

I am blessed with a wonderful family. I thank Louis and Char for being constant sources of inspiration as well as joy. Lastly, I thank my wife Dana, who makes everything good in my life possible.
ABSTRACT

EXPLAINING STABILITY AND CHANGE IN NATURAL SYSTEMS

Stephen Esser
Michael Weisberg

An aim of science is to increase our understanding of the natural world. A primary means for doing so is by providing explanations, which often proceed by tracing the causes of phenomena. How can a causal explanation lead to understanding? While explanations can take many forms, I argue that to succeed they must embody a conception of causation shared with their audience. The challenge then, is to describe this conception and detail its role in explanation. While there is good evidence that scientists employ more than one causal concept, I argue that the concept of productive causation (centered on the notion of bringing about change via a connection) has a primary role in natural science explanations. After critiquing other philosophical accounts, I develop a new theory of productive causation and show how it provides an underpinning for successful explanations. The heart of the theory is a network of persisting processes that possess dispositions toward change-producing mutual interactions. I argue that in a good explanation, the scientific entities, properties and activities invoked will correspond to the theory’s depiction of causal structure. One important dimension of the theory describes how repeated patterns of interaction can give rise to a hierarchy of composite processes. This allows the theory to account for stabilized entities at various spatio-temporal scales. In turn, this enables the approach to be applicable throughout the natural sciences. After starting with simple examples, I show how the theory deals with more challenging cases from physics to biology. I conclude that the approach illuminates how explanations of various forms across diverse disciplines can lead to scientific understanding.
# TABLE OF CONTENTS

ACKNOWLEDGMENT ........................................................................................................ III

ABSTRACT .................................................................................................................... IV

LIST OF TABLES .......................................................................................................... IX

LIST OF ILLUSTRATIONS ............................................................................................. X

CHAPTER 1: INTRODUCTION ......................................................................................... 1

1. Scientific Understanding and the Workings of the World ........................................ 1

2. Advantages over Competing Approaches .............................................................. 4

3. Implications for the Metaphysics of Causation ...................................................... 7

4. Overview of Chapters ............................................................................................ 11

CHAPTER 2: TWO CAUSAL CONCEPTS IN EXPLANATION ..................................... 15

1. Two concepts of causation and two approaches to causal explanation ............... 16
   1.1 Production and difference-making .................................................................... 17
   1.2 The concepts at work in scientific explanation ................................................ 22

2. The debate over explanation and the question of conceptual priority ............... 25
   2.1 From the priority of the physical to the priority of production? ....................... 27
   2.2 Psychological evidence for priority? ................................................................. 31
   2.3 Implications for the debate over explanation ................................................... 34

3. Determining applicability criteria ........................................................................ 35
   3.1 Teleology and difference-making .................................................................... 36
3.2 Normative contexts and the application of the two concepts........................................ 39

4. The missing theory of production.................................................................................. 41

CHAPTER 3: A REVIEW OF CAUSAL PROCESS THEORIES.......................................... 44

1. Russell’s Causal Lines.................................................................................................... 45

2. Salmon’s Mark-Transmission Theory.......................................................................... 48
   2.1 Causal Processes.................................................................................................... 50
   2.2 Propagation............................................................................................................ 53
   2.3 Production.............................................................................................................. 56
   2.4 Probabilistic Causation.......................................................................................... 59
   2.5 Causation and Constitutive Explanations............................................................... 60

3. Criticism of the Mark-Transmission Theory.............................................................. 61
   3.1 Capabilities and Counterfactuals........................................................................... 62
   3.2 Distinguishing Causal Processes from Pseudo-Processes....................................... 65
   3.3 Vagueness Objections and Explanatory Adequacy............................................... 67

4. Conserved Quantity Theories...................................................................................... 69
   4.1 Dowe’s Theory and Salmon’s Revision................................................................. 69
   4.2 Reliance on Scientific Laws/Generalizations......................................................... 73
   4.3 Explanatory Adequacy and Other Criticisms......................................................... 77
   4.4 Contemporary Status of Process Theories............................................................. 79

5. Causal versus Scientific Theorizing and the Implications for Process Theory................. 80
   5.1 Salmon on Unobserved Theoretical Entities......................................................... 80
   5.2 Scientists in the Causal Network.......................................................................... 84
   5.3 An Alternative Path Forward for Process Theory................................................ 85

CHAPTER 4: THE CAUSAL NEXUS AND ITS ROLE IN EXPLANATION......................... 89

1. Goals for a Causal Theory in Service of Explanation................................................. 90
   1.1 The Causal Nexus as Ontic Constraint and Target.............................................. 92
   1.2 The Case of General Explanation........................................................................ 96
   1.3 Summary of the Causal Theory’s Role.................................................................. 98

2. Elementary Causal Processes....................................................................................... 99
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1 Elementary Causal Processes – Definitions</td>
<td>99</td>
</tr>
<tr>
<td>2.2 Simple Physical Examples</td>
<td>102</td>
</tr>
<tr>
<td>3. The Role of Causal Properties</td>
<td>104</td>
</tr>
<tr>
<td>4. Composite Causal Processes</td>
<td>107</td>
</tr>
<tr>
<td>4.1 Composite Causal Processes – Definitions</td>
<td>109</td>
</tr>
<tr>
<td>4.2 Examples</td>
<td>114</td>
</tr>
<tr>
<td>5. Comparisons and Objections</td>
<td>115</td>
</tr>
<tr>
<td>5.1 Comparison to Cartwright</td>
<td>116</td>
</tr>
<tr>
<td>5.2 Comparison with Wimsatt</td>
<td>118</td>
</tr>
<tr>
<td>5.3 Objections</td>
<td>121</td>
</tr>
<tr>
<td>6. Mechanistic Explanation</td>
<td>125</td>
</tr>
<tr>
<td>6.1 Mechanisms and Productive Causation</td>
<td>125</td>
</tr>
<tr>
<td>6.2 Mapping Mechanisms</td>
<td>128</td>
</tr>
<tr>
<td>6.3 Other Critiques of “New Mechanism” Accounts</td>
<td>134</td>
</tr>
<tr>
<td>CHAPTER 5: THE COMPATIBILITY OF CAUSAL PROCESS THEORY AND QUANTUM MECHANICAL EXPLANATION</td>
<td>138</td>
</tr>
<tr>
<td>1. Quantum Worries for the Process Theorist</td>
<td>139</td>
</tr>
<tr>
<td>2. Propensities as a Bridge from QM to the Revised Process Theory</td>
<td>148</td>
</tr>
<tr>
<td>3. Comparison to the Orthodox Interpretation and the Challenge of Entanglement</td>
<td>155</td>
</tr>
<tr>
<td>4. Other Interpretations and the Question of Compatibility</td>
<td>161</td>
</tr>
<tr>
<td>4.1 GRW and Dispositional Ontology</td>
<td>162</td>
</tr>
<tr>
<td>4.2 Bohmian Mechanics and the Reducibility of Dispositions</td>
<td>165</td>
</tr>
<tr>
<td>4.3 The Many-Worlds Approach and Causal Explanation</td>
<td>168</td>
</tr>
<tr>
<td>5. Conclusion</td>
<td>171</td>
</tr>
<tr>
<td>Appendix: Salmon’s Definitions</td>
<td>172</td>
</tr>
<tr>
<td>CHAPTER 6: AIM THEORY AND THE INTERACTIVE CONCEPTION OF CHEMICAL BONDING</td>
<td>174</td>
</tr>
</tbody>
</table>
1. Conceptions of the Chemical Bond in the Wake of Quantum Theory ...................... 175
2. Overview of the AIM Theory .................................................................................. 181
3. Interpreting the AIM Conception of Atoms and Bonding ...................................... 187
4. The Question of Spatial Localization ..................................................................... 191
5. An Alternative Interpretation of ρ ......................................................................... 199
6. Conclusion ................................................................................................................. 203

Appendix: Estimation of ρ through X-Ray Diffraction ................................................. 204

CHAPTER 7: THE CAUSAL EXPLANATORY CHARACTER OF EVOLUTION MODELS .......................................................... 207

1. Introduction .............................................................................................................. 207
2. Causal Controversies: Finding the Right “Level” and the Right Concept .............. 212
3. Productive Population-Level Interactions ................................................................ 216
4. Deploying the Dispositional Causal Process Framework ....................................... 221
5. Interpreting an MS-model ...................................................................................... 226
6. A Brief Look at Four Pillars .................................................................................... 233

BIBLIOGRAPHY ............................................................................................................. 240
LIST OF TABLES

Table 7.1: Beginning of year census by type (annual fitness values derived) ..................228
Table 7.2: Type frequencies and relative fitness (annual fitness values derived) ..............229
Table 7.3: Type frequencies and relative fitness (mean fitness values estimated) .............230
LIST OF ILLUSTRATIONS

Figure 4.1: Etiological and constitutive aspects of explanation…………………………..108

Figure 4.2: Two causal processes form a composite via repeated interaction……………….111

Figure 4.3: Interactions of composite processes……………………………………………….113
Chapter 1: Introduction

1. Scientific Understanding and the Workings of the World

A primary goal of science is to increase our understanding of the natural world. Scientists seek to describe, predict and control phenomena, but providing explanations is crucial for understanding. In one formulation, to understand a phenomenon is to “grasp” an explanation of it: an achievement that has both a cognitive and an epistemic dimension.\(^1\) A subject gains mastery over the contents of an explanation, which in turn convey something about the world.

Explanations themselves may take many forms, reflecting, among other things, the particular goals of the scientist/research program and the identity of the intended audience. But to explain a natural phenomenon is generally to show its place in a structured context. The cognitive dimension of understanding, then, requires that the audience use concepts and reasoning processes to master this structure. And while there is an active debate regarding the possibility of non-causal explanations in science, in what follows I assume that the relevant structure is at least partly causal in character. For causal explanations, I conclude that understanding requires that the conception of causation invoked in a scientific account of worldly structure be broadly shared by the cognizing audience.

---

\(^1\) See Strevens (2013) and Grimm (2006).
What is this conception? Does it differ across disciplines or according to the various theoretical terms and representational devices that are employed? In this work I offer a theory of causal structure and argue that it underpins successful explanations throughout the natural sciences. The theory begins with a causal concept that is also used in non-scientific contexts: that of production (or “productive causation”). This concept can be sketched as bringing about change via a connection. I then refine and extend this basic notion into a fuller conception of nature’s workings—a network of productive causal connections. I argue that the theoretical entities and properties utilized in a good explanation map onto a part of this network (or “nexus”). Here, the project is motivated by the fact that scientific investigations of a particular phenomenon may deploy a variety of models. Yet despite the deliberate use of abstraction and idealization, when models are used to explain a phenomenon we can expect a correspondence of at least some of their features to the causal structure of the target system of interest.\(^2\) I believe this insight generalizes to non-modeling contexts. The role of the causal nexus described by the theory is to provide both the ultimate target for natural science explanations as well as a constraint on their successful formulation. Scientific understanding is possible when the causal elements of an explanation map onto the world’s productive causal structure.

It is controversial whether understanding (in its epistemic dimension), like knowledge, requires truth. Some argue that understanding is closely related to knowledge, or if distinct must still be factive in character. Other philosophers disagree, maintaining that understanding need not be factive.\(^3\) While I won’t provide a detailed epistemological

\(^2\) See the discussion of target-directed modeling in Weisberg (2013).
\(^3\) Grimm (2006) sees understanding as a “species” of knowledge, responding primarily to Kvanvig (2003) who argues they are distinct but sees understanding as factive. Elgin (2007), in contrast, argues scientific understanding need not be factive. Strevens’s (2013) account includes the criterion that understanding requires a “correct” explanation, allowing a limited role for idealization. Potochnik (2015), on the other
account, the present approach implies that strictly false explanations can lead to understanding if their elements map onto broadly similar features in the causal structure of the world. Related to this point, the present project does not depend on a presumption of scientific realism: instead I assume a stance of realism only about the existence of an underlying causal nexus.4

Importantly, in order for the causal nexus to play its role, it must encompass an interweaving of both stability and change. Scientists investigate the composition of persisting entities as complex systems. It follows that the causal structure underlying scientific explanations must provide a target for mapping theoretical entities as well as the changes in properties they undergo.5 This is a topic less explored by many causal theories, but critical to the task. Accordingly, the theory includes a causal account of composition.

While the details differ, the present project owes a debt to Salmon’s (1984) causal approach to explanation. And like Salmon’s account, it inherits a long legacy of applying broadly mechanical approaches to explaining natural phenomena. At the same time, it also includes a connection to a different tradition, exemplified in recent decades by Cartwright (1983, 1989), that sees capacities or dispositions as targets for scientific investigation. The result—dispositional causal process theory—brings together these two strands of philosophical thought to offer a new account of causation and its role in science.

---

4 See discussion below, especially Chapter 3, section 5 and Chapter 4, section 3. I note that some approaches to scientific realism rely on realism about causal structure in their formulations.

5 Here I draw inspiration from work on complex systems due to Simon (1996) and Wimsatt (2007).
2. Advantages over Competing Approaches

Several philosophers of science offer causal theories in the context of scientific explanation. Here, an important distinction divides the field. As discussed in Chapter 2, both philosophical examination of cases (e.g. Hall 2004) and psychological evidence provide strong support for the existence of two broad concepts of causation: production and difference-making. When it comes to scientific explanations, we see evidence that both concepts are put to work. Production includes the idea of intrinsic spatiotemporal connection, and as such it plays a primary role in explanations across the natural sciences. These are explanations that seek to account for phenomena by tracing antecedent causes via a series of interacting entities (such as elementary particles, fields, molecules, cells, or organisms). Seeking this kind of explanation is often described by scientists as a search for mechanisms. The difference-making concept, in contrast, is employed when spatiotemporal causal connections cannot be traced using available techniques, or when a pragmatic decision is made that the additional explanatory value of detailing them can be foregone.6 Difference-making involves comparing scenarios where a candidate cause is present and absent. For instance, a targeted mutagenesis experiment (e.g. a gene knockout) may have value for inferring a causal explanatory relationship between a gene and the expression of a certain phenotypic trait. On the other hand, when a detailed account is sought about how the gene’s presence brings about the development of the

6 For this reason, difference-making will play the leading role in social science contexts.
trait, a chain of productive connections will be traced.\textsuperscript{7} It is this latter kind of explanation that is underpinned by theory of the world’s causal structure presented below.

I argue that productive causation is the primary causal concept in this sense: it is the concept we use when envisioning the world’s basic fabric. Where gaps in our understanding of this causal web exist, we look to fill them in. Productive causal structure plays a ubiquitous role in the formulation of natural science explanations in that it is assumed to underlie phenomena even when the option of explaining via a difference-making relationship is employed.\textsuperscript{8}

Of course, advocates of particular difference-making theories, of which the most prominent is Woodward’s (2003) manipulation-based approach, see the potential for application across the board. In part this ambition has been encouraged by the shortcomings of earlier production-based theories (see below). And, indeed, there are places where a difference-making causal relation can play a role in natural science explanations even beyond the pragmatic use sketched above. As discussed in Chapter 2, when causal claims are made in a context that includes a norm (such as a statistical norm or a norm of proper biological function), there is potential for a difference-making relation to do work that a production relation is not suited for. Allowing for this caveat, a theory based on the concept of production has the advantage of greater accuracy in accounting for how natural science explanations (that seek to trace connections and describe mechanisms) lead to understanding.

\textsuperscript{7} In some cases the presence of back-up pathways means a gene is a productive cause but not a difference-maker with regard to the phenotypic effect (causal preemption).

\textsuperscript{8} This is not to say that difference-making relationships can be reduced to productive relationships (see discussion in Chapter 2).
The theory offered in this dissertation has most in common with Salmon’s mark-transmission causal process theory. As discussed in greater detail in Chapter 3, Salmon’s theory faced a number of criticisms, but for present purposes the most important involve the difficulty in applying the theory to complex systems. A process is defined only in terms of “structure” or “characteristics.” A “mark” (a change in structure or characteristics) is introduced by external interactions and is transmitted through space-time in the absence of further interactions. Now, a biological entity (for example) experiences a steady stream of changes through time, due both to internal processes as well as environmental influences. Which changes are marks as opposed to those driven by internal processes? Also, what properties identified by biologists count as characteristics? Salmon has no theory of properties, per se, and despite a helpful discussion of the notion of constitutive explanation, has no account of composition to apply to complex systems. The theory presented in Chapter 4 provides greater scope through its ability to treat the full range of scientific entities and properties invoked in the natural sciences. This advantage also applies when the theory is compared to the conserved quantity versions of process theory due to Dowe (2000) and the later Salmon (see chapter 3, section 5): these are only readily applicable to explanations in physics.

An interesting consequence of this greater scope is its potential implication for debates about the unity of the sciences. Some see the existence of myriad incompatible theories and models across the sciences as evidence for an underlying pluralism or disunity (Cartwright 1999). I attempt to show through argument and examples that the same causal theory plausibly underpins explanations across physics, chemistry and biology. The different disciplinary terminologies and methods, guided by a variety of explanatory norms, obscure the fact that there is one “ontic constraint” that is shared by good causal explanations. On this point, the present approach is in sharp contrast to that
of the so-called “new mechanists” (discussed in chapter 2 and in more detail in chapter 4, section 6) where a very general framework is filled in by theoretical entities and “activities” drawn from the sciences themselves (Machamer, Darden & Craver, 2000). The absence of a theory of productive causation underpinning their framework means the new mechanists cannot offer an account of why some mechanistic explanations can lead to understanding while others may not.

Finally, while I do not present anything like an exhaustive account of the other virtues and norms that govern explanatory work in the various disciplines, the wide scope of my theory of the ontic constraint provides a starting point to explore the role these play in different contexts. For this reason, the approach offers a promising platform for philosophical investigation of these aspects of scientific explanation.

3. Implications for the Metaphysics of Causation

The theory offered here is not intended as a definitive metaphysical account of causation. But assuming it describes the world’s causal structure as it figures in successful natural science explanations, one might ask about its metaphysical import. Of course, one could take a narrowly naturalistic view of such questions. Perhaps what the theory shows is that natural science explanations mainly succeed by invoking a particular psychological concept/reasoning apparatus. This might, in turn, account for the phenomenon whereby one feels a subjective grasp of such an explanation. But if one takes seriously the idea that these explanations increase our understanding of the world as discussed in section 1, then the theory of causation I develop may have something to say about the world’s real structure.
As I see it, the productive causation inherent in our natural science explanations follows from a special kind of engagement of humans with the world (see also chapter 3, section 6). It reflects an organized activity whereby we, causal entities in our own right (according to the theory), intervene and manipulate the world in carefully planned ways, while the world “pushes back” on us in return (note here that the traditional focus on “observation” is misleading: observing, say, an astronomical phenomenon is an unusually passive activity, but strictly speaking all causal interactions involve mutual change). This engagement with the world might well give us defeasible impressions of the character of the real causal structure we participate it.

Before proceeding, I want to note again the distinction between this suggestion—that we can infer from science something about real causal structure—and traditional scientific realism (pertinent discussions are in chapter 3, section 6 and chapter 4, section 3). Scientific realism involves the truth of particular models/theories, and this involves commitments that go beyond the present project.

Subject to this clarification, can it be argued that my approach to causation picks out a correct or approximately correct fundamental metaphysical theory? It seems we might quickly answer this question in the negative: at a minimum this would be because metaphysicians are typically interested in modal truths that have wider scope than that encompassed by a theory inspired by the scientific investigation of the contingent actual world (more on this below). But I will argue that metaphysical theorizing about causation should be constrained by a theory that underpins good natural science explanations.9

---

9 Here, the type of metaphysics I have in mind is that which theorizes directly about the nature of the actual world and other possible worlds using various kinds of evidence, rather than traditional conceptual
This claim needs to be fleshed out, because it differs from some discussions of the constraints on metaphysics posed by science and/or philosophy of science. Paul (2012) defends metaphysical theorizing in part by arguing it has a different subject matter than science:

Metaphysics investigates, for example, the nature of laws, naturalness, causation, persistence, and properties. Science assumes that we have a pretheoretical grasp on these natures, and then investigates the instances of these natures: it tells us which laws obtain in the natural world, which natural kinds there are and how they are ordered, which other properties and relations are actually instantiated, which objects persist, and what causes what (and how). (Paul, 2012, 6)

Paul claims that a focus on science as a guide to metaphysics is a mistake because “scientific theorizing uncritically assumes the very organizing principles and deep general truths that metaphysics is concerned to prescriptively develop and understand (Paul, 2012, 6, emphasis original). Paul does also offer a limited caveat to these assertions, saying that “science still acts as a constraint upon metaphysics—the metaphysician should want her theory of the world to be consistent with accepted scientific theories of the world—but it should not preemptively define the role or concepts of metaphysics (Paul, 2012, 7).” Elsewhere, in a conciliatory vein, she also says there is a “vague boundary between science and metaphysics” that can provide a “rich opportunity for philosophers of science” to “evaluate the plausibility of metaphysical theories that bump up against the domain of the empirical (Paul, 2012, 8).” In the present case, I argue that my investigation of how causation underpins natural science explanations does not simply import “naïve and
uncritical scientific suppositions (Paul, 2012, 6).” Instead, it is a critical project in philosophy of science and as such it has implications for metaphysics.

By articulating a theory underpinning causal explanations in physics, chemistry and biology, I show that more can be said about the “pretheoretical grasp” that Paul envisions is embodied in the practice of science. Importantly, science is not just about describing and labeling “instances” of what is found in the world. In addition to description (and prediction and control), an aim of science is to provide understanding. I believe that it succeeds to some degree in this objective, and causal explanations are a principle means by which this is accomplished. It is implausible to think that successful explanations have nothing to tell us about the “nature” of causation. The role of philosophy of science as it relates to this particular metaphysical subject is not just to keep watch out for conflicts between the work of metaphysicians and accepted scientific theories. In this case (at least), it can explicate the worldly causal structure that is implied by the scientific understanding we achieve. I submit that any metaphysical theory should be consistent with the theory of the causal nexus presented here.

This conclusion still allows for many degrees of freedom for the metaphysician (some of these are discussed in Chapter 4). It is even plausibly consistent with reducing causation to wholly non-causal theoretical elements—something the present approach does not attempt (although the history of such efforts is grist for pessimism about achieving that goal). Also, my theory has less to say about what Paul describes as “one especially important sort of general truth that metaphysicians are interested in uncovering: modal truth (Paul, 2012, 7).” To see that the approach does not encompass all possible worlds that may feature causal relations (broadly defined), it is enough to point out that it fails to incorporate causal action-at-a-distance. If one can conceive of a
possible world with this feature (and it seems fairly easy to do so), then, assuming
conceivability is a guide to possibility, the present theory fails to embody metaphysical
necessity. It describes our actual world, other “physically possible” worlds, and perhaps
some further “nearby” worlds that share a similar productive kind of causal structure.
Perhaps there is a “deeper” metaphysical theory that can be shown to account for our
world’s structure while also being plausibly true of all possible worlds that feature
“causation.”

On the other hand, I will note that the pursuit of such a goal is complicated by the
evidence for conceptual pluralism. In particular, the notion of difference-making
causation, which is naturally explored philosophically via notions of counterfactual
dependence, is plausibly independent of the concept of productive causation that is
primarily invoked in natural science explanations. Now conceptual pluralism need not
rule out metaphysical monism. On the other hand, while I do not argue for this claim, it
is plausible that while difference-making figures in our reasoning (and I note it can
readily accommodate action-at-a-distance), it has no bearing on the fundamental
structure of our world. In contrast, the theory offered here has the advantage of greater
consistency with the explanations offered in the scientific endeavor to understand the
actual web of nature.

4. Overview of Chapters

The chapters that follow can be divided into two parts. Chapters 2 through 4
develop the causal theory and describe its relationship to scientific explanation. Chapters
5 through 7 test the theory by showing how it applies to and illuminates challenging philosophical topics and controversies involving the natural sciences.

In Chapter 2, I examine how concepts of causation enter into the philosophy of scientific explanation. I describe and defend the thesis that there are two independent concepts of causation (production and difference-making) and discuss how these relate to prominent accounts of explanation. I argue that production is the notion most generally applicable in the natural sciences, although I also describe a proper role for difference-making, particularly in biological contexts. I then note that production theories of causation are less well developed than difference-making approaches in the recent literature, revealing a need for a new account. The chapter concludes with desiderata for a successful theory of production.

Chapter 3 sets the stage for developing a new theory by critically reviewing the most prominent production-based approach: the causal process theories due to Salmon and Dowe. The biggest concern for these theories is their adequacy for supporting explanations. Another set of historical criticisms concern whether the theories are consistent with a certain kind of philosophical empiricism. I argue these latter concerns are misdirected, and discuss how a revised version of process theory in the spirit of Salmon’s approach can achieve success.

Chapter 4 sets out the new theory of causal structure and describes its role in explanation. In the new approach, causal processes are described in terms of dispositions toward interactions. Interactions, in turn, involve mutual manifestation of dispositions of two or more processes. I offer examples to show how the elements of an explanation map onto this conception of the causal nexus. I then present an account of composite causal processes: these correspond to the hierarchy of theoretical entities that appear in
explanations across physics, chemistry and biology. I compare the theory to other accounts and argue in particular for its advantages over the approach of the “new mechanists.”

Chapter 5 explores the consistency of the account with non-relativistic quantum mechanics. I review why this issue was a particular worry of Salmon’s, then describe how the revisions to Salmon’s approach that feature in the new theory make it more suitable for supporting quantum mechanical explanations. The key point is that while causal interactions take place at space-time locations, there is no requirement for spatial localization of dispositions between interactions. I discuss how the mapping of explanations onto the causal structure makes interpretive commitments about the nature of quantum measurements, and draw comparisons to some of the other interpretations in the literature.

Chapter 6 discusses chemical bonding. Here the context is a debate over how the traditional notion of bonding should be reconceived in the wake of modern quantum models of molecular systems. I review an approach called the quantum theory of atoms in molecules, due to R.F.W. Bader and colleagues. I argue that this theory embodies an attractive interactive conception of bonding that is compatible with the theory of composite causal processes presented in Chapter 4.

The final chapter examines a controversy at a much different scale of natural systems. Modern evolutionary theory features mathematical models of population change. Some philosophers argue that these models can be used to explain outcomes but the kind of explanation involved is statistical rather than causal in character. I argue the models are causal when used to explain an evolutionary outcome in an actual population. The key to this result is a proper description of the target system’s causal structure: I show
how a population and its constituent organisms map onto the causal nexus. Evolutionary models explain by virtue of elements that correspond to features of population-level causal structure.
Chapter 2: Two Causal Concepts in Explanation

Philosophical examination of cases and psychological evidence support the existence of two concepts of causation: production and difference-making. These concepts, in turn, are employed in various accounts of scientific explanation. Examining the relationship between the concepts and their applicability in different contexts is crucial for evaluating competing philosophical approaches to explanation. I argue below that we have strong evidence for the independence of the concepts, while there is little support for asserting the priority of one over the other. The conditions governing their applicability will nonetheless differ: there are good reasons to give production relations precedence in physical settings, while difference-making relations become salient in normative and teleological or functional contexts. Both production and difference-making should often figure in biological explanations. In reviewing contemporary approaches to explanation, however, it becomes evident that there is an imbalance in the actual roles played by the two causal concepts. There are prominent, well-developed accounts based on difference-making. But recent approaches to mechanistic explanation, while invoking the concept of production, do not provide a unified theory to support their frameworks. A new theory of production is needed in order to provide an improved understanding of explanation in the natural sciences.

This chapter is organized as follows. Section one summarizes the case for the existence of two independent causal concepts and briefly describes their role in prominent theories of scientific explanation. Section two reviews the debate over priority and concludes neither concept has a good claim to priority or fundamentality versus the other. In section three I consider conditions of applicability for the two concepts in
explanation: production is generally aptly applied in physical contexts, but difference-making becomes salient in teleological and normative settings. Section four concludes by noting the need for a new account of production to support scientific explanation and lists desiderata for such a theory.

1. Two concepts of causation and two approaches to causal explanation

A recent article presenting research on vertebrate limb development begins by contrasting two approaches to investigating morphogenesis (the development of organ shape):

In the last few decades, developmental biologists have identified key genes necessary for organ development and performed functional analyses of these genes by forward and reverse genetics. These genetic analyses have enabled us to determine the genes that are required for the morphogenesis of tissues. While these molecular approaches have identified static one-to-one correspondences between gene function and final phenotype, such correspondences are not enough to elucidate the mechanisms of the morphogenetic process itself. How do stage- and region-specific cellular behaviors, regulated by different molecules, cause tissue-level deformation and determine the overall shape of an organ? (Suzuki & Morishita, 2017, 108)

The first approach examines how the presence or absence of genes makes a difference to the final shape of an organ. In forward genetics, a mutation is observed in the phenotype and the associated genetic change is found, while in reverse genetics, a change in genome
is experimentally introduced: for instance, a gene may be rendered ineffectual (gene knockout) and the associated phenotypic outcome observed (see Brooker, 2018, 660-1). The authors then sketch another approach (which they aim to pursue), which is to “elucidate” the mechanisms involved in morphogenesis: this means providing a detailed account of how molecules and cells interact and change in time and space to produce the final shape. This discussion highlights the presence of two kinds of explanatory strategies used in the natural sciences. These, in turn, invoke two distinct concepts of causation.

1.1 Production and difference-making

For philosophers, the traditional goal of providing a conceptual analysis of causation has proven elusive. Conceptual analysis itself has perhaps receded as a central focus of the discipline, but related successor projects (e.g. attempts to provide a fundamental metaphysical theory, even at the cost of revisions to the folk concept) have similarly faced stubborn obstacles. It has been proposed that one of the lessons learned from the various attempts is that our concept of causation is not univocal (Cartwright, 2004, Hitchcock, 2007a, Godfrey-Smith, 2009, cf. Anscombe 1971). In particular, Ned Hall (2004) offers a strong case for the existence of two distinct concepts of causation. The first, difference-making, concerns relevance or dependence relations, while the second, production, emphasizes spatio-temporal connection and the generation of change.10

Philosophers of causation test theories by applying their intuitions to cases. Assume Suzy throws a rock at a window and it shatters. In this case, a sketch of a theory

---

10 I have substituted “difference-making” for Hall’s “dependence” to connect the discussion to the widespread use of the former term. It should be noted Hall himself is not committed to the two concepts thesis, continuing to explore avenues toward a unified reductive account. Hall and L.A. Paul concede, however, that the prospects are “dim” (Paul & Hall, 2013, p. 249).
of causation-as-production (such as a simple transference account) might say that Suzy’s action brings about the shattering via the connection provided by the flying rock, which possesses momentum provided by the throw. Alternatively, a sketch of a difference-making (DM) theory (a simple counterfactual account) would note that the shattering depends on Suzy’s throw, since if the throw had not occurred, then the shattering would not have happened. Production and difference-making are both compatible with simple scenarios like this one. In other cases, however, intuitions appear to track only one of the two concepts. The seeming inability of theories based solely on either difference-making or production to provide a complete analysis prompts Hall’s proposal.

The most prominent DM theories have been developed in terms of counterfactual dependence (Lewis 1973, 2000; Collins, Hall & Paul 2004). The effort to analyze causation in this way grew into a large philosophical research program over several decades, but has fallen short of a successful analysis. A central concern surrounds scenarios such as so-called “late preemption”. Here Suzy throws a rock at the window, and Billy does likewise a split-second later. Stipulating that either rock alone has the ability to shatter the window, Suzy’s throw makes no difference to the outcome (if she had not thrown, the window still would have shattered). Yet our intuition is that in the actual case, Suzy’s throw did cause the shattering (Hall, 2004, 235). Here the production notion easily accounts for our intuition, while difference-making fails. An important distinction between the two notions is that production appears to be intrinsic to the

11 Another class of difference-making theories focuses on changes in the probability of one event given another, but these approaches share relevant features with counterfactual strategies as it relates to the present discussion. Regularity theories (which won’t be discussed at length here) might also be seen as difference-making accounts – but ones that also attempt to capture at least part of the notion of production with necessitation/sufficiency relations and/or laws taking the place of connections between cause and effect (Mackie 1974, Davidson 1967).

12 Paul & Hall (2013) offer an in-depth review.
sequence (Suzy’s throw, the flying rock, the shattering), while difference-making requires attention to alternative scenarios involving extrinsic factors. In response to cases like this, various revisions to the counterfactual account have been proposed. For instance, one might argue the preempted shattering event is distinct from the actual one, although more careful counterexamples put pressure on this gambit. Other strategies add rules for choosing counterfactuals that hold certain facts unchanged (such as the fact that Billy’s rock did not, in fact, hit the bottle) or otherwise make the choice based on facts about context. While some of these latter strategies are highly useful in causal modelling contexts, various objections have undercut them as paths to a complete analysis.

There have also been efforts to analyze causation via a theory of production, notably process and transference accounts, which feature interactions involving persisting objects or entities (Fair, 1979, Salmon 1984, 1994, Dowe 2000; cf. Ehring 1997). These theories face challenges specific to their formulation, but a general concern with production theories has been the difficulty of accommodating cases where connecting processes are absent. Philosophers have pointed in particular to our causal intuitions about scenarios involving omissions and preventions. The DM concept more readily accommodates the intuition that, say, Billy’s failure to water his houseplant caused it to die. Here defenders of production theories make a number of points. First, they point out that there is a corresponding story to be told in productive terms: one can give a positive account of the death of the plant in terms of the processes that ensue in the absence of watering, and we can do the same in describing the movements of Billy in his

---

13 Related to this point is the fact that production accommodates the intuition that causation is transitive, while difference-making does not (transitivity was stipulated by Lewis in his 1973 paper).
14 A stumbling block is the need to evaluate these more complex counterfactuals without implicitly incorporating causal information (see Paul & Hall, 2013, pp. 111-124).
15 Although fleshing out a counterfactual theory that handles cases of omission and prevention is also problematic (see Paul & Hall, 2013, pp. 212-214).
non-watering activities. Dowe argues that omissions and preventions can be understood in terms of positive causal processes coupled with counterfactuals describing other, absent, positive causal processes (such as, in this case, the successful watering). Dowe encountered opposition to his proposal based on his refusal to consider these relations involving absences to be genuinely causal. Certain cases, such as those involving “double prevention” are particularly salient in this debate. Hall’s example features an air battle: Suzy, flying a bomber plane, successfully bombs her target; Enemy fighter plane would have prevented this by shooting Suzy down, if not for the fact that Billy (flying a fighter plane escorting the mission) succeeded in shooting Enemy down first. Many have the intuition that Billy’s action serves as a cause of the bombing.

All of these approaches have attempted to analyze what singular causation “is”, but it should be acknowledged that causal explanations do not simply attempt to track singular causal relations. Among other things, they attempt to describe general patterns that might guide broader conclusions about how things happen or why things change. For their part, regularity theories of causation give priority of place to general patterns, and reflect a tradition that stresses the epistemic role regularities play in identifying causal links. But while the drive to uncover generalizations enters into theorizing about causal explanations, there is an advantage in beginning with an assessment of singular causation at the fundamental conceptual level: if successful, other elements can be added in order to formulate an account of causal explanation. But with which causal concept should one begin?

16 Here counterfactuals are not being used to analyze causation itself, but to construct a derivative kind of causation (Dowe calls it causation*) to cover omissions and preventions (see Dowe, 2000, Ch. 6.).
17 Some have argued that problems with conceptual analysis reveal a need to change focus to an epistemological investigation of causation. See De Vreese (2006) and Williamson (2006, 2013).
Another, naturalistic, approach to this question might look to psychological evidence regarding our causal reasoning. Perhaps we should develop a causal theory that (at least in part) tracks this capacity. Interestingly, the psychological literature also supports the conclusion that there are two general causal concepts at work. One prominent strand of psychological research concerns the phenomenon of causal perception. Beginning with the work of Michotte (1963), researchers have identified a “launching effect”, whereby subjects ascribe causation to depictions of billiard-ball type collisions (typically featuring moving colored shapes). The perception of causation in these cases appears to be “salient, immediate, and irresistible (Scholl & Tremoulet, 2000, 302).” The phenomenon appears to be present in six-month-old infants (Leslie & Keeble 1987). While attributing a concept such as causation might seem characteristic of a higher-level cognitive activity, this sort of causal perception seems to depend “only on visual input, and not on higher-level cognition; you cannot, for example, choose not to see the classic launching effects as causal (Danks, 2009, 451).” Causal perception appears to be closely linked to the concept of production. Furthermore, it is plausible that causal perception, coupled with other elements of primitive object perception, underpins the understanding of mechanisms. This would expand the application of production from individual interactions to the broader mechanistic domain.

---

18 See, e.g. Spelke, Breinlinger, Macomber, & Jacobson (1992) and Baillargeon (1994).
19 However, as children are able to learn more about mechanisms independently of observation, they will encounter conflicts between the spatio-temporal cues present in causal perception and the underlying knowledge (for instance in experiments involving unseen apparatuses). In such cases younger children may incorrectly follow the evidence from, say, temporal contiguity, while older children (correctly) give priority to knowledge of the unseen mechanisms: see Schlottmann (1999), cf. Schultz (1982). This research is consistent with the thesis that “perceptual causality could promote rapid acquisition of mechanical knowledge” but may also point to the need for cognitive resources that make mechanism learning a “separable” process from causal perception (Schlottmann, 2001).
With regard to the DM concept, this appears to connect to another set of research programs that investigate into how causal inferences are made by subjects confronted with data involving co-varying candidate causes and effects. There are a large number of competing models of causal inference (see Gopnik & Schulz, 2007 and the summary in Danks, 2009, 451-455). These have included traditional associative learning models, and models which propose that subjects treat the degree of covariance as an indicator of the strength of causal powers (Cheng 1997). More recently researchers have employed causal models (particularly causal Bayes nets) in a variety of investigations, including the formulation of frameworks for causal learning in children (Gopnik, et al., 2004) and general causal inference (Griffiths & Tenenbaum, 2005). Here, the connection to how philosophers have theorized about difference-making is clearest, as these researchers are deploying formal DM frameworks.\(^\text{20}\)

1.2 The concepts at work in scientific explanation

Despite there being no agreed upon theory of singular causation, the business of developing frameworks for causal scientific explanations has proceeded apace. And those pursuing these projects rely on one of the two causal notions. Several prominent causal explanatory programs utilize difference-making. In developing these approaches, theorists take different stances on the relationship between the explanatory project and the unfinished work of understanding what causation is. For instance, Michael Strevens (2008) presents a difference-making-based methodology for causal explanation (the “karietic approach”), but is agnostic about the correct theory of causation. Importantly, he argues that whatever the right theory is, it is needed only to characterize a web of

\(^{20}\) This research approach, which is currently very prominent, is interesting in that a normative framework is being deployed toward a descriptive goal.
“causal influence” at a fundamental physical level, and difference-making explanations involving higher-level facts do not depend on the details. Other difference-making frameworks are also employed without attempting to address deeper issues. This is true of popular causal modelling approaches, for instance, which utilize formal patterns of dependence (Pearl 2000, Halpern & Pearl, 2005, Spirtes, Glymour & Sheines 2000). Some philosophers originally thought these frameworks would aid in the analysis of causation, but it is now generally conceded that their application requires background causal knowledge. Despite this, they are widely employed as methodologically valuable tools for a variety of purposes. The influential manipulation/intervention approach to causation, closely allied with causal modelling techniques, offers a more developed stance. As articulated by James Woodward, it defines causation in terms of counterfactuals that represent the effect that hypothetical ideal interventions or manipulations on one (type-level) variable have upon another given certain background conditions (Woodward, 2003). It is important to note that since the notions of intervention and manipulation are causally laden, the account is not attempting an analysis or reduction of causation. One reason for the influence of Woodward’s approach in the literature on scientific explanation is its integration of an illuminating (if non-reductive) theory along with a broad set of proposals for its application.

There has been another prominent set of explanatory programs advanced recently that rely to some degree on the concept of production: they focus on the identification and elucidation of mechanisms. The mechanistic approach has been explored by a number of authors, most often in context of biology.21 Here philosophers take as a

---

21 The causal process theories due to Salmon and Dowe are sometimes referred to as mechanistic, but here I am reserving the term for the work of recent authors especially prominent in the philosophy of biology (sometimes referred to as the “new” mechanists).
starting point the fact that scientists characterize their own work as a search for and description of mechanisms to explain phenomena. In this literature, the authors also do not attempt a detailed grounding of their frameworks in a theory of what causation is. In the case of Machamer, Darden & Craver (2000 – hereafter “MDC”), mechanisms are “entities and activities organized such that they are productive of regular changes from start or set-up to finish or termination conditions (MDC, 3).” Activities in particular are the “producers of change (MDC, 4).” Despite invoking the concept of production, the authors allow for a pluralism of entities and associated activities (described using terms drawn from the sciences themselves) without attempting to subsume them under a single causal theory. For instance, they list four kinds of basic or “bottoming-out” activities that typically figure in molecular biological mechanisms: spatio-temporal, electro-chemical, energetic, and electro-magnetic (MDC, 14). Importantly, extant theories of production, such as Salmon’s and Dowe’s process accounts, are not viewed as helpful in supporting the (new) mechanistic approach. This is principally because they are not viewed as having the resources to support explanations featuring the “diverse kinds of production” involved in biology (MDC, 7). Also, MDC choose not to take the path of describing entities as possessing causal properties, or having “capacities (Cartwright, 1989) or dispositions to act”, in favor of their focus on activities (MDC, 4). In the absence of a more developed causal theory, some authors have supplemented their accounts of how parts of mechanisms interact by drawing on Woodward’s difference-making approach

22 See, for examples, Bechtel (2006), pp. 1-5 (examples from cell biology), and Craver (2007), pp. 2-3 (neuroscience).
23 An exception here is Glennan’s (1996) approach that attempted to analyze causation in terms of mechanisms. But this account ultimately relied on an unanalyzed notion of causal law. Glennan (2017) recently explores other possible ways of understanding how mechanisms could play this metaphysical role.
25 For exploration of the notion of non-specific productive activities, see Machamer (2004), Bogen (2008), and Waskan (2011).
(Craver 2007). Still, the notion of production is clearly at work in the formulation of mechanist explanatory frameworks, even if scientifically defined activities are used to fill out its role.

Given the competing approaches to explanation on offer, it is important to examine the causal concepts that lie at their foundation. Are the two concepts really independent? If so, are there distinct contexts in which they should be applied? On the other hand, might it be that one of the concepts is more primitive or basic, while the other is derivative? If so, would this fact favor an approach to explanation based on the more fundamental concept?

2. The debate over explanation and the question of conceptual priority

In the recent philosophy of science literature, DM approaches to explanation (mainly Woodward’s) and mechanistic approaches have often been pitted as rivals, particularly on the battleground of biology. The most salient points in the debate seem to be the following: Mechanists tend to view their approach as providing relevant details that DM frameworks cannot provide (such as spatio-temporal or structural information). On the other hand, mechanistic accounts are accused of lacking an

---

26 See also Glennan (2010), who argues for using two kinds of causation (production and “relevance”) as part of the description of mechanisms.

27 I note that the MDC (2000) article uses the term produce (or cognates) several dozen times to describe mechanistic explanation. For further discussion of the new mechanists’ approach, see Chapter 4, section 6.

28 For a recent example of this kind of argument, see Kaiser (2016). Another potential difficulty for DM approaches is the existence of redundant or back-up causes, which are common in biological contexts. In such cases, a DM relation may falsely describe the situation—see the discussion in Love & Nathan (2015, pp. 766-768). With that said, moving DM relations from the token to the type level may often address the issue.
adequate theoretical grounding absent recourse to counterfactual-based resources. The nature of “activities” and of the connection between entities or parts is an ongoing focus of concern, as in a 2004 evaluation by Psillos:

I take this to be a crucial problem of the mechanistic approach. In a sense, this approach fills in the ‘chain’ that connects the cause and the effect with intermediate loops. But there is still no account of how the loops interact. Here, it might well be the case that the most general and informative thing that can be said about these interactions is that there are relations of counterfactual dependence among the parts of the mechanism. Even if we posited activities, as MDC do, we would still need counterfactuals to make sense of them, as we have just seen. (Psillos, 2004, 314-315).

The decomposition of mechanisms into parts or entities on the one hand and activities on the other is also a source of worry: any lack of clarity in the nature of activities impacts the correlative notion of “entities.” For instance, the stabilization of biological entities arguably involves causation, just as much as the identified activities do.29

The debate has continued, with Woodward in particular responding to the idea that mechanistic approaches offer more detail by elaborating his own framework for application to biological phenomena (Woodward 2010). Woodward also acknowledges that spatio-temporal structure or organization is important and his DM framework needs to incorporate this (Woodward, 2011a, 2013). On the other hand, while the literature on

29 See Dupré (2013). For a defense of the dualism of entities and activities, see Illari & Williamson (2013). Franklin-Hall (2016) has recently offered a sharp critique of the theoretical underpinning of mechanistic approaches, including the causal notions employed and the process of identifying parts. The Dupré and Franklin-Hall critiques are further discussed in Chapter 4, section 6.
mechanistic explanation has expanded rapidly, the program still faces concerns regarding how its basic causal elements are to be understood.

In evaluating this debate, it would be good to know whether there is a reason to think that either the concept of production or of difference-making is more fundamental than the other. If production is more fundamental, then the failure of extant mechanistic approaches to satisfy critics like Psillos might be due to the fact that they lack an adequate theoretical bridge between their frameworks and the concept that inspires the causal element in their explanations. On the other hand, if there is no reason to think production is more fundamental, then we must view the debate in a different light: perhaps difference-making is more fundamental, favoring Woodward and other DM advocates. Or perhaps the two concepts are independent, but with neither having a priority over the other. In this case other criteria must be employed to judge the applicability of the alternatives.

2.1 From the priority of the physical to the priority of production?

Production applies most clearly when actual worldly connections between phenomena appear to be present. Our notion of productive connections begins with sequences involving macro-sized objects coming into contact. Authors such as Salmon and Dowe refine the folk physical notion by explicating production using scientific descriptions of particles and fields exchanging conserved quantities (e.g. energy and momentum). For a first pass, then, one might presume that production is the appropriate concept to apply to natural phenomena, which after all are rooted in the physical world. The DM concept, on the other hand, does not require an actual connection between

---

30 See the discussion above of the phenomenon of causal perception.
events. While it can be applied to scenarios involving physical entities, this comes with
the important caveat that confounding issues related to preemption or over-
determination must be avoided. On the other hand, DM is potentially applicable in
situations where we cannot readily trace physical connections. As such, it can play a role
in other domains, such as the social sciences, or in discussions of mental causation.\textsuperscript{31}

These differences may prompt an argument for the priority of production. It is
widely thought that physical facts provide a basis or foundation for all other facts. If so,
then analogously, the causal facts about the physical world provide the foundation for all
causal facts.\textsuperscript{32} And if a theory based on the concept of production is the best way to
construe physical causation, then, production is ultimately at the foundation of all causal
facts including facts about difference-making. Related to this line of reasoning, Alyssa
Ney recently advanced an argument for “foundationalism” regarding the relationship of
physical causation to difference-making:

In my view, the difference-making facts depend on the facts of physical causation;
whether or not one event makes a difference to another depends not only on the
obtaining of certain counterfactual or probabilistic dependencies, but in addition
on facts about the obtaining of a physical causation relation. (Ney, 2009, 740)

As a premise, Ney assumes that something like Dowe’s conserved quantity theory
describes physical causation. She then argues that difference-makers can be viewed as
subsets of physical causes, selected to aid goals of explanation or prediction. One’s

\textsuperscript{31} Difference-making causation might even be used to describe causation in conceivable worlds that differ
physically from our own, for instance where action-at-a-distance occurs.

\textsuperscript{32} I will not be discussing the line of argument that denies that there is such a thing as physical causation—
famously found in Russell’s “On the Notion of Cause” (1913). Ney (2009) includes a critique of this view.
For another recent discussion see Frisch (2014).
favorite difference-making framework may serve this interest, as long as the difference-makers selected are also physical causes. Thus, the causal nature of difference-makers depends on their also being physical causes.

The defender of this view has to respond to the obstacle represented by our intuition that omissions and preventions can be causes. Ney discusses strategies (including those mentioned above) that the advocate of physical causation can turn to here, but it remains a topic of controversy. Another challenge is posed by Glynn (2013), who objects to Ney’s paper by pointing out that her argument requires that physical causation does not itself involve difference-making facts. Given that theories such as Dowe’s and Salmon’s have been subject to various criticisms, a skeptic can assert that a role for difference-making at the physical level has not been ruled out. These challenges may still create an opening for those who wish to argue for the priority of DM as a causal concept.

Even if Ney’s argument is accepted, it does not suggest that the DM concept is idle or somehow collapses into the production concept. Rather, Ney sees it as a tool for isolating relevant causes from among multiple possibilities. Here, it is helpful to compare how Streven’s account utilizes difference-making. If causation at the fundamental level of physics worked according to Dowe’s theory (Streven takes this as an open possibility, but is officially non-committal) Streven’s would assert that his difference-making methodology would still be the right method to use for high-level

---

33 On this point, critics will note that Salmon’s original process theory (Salmon, 1984) employed counterfactuals (see discussion in Chapter 3).

34 Assigning this role for DM is in keeping with a criticism of extant physical theories of causation—that they lack the resources to support “higher-level” explanatory tasks. Critics focus on cases where there may be energy/momentum transfer between two objects, but it is irrelevant to explanation (see discussion in Chapter 3).
explanations. What would this mean for the question of asserting a foundational role for physical (production-style) causation? For Strevens, Dowe’s theory would indeed describe the world’s “single, homogeneous causal structure.” For its part, the high-level DM relation is arrived at through a particular process of abstraction (the kairetic method), which means “difference-making summarizes pertinent information about the fundamental stuff of causation (Strevens, 2013, 314)”.

As in Ney’s proposal, difference-making would retain a basis in physical causation. However, given the nature of Strevens’ method, something is missed if one characterizes such a DM relation as either an abstraction from (in the sense of omitting information) or just a summarization of physical (production-style) causal relations. This is because the method utilizes conceptual elements distinctively different from the notion of production. The key distinction involves the need to consider alternative scenarios. Strevens himself points out that all difference-making accounts of explanatory relevance compare the actual scenario with non-actual scenarios where the cause is not present (see Strevens, 2008, 55-56). His kairetic account features a sophisticated variant of this removal and comparison step, featuring what he calls “abstraction”: for instance, if the description of an actual scenario features a 10 kg cannonball breaking a window, one would substitute a proposition involving cannonballs with a range of masses that would have also caused a break (2008, 96-99). But the process of determining the appropriate range still implicitly requires a removal and comparison process characteristic of DM methods, and this is not something which is part of the production concept.

When it comes to the present discussion, then, it doesn’t appear that Ney’s argument for causal foundationalism quite settles the question asked above regarding the conceptual priority of production over DM. Even if DM facts are also physical causal facts
given the assumed universality of physical causation, the DM concept has features that go beyond those present in the production concept.

2.2 Psychological evidence for priority?

In an interesting recent exchange between Woodward and Jonathan Waskan, the debate over explanatory frameworks centers on their conceptual foundations. Waskan (2011) wants to defend mechanistic explanation by arguing that causation is plausibly at root a mechanistic concept. Woodward (2011a) believes that the concept of mechanism is problematic,\(^{35}\) and he defends the importance of his interventionist brand of difference-making. Both authors then turn to psychological research to shed light on the matter (Woodward also pursues the question in his 2007 and 2011b). The debate comes down to the question of whether one of the concepts is cognitively and perhaps developmentally foundational to our causal reasoning. If so, then this might support the claim that (normative) theoretical frameworks linked to that notion offer deeper or superior causal explanations.\(^{36}\)

Recalling the discussion above of causal perception and causal inference (and assuming there is a link between these psychological phenomena and the production/difference-making distinction), the fact that causal perception is present very early in development may suggest it plays a necessary role in the emergence of causal inference. Waskan suggests that causal perception, supplemented by other perceptual

\(^{35}\) This is since it is “used in a variety of different ways or at least has amorphous boundaries (Woodward, 2011a, 410).”

\(^{36}\) Woodward explores the relationships between philosophical theories of causation and psychological evidence in his (2012). He begins with the premise that human causal cognition is “fairly successful in enabling us to cope with the world”. So, if a normative theory proposes an approach to causal reasoning and people do in fact seem to reason in a roughly compatible way, then the theory might help explain “why (and to what extent) people are successful causal reasoners (Woodward, 2012, 962-3).”
expectations about the nature of objects, underlies the forming of expectations about when one event makes another happen in mechanistic settings. Causation, he proposes, is a “non-specific” mechanistic concept “abstracted from instances in which we integrate a variety of early perceptual expectations in order to understand how events in specific mechanical systems unfold (Waskan, 2011, 390).” Crucially, he next suggests that achieving this kind of mechanistic understanding incorporates “tacit knowledge” of what might happen in alternate scenarios, involving, say, interference with a given mechanism (402). If this is right, then this mechanistic concept of causation also allows for “an ability to determine the consequence of counterfactual interventions (390).” This would obviate the need to invoke the DM concept in mechanistic explanation. Going beyond the goals of Waskan’s discussion, one might also suggest that this concept of causation could be the general foundation for adult causal inferences outside the mechanistic domain.\(^{37}\)

Here, then, we have a sketch of the case for the developmental and cognitive priority of production over DM: causal perception is the key element in forming a productive/mechanistic concept of causation, which in turn is the key element in domain-general adult causal inference. However, there is room for doubt about this proposal. First, with regard to drawing conclusions about the order of development, Danks points out that the date of emergence of adult-style causal inference is not known (and is difficult to test in very young children) (Danks, 2009, 456). So, we cannot say with certainty that causal perception precedes it in time. Also, when considering the process of learning about mechanisms in children, it isn’t necessarily clear what role is played by

\(^{37}\) Some psychologists, dissenting from the approaches to causal inference discussed above, do advance a foundational role for notions related to production or mechanisms in causal reasoning. In particular, Ahn & Kalish (2000) posit a crucial role for mechanistic beliefs in adult causal inferences, while Wolff (2007) argues that subjects represent causes (even involving non-physical factors such as human intentions) in terms of force dynamics. On a related note, Strevens (2007) outlines a number of ways in which thinking in mechanistic terms is useful to our reasoning.
leveraging causal perception and basic object knowledge vs. the utilization of a distinct domain-general inference process the children might already possess (see discussion in Kushnir & Gopnik 2007). Further, even if causal perception is necessary for the development of domain-general causal inference in children, perhaps mediated via learning about mechanisms, it may not mean these remain foundational elements for inference in adults. We may have developmental priority without cognitive priority. There appear to be relatively few studies that directly address the relationship between causal perception and causal inference in adults, but Schlottmann & Shanks (1992) is relevant. Their experiments presented subjects with displays of successful and unsuccessful launches of one colored square by another that were coupled in a variety of ways with a non-launching predictor of movement (a change in color signaling the immanent movement of the second square). The authors reach a couple of conclusions. First, the results reinforce earlier studies in that a successful launch irresistibly prompts the ascription of causation: this effect is present regardless of how reliably the launching succeeds in a sequence. Second, the subjects at the same time were able to distinguish judgments of how necessary the launch versus the color change was to a successful effect. This result suggests that the process of causal inference (what the authors call causal judgment) has a high degree of independence from causal perception. While the authors speculate that intuitions derived from causal perception do help guide adult causal inferences, Danks sees as the main lesson of the experiments the apparent separability of the two processes of perception and inference (Danks, 2009, 457).

Woodward (2011a) also argues that the evidence is consistent with the relative independence of the two processes and thinks this tells against the idea that the capacity for causal inference is somehow derived primarily from causal perception. He then goes further to propose that the evidence is also consistent with difference-making being the
principle ingredient in the adult concept of causation: while causal perception and learning about mechanisms may play a developmental role, they aren’t crucial elements in the adult concept. Woodward feels in particular that understanding the significance of potential manipulations is important for attributing causation. And given this emphasis on manipulation, he raises a concern about the linking of the concept of causation with causal perception given the disassociation of perception and action present in very young infants and non-human animals (Woodward, 2011b, 254-256).

Woodward’s turnabout suggestion that the adult concept of causation is based primarily on difference-making seems highly improbable, however, given the continuing strong tie between causal perception and ascriptions of causation in adults. On the other hand, the idea that the difference-making notion is somehow inherent in or principally based upon causal perception, perhaps leveraged through the understanding of simple mechanisms, lacks adequate support. What seems best supported is the relative independence of the concept linked to causal perception and the one involved in causal inference (although they are surely also integrated in complex ways in many real-life causal claims). The philosophical case for two concepts of causation, as outlined by Hall, is bolstered by the empirical evidence, but the case for asserting the priority of one of these is not.

2.3 Implications for the debate over explanation

To summarize, production is a good fit for the physical realm, and it also commonly thought that all of reality depends on a physical foundation. Therefore, it is

38 Some research has suggested an important role for interventions in causal learning in children (see Kushnir & Gopnik, 2005. Schulz, Kushnir & Gopnik, 2007).
39 Also see the strong experimental support for the intuitions of causation in late-preemption scenarios discussed in section 3.1 below.
tempting to view production as the more fundamental causal notion compared to
difference-making. But there is no persuasive argument that supports the claim of an
explicit dependence relation between the two. While future work might address this lack,
the evidence from psychology also supports the view that the concepts are best seen as
independent from one another: the case for difference-making having a developmental or
cognitive basis in production has not been convincingly made. Now, in cases where both
causal notions can apparently be applied (i.e. no confounding factors are present), it is
certainly true that a difference-making relation leaves out information compared to the
production relation (at a minimum information about connections). But the difference-
making relation cannot be viewed as an abstraction from the production relation:
substituting a difference-making relation for a production relation is not just leaving out
detail: it is moving to a distinct concept. Revisiting the rivalry between alternative
explanatory approaches, we might ask the following: if we (somehow) had available
equally good theoretical approaches to explanation based on the two concepts, what
would be the right way to apply these (limiting our attention to the natural sciences)?

3. Determining applicability criteria

To address the hypothetical question, we can turn our attention back to the fact
that when the two notions are differentially salient, they have priority in distinct kinds of
cases. Clearly, production dominates our intuitions where identifiable processes connect
events. One source of evidence is our strong willingness to ascribe causation to the
preempting cause in preemption scenarios where such a tie is apparent. Another source
of evidence is the durable link between causal perception and causal ascription when
subjects view depictions of launching processes. To the extent mechanistic explanations effectively build on the concept of production, they will tend to be favored where physical connections are thought to be traceable, even in the case of more complex phenomena. And as a number of authors have stressed, mechanistic explanations are indeed pursued by researchers throughout the natural sciences.

What about difference-making? Clearly, pragmatic factors may favor employing difference-making even in cases where physical connections are known to be present: it may be impractical to trace the productive causal pathways. But there also may be cases where productive connections are either absent or else are deemed relatively unimportant to explanation. There may be many contexts where the latter is the case. For one example, consider that connections are less important in scenarios involving the actions of agents: in seeking explanations the most important element is often the intention to accomplish an end. Also, recent philosophical work has highlighted that in cases where we are inclined to appeal to absences or omissions as causal relata (making the production concept idle), a particular normative context is present (see Beebee 2004, McGrath 2005). Billy’s failure to water the plant caused it to die because he was expected to water it. If this wasn’t the case, then we could equally well label as causes a huge number of absent potential sources of watering. These examples suggest that the DM concept may play an apt role when teleological or normative considerations are germane to the context of an explanation.

3.1 Teleology and difference-making

In the context of biology, references to causation by omission and prevention do occur, but norms (of proper or natural function) will also be present here. Consider
Woodward’s discussion of the lac operon gene regulation model for E. coli, which is presented as an instance of double prevention appropriately handled by a DM approach:

When lactose is present in its environment, E. coli produces enzymes that metabolize it, but when lactose is absent, these enzymes are not produced. What determines whether these enzymes are produced? According to the model proposed by Jacob and Monod, there are three structural genes that code for the enzymes as well as an operator region that controls the access of RNA polymerase to the structural genes. In the absence of lactose, a regulatory gene is active which produces a repressor protein which binds to the operator for the structural genes, thus preventing transcription. In the presence of lactose, allolactose, an isomer formed from lactose, binds to the repressor, inactivating it and thereby preventing it from repressing the operator, so that transcription proceeds. (Woodward, 2003, 225-226)

Note that the description is full of productive language describing the process and its various steps. However, given our interest in explaining the larger biological phenomenon, it is also salient that the enzyme production depends on the allolactose. It makes a difference to the outcome, and it seems natural to identify it as a cause. The context plays a role in priming this intuition: the ability to metabolize lactose when available is an evolved biological function.

40 E. coli “produces” the needed enzymes, an active process that begins with the transcription of the lac operon genes; the regulatory gene “produces” the repressor protein, which (absent allolactose) “binds” to the operator for the structural genes; while allolactose (when present) “binds” to the repressor protein. 41 Woodward: “A causal relationship is clearly present between the presence of allolactose and the production of the enzymes, and the former figures in the explanation of the latter, but there is no transfer of energy from, or spatio-temporally continuous process linking, the two (Woodward, 2003, 226).”
There is some empirical support for the conclusion that this kind of context plays a crucial role. Turning back to the psychological literature, another kind of research asks subjects to render judgements on causation in linguistically described vignettes. Some studies have featured preemption and double-prevention scenarios. Lombrozo (2010) is particularly interesting: the article references the philosophical debate regarding the two concepts of causation (which Lombrozo terms “dependence” and “transference”) and presents the results of experiments that address the question of their roles in explanation when intentional/teleological contexts are absent and present.

Lombrozo summarizes prior research as showing that subjects have a fairly strong tendency to attribute causation to the preempting cause in preemption cases (favoring the production concept). On the other hand, there has been a mixed verdict regarding whether subjects attribute causation in double-prevention cases where production would not apply (306-308). Lombrozo sets out to examine these kinds of causal ascriptions under two different “explanatory modes” (309) – mechanical and teleological. Specifically, the experiments feature preemption and double-prevention scenarios some of which involve teleological elements while others do not. A first set of experiments uses human goal-directed behavior vs. human accidental actions to distinguish the teleological vs. merely mechanical modes. In cases of preemption, subjects gave the preempting factor a high “causal rating” compared to the preempted factor regardless of whether intentions were included or not. In double prevention scenarios, Lombrozo finds that the existence of intentions significantly raises the causal ratings subjects attribute to the double preventer. Without intentions, the double preventer is rated lowly, but with intentions its rating is comparable to the rating given to the proximate productive factor. A second set of experiments changes the teleological element by removing the presence/absence of human intentions in favor of a contrast between scenarios involving outcomes that fulfill
a function (including both biological and human-designed artefactual scenarios). For example, one double prevention scenario with a teleological character features interacting chemicals in the food sources of a shrimp: the result of these interactions enables the realization of a trait (skin reflectance of UV radiation) that the subjects are told evolved because it conferred an advantage. The results of this second set of studies roughly parallel those of the first set, although the relative impact on causal ratings of adding the teleological contexts in the double prevention scenarios was not quite as strong.

The evidence appears to be consistent both with the proposal that difference-making and production are independent concepts underlying causal ascriptions, and that the two concepts have different conditions of applicability.\textsuperscript{42} \textsuperscript{43} While experiments featuring preemption show the relative strength of the production concept where connections are present (regardless of context), where connections are missing there is difference: in non-teleological settings, the production notion remains dominant, but in teleological contexts, the difference-making notion becomes more salient.\textsuperscript{43}

3.2 Normative contexts and the application of the two concepts

Contexts have entered into in theories of causation and explanation in a number of ways. In particular, the description of events as normal vs. abnormal (or default vs. deviant) has sometimes played a role in attempts to analyze causation, and for that reason

---

\textsuperscript{42} A caveat is that all the studies discussed here were conducted under prevailing standards governing statistical significance, which have come under greater scrutiny recently. It is not known what the impact of even tighter standards would be for these research results.

\textsuperscript{43} An alternative explanation considered by Lombrozo and favored by Woodward (2012) is that the varying ratings on double prevention had to do with the perceived stability of the causal relation (with the intentional/functional causes being viewed as more consistent over counterfactual scenarios). This interpretation is perhaps given a boost by the accidental elements that were present in the mechanistic scenarios. However, the results are clearly compatible with the idea that the two notions of causation have different conditions of applicability.
these approaches have been subject to criticism in terms of their ability to provide a neutral or objective account. The approach of those who offer a difference-making approach to causal explanation has generally been to draw a distinction between a collection of difference-makers that can be identified objectively, and a set of contextual considerations that are subsequently applied to pick out the most relevant candidates. In the case of Woodward, he acknowledges that various contextual considerations enter into judging what interventionist counterfactuals are relevant: these may include “moral requirements, expectations and customs (Woodward, 2003, 88).” But he says this does not serve as the basis for criticizing the objectivity of his approach, since an objective structure of interventionist counterfactuals holds independently of the application of these additional criteria (90). Streven highlights the distinction between claims about difference-making simpliciter (a relation drawn from a fundamental and objective web of “causal influence”) and ordinary causal claims that are typically “assertions of framed difference-making (Strevens, 2013, 315, emphasis added).” Here the difference-makers are relative to a framework, which is a set of propositions that will usually include certain normative facts (317). Similarly, Hitchcock & Knobe (2009) argue that in cases where we identify a singular cause, it is selected from an “egalitarian” multifaceted (difference-making) causal structure: criteria for selection are based on normative considerations (regarding statistical norms, norms of good/bad, or norms of proper function).

The present proposal can be sketched as follows: in cases where applying a production-based account is appropriate, that is, when physical connections are present, then productive relations should provide the primary template for identifying causes.

---

44 See, for example, the discussion of Hitchcock’s (2007b) account by Paul & Hall (2103) pp. 205-209.
This would take the place of the objective (or egalitarian) step of identifying causal structure discussed by the difference-making theorists above. Contextual considerations will then enter into selecting a subset of productive causes deemed relevant; this will be driven, for instance, by the priorities set by the scientist/research program. However, applying the concept of difference-making may also enter the picture at this stage when the context includes normative elements. When it comes to biological settings in particular, there is reason to think difference-making should naturally play a role in contexts involving norms of proper function.

Returning to the morphogenesis example at the beginning of section 1, the authors note that the program of identifying gene function has been prominent in past research. In this context, difference-making relations (found through forward and reverse genetics) were salient, while the productive mechanisms involved in development, in contrast, were deemphasized (of course pragmatic obstacles to detailing the mechanism may have played a role in the relative emphasis biologists have placed on the two approaches). The authors, in turn, highlight the explanatory value that can come from a mechanistic analysis that fills in the productive connections. Given the conclusions drawn in this section, we can see that good explanations in the context of biological function may feature both productive and difference-making elements.

4. The missing theory of production

The foregoing conclusion was conditional on the assumption that equally good theories based on the two concepts were available. But when surveying the rivalry between mechanistic and DM explanatory frameworks, there is a significant asymmetry
in the critiques offered. The DM frameworks are accused of a limitation in their ability to offer pertinent details. The mechanistic frameworks, on the other hand, are accused of deficiencies in their theoretical foundation. While the concept of production is an inspiration for their approach, advocates of mechanistic explanation take a pluralistic approach to filling in their frameworks, and don’t provide a grounding in the more basic features of production. Extant production theories are deemed inadequate or unsuitable, and hence the approaches are left without an essential underpinning. \(^{45}\) Whether or not one endorses them, the rival DM approaches have well-developed theories that describe what difference-making is as well as how it can figure into explanation. In order for the two concepts to be deployed as the above discussion suggests, this deficiency on the part of the production-inspired frameworks needs to be addressed.

The following are desiderata for a theory of production.\(^{46}\)

A. The theory should provide a unified account of the various productive connections found in the phenomena.

B. To do this, the theory should also provide an account of the persistence or stability of entities that enter into productive connections.

C. In addition to connections, it should also include an element corresponding to the notion of “bringing about” or generation that is part of the concept of production.

---

\(^{45}\) Illari also identifies a need for an improved theory of production, and in her (2011) explores the idea of an account based on the transmission of information.

\(^{46}\) Absent from this list is a requirement that the theory be reductive. Difference-making theorists have shown that illuminating causal explanatory accounts need not include a reduction of causation to non-causal elements.
D. The theory should respect the intrinsic character of production by restricting its domain to actual phenomena. Counterfactuals should not play a foundational role.

E. Finally, the theory must be coupled to an account of its role in explanation. This should show how it properly underpins etiological and constitutive explanations in physics, chemistry and biology.
Chapter 3: A Review of Causal Process Theories

The most prominent theories of production in recent decades have been causal process theories, versions of which are due to Wesley Salmon and Phil Dowe. Salmon’s original (1984) mark-transmission account was meant to provide both a new theory of causation and a framework for scientific explanation: under Salmon’s ontic conception, to explain a phenomenon is to fit it into the world’s causal structure. To describe this structure, Salmon introduced as a basic entity the spatio-temporally extended causal process. A causal process is something that can transmit a change or “mark”: the interaction of two or more processes is responsible for these productive changes. This account faced two main lines of criticism. First, as a theory of causation, Salmon’s account fails to meet certain constraints associated with empiricism. The main idea here was that causal processes should be distinguishable from non-causal phenomena using only criteria derived from observation. It follows that the definition of a causal process should avoid use of counterfactuals or the invoking of “hidden powers.” Salmon’s formulation did use counterfactuals, and was also subject to counterexamples involving so-called pseudo-processes that appeared to meet his definition of a causal process. The second line of criticism involved the theory’s adequacy to underpin explanations. Here the formulations were seen as vague and ill-suited for explaining complex phenomena. Salmon eventually abandoned his theory in favor of a different approach, similar to that

47 A somewhat related approach is the transference theory of causation, an example of which is due to David Fair (1979). This approach has generally been seen as less promising than process theory, although see Kistler (1998, 2006) for a recent defense. A metaphysical theory with some similar elements to the process account is Ehring’s (1997) trope-persistence theory. Ingthorsson (2002) discusses the notion of production as interaction in comparison with traditional ideas about causation.
advanced by Dowe, which characterizes processes in terms of conserved quantities. Here
the formulations are clearer and avoid appeal to counterfactuals. However, the approach
has other shortcomings. It bases causation on contingent and possibly false assumptions
about physical quantities, and also has a narrow scope of application for explanation.
Below I argue that Salmon’s original theory presents an attractive approach to thinking
about productive causation. However, the approach is not compatible with the kind of
constraints on causal theorizing that Salmon and contemporaries thought should apply.
This assessment suggests a different path forward for the theory’s development.

After discussing a related predecessor theory due to Bertrand Russell (section 1),
Salmon’s mark-transmission approach and its motivations are described in section 2.
Section 3 reviews the main criticisms faced by Salmon’s theory. Section 4 outlines the
conserved quantity versions of the theory and highlights some shortcomings of this
approach. Finally, Section 5 argues that Salmon’s original ideas about causation and
explanation are promising, and that a different formulation of process theory and its
explanatory role can be successful.

1. Russell’s Causal Lines
In *Human Knowledge: its Scope and Limits* (Russell 1948), Russell proposes that certain causal notions play a key role in scientific inferences. First, he sketches a very general notion of a causal law:

A “causal law,” as I shall use the term, may be defined as a general principle in virtue of which, given sufficient data about certain regions of space-time, it is possible to infer something about certain other regions of space-time. The inference may be only probable… (Russell, 1948, 326)

Russell then introduces a distinction that anticipates Salmon’s later introduction of two basic conceptions of propagation and production:

Causal laws are of two sorts, those concerned with persistence and those concerned with change. The former kind are often not regarded as causal, but this is a mistake. (Russell, 1948, 327)

Russell gives as examples of the first sort Newton’s first law, and the general persistence of matter and energy. Russell fleshes out his notion of persistence (or “quasi-permanence”) with the introduction of “causal lines:”

A “causal line,” as I wish to define the term, is a temporal series of events so related that, given some of them, something can be inferred about the others whatever may be happening elsewhere. A causal line may always be regarded as

---

48 This constructive view of causation’s role in science appears to place *Human Knowledge* in opposition to the arguments of Russell’s own “On the Notion of Cause” from 1913. There, Russell argues that the prevailing philosophical notion of cause does not play a role in advanced sciences. The disagreement between the two positions might be seen as limited: the later Russell is developing causal notions that better suit the practice of science as compared to the traditional concept he criticizes in the earlier article. However, the book’s conclusions clearly imply that the earlier paper overreached in its appeal to the use of functional relations to dismiss the importance of causation to science (see Russell, 1948, p. 475).

49 Discussed below, and see Salmon (1984), pp. 138-139.
the persistence of something—a person, a table, a photon, or what not. Throughout a given causal line, there may be constancy of quality, constancy of structure, or gradual change in either, but not sudden change of any considerable magnitude. (Russell, 1948, 477)

Note that in contrast to Salmon’s theory discussed below, Russell is working with an event ontology, so the persisting causal line is comprised of events connected according to a causal “law” of the first kind above.  

While the term “lines” might suggest a fairly simple sort of continuity, Russell discusses how persistence may take the form of an abstract notion of “structure” shared by events or event-complexes. He talks about the kinds of inferences that can be made from observing certain configurations of structurally similar events (regarding common causal origins, spatio-temporal patterns, and the presence of laws fostering the stability of these – pp. 498-499). Russell suggests that these inferences account for most of those that are utilized in the development of physics (499). Having discussed persistence, Russell turns to the question of change, but he has less to say here. Changes are due to interactions, and he believes laws governing interactions are formulated using quantitative enumeration and induction coupled with the identification of the causal lines involved. It follows from the earlier discussion that if laws of interaction are deemed to be causal laws, then the associated conception of cause would be distinct from that which connects events in causal lines. Russell does stipulate that causal laws include those which only confer probability to the inferred effects (327).

---

50 He also refers to a causal line as a “more or less self-determined causal process (Russell, 1948, 477).”
51 This portion of the discussion has much in common with that of Russell’s earlier Analysis of Matter (1927).
In the book’s penultimate chapter, Russell formulates five postulates that are “required to validate scientific method (506).” Four of the five are premised on the existence of causal lines (the fifth is a “postulate of analogy”). For Russell, the framework of persisting and interacting causal lines is the key to gaining scientific understanding of the world.

2. Salmon’s Mark-Transmission Theory

Salmon did extensive work on the problem of scientific inference, and engaged with Russell’s ideas in works such as his *Foundations of Scientific Inference* (1966). In this book, Russell’s views are discussed as part of a general survey of efforts to overcome the traditional problem of induction. As such, Salmon does not deal there with Russell’s causal concepts in depth. Later, Salmon did engage with these causal concepts by approaching them from a different direction: that of developing an account of scientific explanation. Salmon had worked on scientific explanation since the 1960’s, culminating in his statistical relevance account (Salmon 1971). Beginning with papers published in the mid-1970’s, however, he began to introduce a new approach to explanation, which had its

---

52 Russell’s last chapter is a critique of empiricism. The relationship of empiricism to the development of causal process theories will be discussed below.
53 In Salmon (1966) he characterizes Russell’s five postulates as an attempt to justify “ampliative” inferences. In Salmon (1974) he notes that Russell’s formulation of his postulates constitutes a rejection of empiricism, and he characterizes Russell’s reasoning as akin to a Kantian transcendental deduction (a difference being that Russell did not insist on the truth of this postulates).

In this book, Salmon describes his approach as embodying the “ontic” approach to explanation: to explain an event is to fit it into a “discernible natural pattern” that obtains in the world (Salmon, 1984, 121).\(^{55}\) This pattern of relations is specifically construed in causal terms: “To provide an explanation of a particular event is to identify the cause and, in many cases at least, to exhibit the causal relations between this cause and the event-to-be-explained (121-2).”\(^{56}\) Salmon sees it as crucial for his approach, then, to carry out an analysis of the causal relations that prevail in the world: failure to do so would be a “major shortcoming” (122). Hume provided a critique of the causal relation, which according to Salmon, must be faced “squarely” (135). The force of this critique, as sketched by Salmon, would rule out ‘necessary connections’ between causes and effects or ‘hidden powers by which a cause ‘brings about’ an effect (137). A fundamentally different conception of causation is therefore necessary. Salmon’s alternative framework makes use of two fundamental causal concepts, which he labels *propagation* and *production* (139). It also introduces a new basic entity, the *causal process*.

---

\(^{54}\) Salmon’s development of the causal process account throughout the period was intertwined with the continuing development of the statistical relevance approach. This latter work was not essential to the causal process theory, however, and will not be a focus of the present discussion.

\(^{55}\) Salmon assesses the landscape of accounts of explanation offered by philosophers of science, and sorts them according to three conceptions of explanation: the epistemic, the ontic, and the modal. In Salmon (1989), he discusses his adoption of this framework, noting debts to articles by J.A. Coffa and D.H. Mellor (see Salmon, 1989, p. 118). The launching point for most of the alternatives canvassed was in the form of critical responses to the classic Deductive-Nomological (D-N) account made prominent by Hempel (Hempel & Oppenheim, 1948). The D-N account (informally, the “covering-law” approach) is categorized by Salmon as “epistemic” because the combination of laws and initial conditions are used to demonstrate that an *event-to-be-explained was to be expected* (Salmon, 1984, 16, emphasis original).”

\(^{56}\) Salmon at times also describes the ontic approach to explanation as a search for “mechanisms” (see Salmon, 1984, pp. 123-4). Late in the book he also characterizes his work as an updated version of “the mechanical philosophy” (pp. 239-241).
2.1 Causal Processes

Compared to an event, a causal process is typically extended in time as well as space: “A baseball colliding with a window would count as an event; the baseball, traveling from the bat to the window, would constitute a process (139).” Salmon explains that this notion is similar to Russell’s causal line (see above). Both conceptions encompass material objects at rest, in addition to bodies or waves that evolve in time.

Going beyond Russell, Salmon recognizes a need for distinguishing causal processes from other kinds of processes that may display regularities (labelled pseudo-processes). After a discussion of the compatibility of a process ontology with special relativity (140-141), Salmon uses relativity’s limitation on the velocity of light to introduce the distinction:

It is a fundamental principle of that theory that light is a first signal—that is, no signal can be transmitted at a velocity greater than the velocity of light in a vacuum. There are, however, certain processes that can transpire at arbitrarily high velocities—at velocities vastly exceeding the speed of light. This fact does not violate that basic relativistic principle, however, for these ‘processes’ are incapable of serving as signals or of transmitting information. Causal processes are those that are capable of transmitting signals; pseudo-processes are incapable of doing so. (Salmon, 1984, p.141)

Salmon’s standard example is that of a very large circular building with a rotating spotlight at its center. A beam of light travelling to the wall is a causal process. If the spotlight is rotated, the spot of light cast upon the wall will move around the perimeter of the building in a highly regular fashion. In fact, if the building is large enough and the rotation is fast enough, the spot may trace a path along the circular wall at a speed
exceeding that of light. It may seem to meet the definition of a process, but it will fail to be a causal process, since it is incapable of transmitting a signal, or “mark”\textsuperscript{57}.

A causal process is capable of transmitting a mark; a pseudo-process is not. Consider, first, a pulse of light that travels from the spotlight to the wall. If we place a piece of red glass in its path at any point between the spotlight and the wall, the light pulse, which was white, becomes and remains red until it reaches the wall. A single intervention at one point in the process transforms it in a way that persists from that point.\textsuperscript{58} (Salmon, 1984, 142)

The light pulse, capable of transmitting a mark, is a causal process. Contrast this with the rotating spot of light on the wall:

There are a number of ways in which we can intervene to change the spot at some point; for example, we can place a red filter at the wall with the result that the spot of light becomes red at that point. But if we make such a modification in the traveling spot, it will not be transmitted beyond the point of interaction...the mark can be made, but it will not be transmitted.\textsuperscript{59} (Salmon, 1984, 142)

Another example of a pseudo-process is a shadow:

Consider a car traveling along a road on a sunny day. As the car moves at 100 km/hr, the shadow moves along the shoulder at the same speed. The moving car,

\textsuperscript{57} Salmon says that the inspiration for the mark criterion is owed to Hans Reichenbach, who used the notion of a mark as an indicator of causal relevance in The Direction of Time (Reichenbach, 1956).

\textsuperscript{58} He goes on to say the following: “We shall say, therefore, that the light pulse constitutes a causal process whether it is modified or not, since in either case it is capable of transmitting a mark (Salmon, 1984, 142).” This notion of being “capable”, however, is potentially problematic, as discussed below.

\textsuperscript{59} Note that while a red filter placed at the source will turn the moving spot red at all points, this does not constitute a local intervention in the “spot-process” itself (Salmon, 1984, p. 142).
like any material object, constitutes a causal process; the shadow is a pseudo-process. If the car collides with a stone wall, it will carry the marks of that collision—the dents and scratches—along with it long after the collision has taken place. If, however, only the shadow of the car collides with the stone wall, it will be deformed momentarily, but it will resume its normal shape just as soon as it has passed beyond the wall. (Salmon, 1984, 143)

Salmon goes on to describe the distinction using a loose definition of the “structure” of a process. Given the sense of uniform behavior that characterizes a process, they can be said to possess a certain structure: “The difference between a causal process and a pseudo-process, I am suggesting, is that the causal process transmits its own structure, while the pseudo-process does not (144).” The regularity in behavior that characterizes a pseudo-process, like the rotating spot of light, will be “parasitic” on other, truly causal processes.

Causal processes, so defined, transmit signals, information or, more generally, causal influence: “Such processes are the means by which causal influence is propagated in our world (146).” The structure of a casual process will be responsible for certain effects when it interacts with other processes: “A causal influence transmitted by a flying arrow can pierce an apple on the head of William Tell’s son. A causal influence transmitted by sound waves can make your dog come running (146).”

Salmon discusses the improvements in his characterization of causal processes vs. Russell’s causal lines. Russell does not offer an account of how causal lines differed from other regularities that seem to fit his criteria, which Salmon provides. Another issue is that Russell’s characterization utilizes epistemic language, given his focus on causal lines as a foundation for inferences. Salmon intends his account to be describing objective
causal processes, since “the existence of the vast majority of causal processes in the history of the universe is quite independent of human knowers (Salmon, 1984, 147).”

2.2 Propagation

A further difference between Russell and Salmon is the latter’s attempt to flesh out the concepts of persistence and change (propagation and production), which Russell somewhat vaguely attributed to the existence of two kinds of causal laws. Turning to the first of these concepts, Russell’s employment of an event ontology implied the need to connect a temporal series of events comprising a causal line. Salmon, envisioning a Humean objection to postulating a connecting relation between events, develops a different notion. In his account, causal processes are extended in time as part of their basic nature (this, in turn will allow the processes themselves to play the fundamental role of causal connections in nature): he called this feature of the account the “At-At Theory of Causal Propagation” (Salmon, 1984, 147).

Salmon says that in response to the question of how a causal process can transmit a mark, the “astonishingly simple answer” is as follows: “the transmission of a mark from point A in a causal process to point B in the same process is the fact it appears at each point between A and B without further interactions (148, emphasis original).” The structure of a causal process will propagate through space-time in the absence of interactions.60 This “at-at” theory of propagation is key to Salmon’s attempt to defuse the Humean critique. Since temporal extension is basic to a causal process, there is no need for powers or necessary connections to account for propagation. Still, Salmon is sensitive to the concern that a “mysterious power” is implied by the fact that in his formulation,

60 Salmon explains that he was inspired by contemplation of one of Zeno’s paradoxes—that of the flying arrow (see Salmon, 1984, pp. 151-3).
the “ability to transmit a mark is the criterion of causal processes (153, emphasis original).” He defends this formulation by asserting that a series of experiments can establish this feature (147-148).61

Salmon presents what he calls the basic thesis about mark transmission as follows (MT):

**MT:** Let P be a process that, in the absence of interactions with other processes, would remain uniform with respect to a characteristic Q, which it would manifest consistently over an interval that includes both of the space-time points A and B (A ≠ B). Then, a mark (consisting of a modification of Q into Q'), which has been introduced into process P by means of a single local interaction at point A, is transmitted to point B if [and only if]62 P manifests the modification Q' at B and at all stages of the process between A and B without additional interventions. (Salmon, 1984, 148, emphasis original)

Salmon notes the inclusion of a counterfactual condition in the first sentence of the MT thesis.63 This is intended to rule out a scenario such as one where a red filter is placed on a spot on the wall a fraction of a second after one is placed at the rotating source: it would appear that the pseudo-process met the subsequent criterion for mark transmission (148). Salmon realizes the need to justify the use of a counterfactual, and he again appeals to experimentation (148-150). This feature of his formulation would be a focus of some critics and would trouble Salmon himself in subsequent years.

---

61 Whether Salmon’s formulations can overcome this sort of Humean concern will be discussed in section 3 below.
62 The ‘only if’ was omitted by Salmon, but is implied by his discussion.
63 Salmon says that the need for this was pointed out to him (in conversation) with Nancy Cartwright after a discussion of previous formulations (Salmon, 1984, p. 148).
Salmon summarizes the ideas of transmission and propagation with the principles of structure transmission (ST) and propagation of causal influence (PCI):

ST: *If a process is capable of transmitting changes in structure due to marking interactions, then that process can be said to transmit its own structure.* (Salmon, 1984, 154, emphasis original)

PCI: *A process that transmits its own structure is capable of propagating a causal influence from one space-time locale to another.* (Salmon, 1984, 155, emphasis original)

Salmon asserts that his account of propagation provides an explanation of causal connections in nature: “causal processes constitute precisely the physical connections between causes and effects which Hume sought—what he called the ‘the cement of the universe’ (156).” He believes these connections provide the basis for superior causal explanations:

We might say, for instance, that turning the key causes the car to start…but I think that we can make sense of a cause-effect relation only if we can provide the causal connection between the cause and the effect. This involves tracing out the causal processes that lead from the turning of the key and the closing of an electrical circuit to various occurrences that eventuate in the turning over of the engine and the ignition of fuel in the cylinders. (Salmon, 1984, 155-56)

At this point, Salmon notes, however, that the notion of “interaction” has been used repeatedly in the discussion of causal processes and mark transmission, and this is in need of analysis.
2.3 Production

Two processes may intersect in space-time without causing any change in either. In the case of pseudo-processes, such as two shadows crossing, they will “move on as if no such intersection had ever occurred (169).” In the cause of interacting causal processes, such as colliding pool balls, on the other hand, mutual modification may occur:

In the case of the two pool balls, however, the intersection of their paths results in a change in motion of each that would not have occurred if they had not collided. Energy and momentum are transferred from one to the other; their respective states of motion are altered. Such modifications occur, I shall maintain, only when (at least) two causal processes intersect. If either or both of the intersecting processes are pseudo-processes, no such mutual modification occurs. However, it is entirely possible for two causal processes to intersect without any subsequent modification in either…light rays normally pass right through one another without any lasting effect upon either one of them. The fact that two intersecting processes are both causal is a necessary but not sufficient condition of the production of lasting changes in them. (Salmon, 1984, 169)

When causal processes intersect and mutual modification does occur, this is a “causal interaction,” which “produces” a change in the respective processes. Salmon sets out the principle CI (for causal interaction) as follows:

CI: Let P1 and P2 be two processes that intersect with one another at the space-time point S, which belongs to the histories of both. Let Q be a characteristic that process P1 would exhibit throughout an interval (which includes subintervals on both sides of S in the history of P1) if the intersection with P2 did not occur; let R be a
characteristic that process P2 would exhibit throughout an interval (which includes subintervals on both sides of S in the history of P2) if the intersection with P1 did not occur. Then, the intersection of P1 and P2 at S constitutes a causal interaction if:

(1) P1 exhibits the characteristic Q before S, but it exhibits a modified characteristic Q’ throughout an interval immediately following S; and

(2) P2 exhibits the characteristic R before S, but it exhibits a modified characteristic R’ throughout an interval immediately following S. (Salmon, 1984, 171, emphasis original)

Salmon’s formal definition of a causal interaction again includes counterfactuals: the alteration in the characteristics of the processes involved would not have occurred absent the intersection.⁶⁴

Causal events, then, are derivative notions on this account (there may be other sorts of events that are non-causal, such as intersections involving pseudo-processes). Salmon explains that when we speak of cause and effect events, we typically will be referring to interactions of causal processes. The cause event, then, will be linked to the effect event by means of a causal process:

⁶⁴ There are a couple of technical issues worth noting: first, the use of “if” instead of “if and only if” was deliberate this time—Salmon explains that other characteristics could have been altered in the requisite way even if Q and R were not (Salmon, 1984, p. 174); second, CI and MT may seem to offer circular definitions, with interactions introducing marks in the MT thesis and the introduction of transmitted marks defining interactions in CI. Later (in Salmon 1998), Salmon reformulates the principles of the theory in a way that avoids circularity (Salmon was responding to the critique in Dowe, 1992). This involved taking care to distinguish causal interactions from intersections, and also clarifying that a mark may be a transient feature of intersecting processes: the key for a causal interaction is that the marks persist beyond the intersection point. One point of interest in this revised formulation is that it clarifies that the causal interaction is logically prior to the causal process in the definitional hierarchy, reversing the order of presentation in the 1984 book (see Salmon, 1998, pp. 178-9).
We say, for example, that the window was broken by boys playing baseball. In this situation, there is a collision of a bat with a ball...the motion of the ball through space (a causal process), and a collision of the ball with the window. (Salmon, 1984, 178)

In this case the ball provides the causal connection between the events of interest. This basic type of interaction, involving two processes, is labelled an interactive fork (or “x” fork). Note that this interaction always involves mutual modification (even if the alteration in one of the participants is not the focus of our interest—for instance the change that occurs in the bat in the baseball scenario). Salmon also notes the implications of the account for the connection of causal concepts with the passage of time. In a typical scenario such as the baseball example, the “cause” must precede the “effect” in time, since the connecting causal process is temporally extended. On the other hand, the mutual modification that characterizes a causal interaction occurs simultaneously for the two processes: “Basically, propagation involves lapse of time, while interaction exhibits the relation of simultaneity (182).”

Salmon goes on to discuss other cases that are either more complex combinations of interactions, or have different structures involving merging or emission of processes (the window in the baseball example would be succeeded by a number of processes corresponding to its fragments). With this discussion, Salmon concludes his main

---

65 Salmon notes that the formulations could be reworded to leave out a commitment to temporal asymmetry. This would leave open the possibility of formulating a causal theory of time. An account of causal asymmetry could be developed according to which it is a contingent large-scale feature of the causal network.

66 See Salmon (1984), pp. 180-182. Those familiar with the 1984 book will note that in the present discussion I am passing over Salmon’s extensive discussion of the statistically characterized “conjunctive fork,” which is linked to Reichenbach’s principle of the common cause. Salmon also sketches a statistical characterization of the interactive fork. As Salmon himself later makes clear, these discussions are not
integral to the causal process account, stating that causal processes and causal interactions “cannot be characterized in terms of relationships among probability values alone (Salmon, 1990, 95).”

explication of the “causal structure of the world:” a network of interacting causal processes.

2.4 Probabilistic Causation

Salmon offers some comments on the relationship between the process approach and probabilistic causation. After reviewing the challenges faced by extant probabilistic theories (such as that of Patrick Suppes), Salmon explores the notion of transmission of probabilistic causal influence. He considers several examples, involving both everyday scenarios and scientific ones:

If a wave—produced by a rock dropped into the water nearby—approaches the [toy] boat, there is a certain probability that the boat will capsize when the wave reaches it…The boat is a causal process, and the wave propagating over the surface of the pond is another. When they intersect, there is a certain probability for a certain type of interaction, and a certain probability that it will not occur […]

If radiation of suitable frequency impinges upon a hydrogen atom, there is a certain probability that the radiation will be absorbed and the atom will exist for a time thereafter in an excited state. (Salmon, 1984, 202-203)

He pictures in these cases a causal process “that carries with it probability distributions for various types of interactions (203).” He says “it seems to be altogether appropriate to refer to these probabilities as propensities (203, emphasis original):”
A given alpha particle, impinging upon a gold foil, has propensities of given magnitudes for no interaction, for small deflection, and for large deflection. (Salmon, 1984, 203)

Salmon goes on to say that his comments must be distinguished from an endorsement of the propensity interpretation of the probability calculus, which is a different claim. However:

I am inclined to think that this idea of propensity, as a probabilistic disposition, is valuable. It is just such dispositions that seem to me to lie at the foundation of probabilistic causality. (Salmon, 1983, 204)

Despite Salmon’s earlier worries about “hidden powers,” he does not try to link this discussion to his earlier formulation of processes in terms of “structure” or “characteristics:” it is a straightforward attribution of power-like properties to causal processes (see discussion in section 3.1 below).

2.5 Causation and Constitutive Explanations

A distinction Salmon discusses is that between etiological explanations and constitutive explanations. Etiological explanations are ontic explanations that place the explanandum in a causal network, tracing processes from relevant preceding causal interactions. A constitutive explanation, on the other hand, is one that makes reference to the causal interactions and processes that compose a larger entity:

---

67 In another article, Salmon explicitly ties propensities to quantum mechanical phenomena: “My suggestion is that the quantum mechanical wave is a wave of propensity (Salmon, 1988, 15).”
68 Salmon explains he is “borrowing” the term from Larry Wright (Salmon, 1984, p. 269).
A constitutive explanation is thoroughly causal, but it does not explain particular facts or general regularities in terms of causal antecedents. The explanation shows, instead, that the fact-to-be-explained is constituted by underlying causal mechanisms. (Salmon, 1984, 270)

Salmon does not offer a detailed discussion of how causal processes (and their interactions) can constitute other causal processes, although he clearly thinks his theory provides the tools to develop an account of composition.69

3. Criticism of the Mark-Transmission Theory

Salmon’s theory was welcomed in the literature as an important contribution, but was subject to significant criticism, some of which was ultimately taken by Salmon himself to be decisive. A first line of criticism focuses on whether the theory meets empiricist constraints. As discussed, Salmon was sensitive to the need to avoid theoretical elements that could not be derived from observation or experimentation. A related objection questions whether causal processes and their interactions as defined by Salmon can be distinguished from pseudo-processes. A second line of criticism alleges that Salmon’s formulations are too vague and provided an inadequate basis for constructing explanations, particularly for complex phenomena.70

69 In this he also follows Russell, who believed that modern scientific understanding was incompatible with the traditional conceptions of physical objects and material substance. See Russell (1948), p. 333.
70 A further worry (shared by Salmon) concerns whether the account was fully consistent with science, particularly in the domain of quantum mechanics. This topic will be addressed in Chapter 5.
3.1 Capabilities and Counterfactuals

At times, Salmon’s causal account appears to violate his own restrictions regarding the inclusion of ‘hidden powers’ or ‘necessary connections,’ either explicitly, or implicitly via his employment of counterfactuals. As noted above, Salmon states his criterion for causal processes by saying they are those processes that have the ability to transmit a mark: “A causal process is capable of transmitting a mark; a pseudo-process is not (Salmon, 1984, 142).” In the actual world, a given causal process may or may not in fact transmit a mark. So how can Salmon account for this presence of this capability? He appeals to experimental evidence:

Ability to transmit a mark can be viewed as a particularly important species of constant conjunction—the sort of thing Hume recognized as observable and admissible. It is a matter of performing certain kinds of experiments. If we place a red filter in a light beam near its source, we can observe that the mark—redness—appears at all places to which the beam is subsequently propagated. This fact can be verified by experiments as often as we wish to perform them. (Salmon, 1984, 147)

But can repeated experimentation secure the “ontic” conclusion that the capability exists? This is perhaps an odd claim from a philosopher who wrote a book on scientific inference centered on Hume’s other famous skeptical challenge – the problem of induction (Salmon, 1966).

71 In a book review, Ronald N. Giere tersely concludes “the analysis employs counterfactual notions such as the capability of transmitting information. Thus Salmon’s analysis runs counter to the empiricist tradition going back to Hume (Giere, 1988, 446).”
Also, as noted above, when Salmon turns to probabilistic cases, he makes an explicit attribution of dispositions to causal processes. James Fetzer highlighted this in his review of the 1984 book:

While [Salmon] overtly identifies indeterministic causal mechanisms and dispositions of probabilistic strength, he does not explicitly relate deterministic causal mechanisms to dispositions of universal strength. Since dispositions of universal strength, no less than those of probabilistic strength, are tendencies to produce (or, to bring about) specific outcomes under appropriate conditions—concerning what effects would (invariably or probably) be brought about if various conditions were to obtain—such a conception is at least consistent with (and might actually be required by) Salmon’s causal/mechanistic theory of explanation. (Fetzer, 1987, 606, emphasis original)

A fairly natural interpretation might be that causal processes possess the capability to transmit a mark and they also have dispositions toward introducing persisting marks in certain other processes upon interaction. Are processes causal by virtue of possessing causal powers?

The careful formulations of the MT and CI principle do not refer to capabilities or dispositions, but then the concern moves to Salmon’s use of counterfactuals. In MT, mark transmission depends on a counterfactual assertion that the mark would not have appeared in the absence of the associated interaction. Similarly, in CI, the marking of the two processes depends on the claim that they would not have been so marked in the

---

72 Fetzer put it this way: “the critical question that arises at this juncture, of course, is whether or not these commitments to subjunctives and counterfactuals violate the Humean heritage that Salmon has so faithfully articulated heretofore (Fetzer, 1987, 607).”
absence of the associated intersection. As Phil Dowe points out, these seem to imply the presence of necessary connections (Dowe, 2000, 84). Specifically, the counterfactuals suggest that the particular modifications that occur have a necessary connection to the particular intersection of processes involved. Salmon’s response, again, is to appeal to experimental testing:

Science has a direct way of dealing with the kinds of counterfactual assertions we require, namely, the experimental approach. In a well-designed experiment, the experimenter determines which conditions are to be fixed for the purposes of the experiment and which are allowed to vary. The result of the experiment establishes some counterfactual statements as true and others as false under well-specified conditions...We set up the following experiment. The [spotlight] will be turned on and off one hundred times...[a] second experimenter uses his device to make a random selection of fifty trials in which he will make a mark...if all and only the fifty instances in which the marking interaction occurs are those in which the spot on the wall is red, as well as all the intervening stages of the process, then we may conclude with reasonable certainty that the fifty cases in which the beam was red subsequent to the marking interaction are cases in which the beam would not have been red if the marking interaction had not occurred. (Salmon, 1984, 149-150)\(^{73}\)

---

\(^{73}\) Another discussion in the same vein is offered for the counterfactual present in CI (Salmon, 1984, pp. 171-174).
The question becomes whether this justification of counterfactuals is adequate, or whether implicit non-Humean attributes of causal processes are needed to support the counterfactuals.\textsuperscript{74}

Philip Kitcher provides a detailed review of Salmon’s account and how it deals with these issues (Kitcher, 1989). Kitcher sees Salmon’s proposal for experimental justification of counterfactuals as inevitably falling short: “Even if the control group and the test group are “similar,” the fact that they differ with respect to the presence or absence of [one characteristic] will mean that they differ with respect to many other characteristics (Kitcher, 1989, 473).” He goes on to discuss other examples that involve less trivial choices in how to set up the control group—this highlights that in practice we design control experiments by drawing on background causal knowledge: “But if we are looking for a theory of how we justify counterfactuals from scratch, then the appeal to the method of controlled experiments is of no avail (Kitcher, 1989, 475).”\textsuperscript{75}

3.2 Distinguishing Causal Processes from Pseudo- Processes

In his critique, Kitcher also offers a good summary of the general challenge which a traditional empiricist approach to scientific knowledge presents for any causal theory and how it plays out in Salmon’s case. First, it is taken as a given that we do not directly observe causation, so “we come to make justified causal judgements by observing that certain conditions obtain, and…inferring causal claims from the premises that record our

\textsuperscript{74} Salmon shows no interest in utilizing a possible-worlds semantics, considering counterfactuals justified in such a way as “unsuited for the explication of causality or scientific explanations (Salmon, 1997, 476).”

\textsuperscript{75} Kitcher (1989) argues, in fact, Salmon’s whole approach might be re-framed to show that causal processes and interactions were only secondary notions, with counterfactuals playing the indispensable role in underpinning causation.
observations (Kitcher, 1989, 460).” This leads to the demand for an analysis of causation in terms of non-causal concepts licensed by observation:

Hence we arrive at the project of giving necessary and sufficient conditions for the obtaining of causal relations, formulating those conditions in ways that will dissolve the epistemological mysteries surrounding causation by deploying only concepts whose satisfaction is observationally ascertainable. (Kitcher, 1989, 460)

Salmon argues that his process approach offers a way to overcome the challenge. The traditional central causal concept was the notion of a causal relation between distinct events, and he replaces this with his notion of a spatio-temporally extended causal process. But, empiricist-style critiques can be carried out on his new concepts of causal processes and causal interactions. Given the mark-transmission account, how can observations distinguish causal processes from pseudo-processes? How can they distinguish causal interactions from mere intersections? Several objections to Salmon’s account focused on these questions, and proceeded by providing examples where the mark criterion appeared to fail in distinguishing causal processes from pseudo-processes.

Kitcher offers several examples involving shadows. One of these pictures a moving vehicle on an ice-rink casting a shadow on the ice. Then, “a projectile is thrown in such a way that it lands at the edge of the shadow with a horizontal velocity equal to that of the shadow of the vehicle (Kitcher, 1989, 464).” In this case there is an intersection of the projectile and the shadow that leaves a persisting mark on both: the shadow’s shape is distorted and the projectile (which continues to track the vehicle) is now partly in shadow. By MT, the shadow meets the standard of a causal process, and by CI, the intersection meets the criteria for a causal interaction. Along with other examples, Kitcher believes that he can conclude there are “general troubles in providing an
empiricist account of the justification of causal claims” using Salmon’s formulations (Kitcher, 1989, 469). Salmon would presumably want to appeal to the derivative nature of the shadow’s behavior in relegating it to pseudo-process status, but in response Kitcher could demand an analysis of how the background causal knowledge involved in this judgement was derived from observation.

3.3 Vagueness Objections and Explanatory Adequacy

Several concerns arise from Salmon’s characterizations of causal processes. A causal process transmits its structure, or remains uniform with regard to a characteristic, for an extended period in the absence of interactions. Phil Dowe expressed the concern that this might “exclude many causal effects which are relatively short-lived (Dowe, 2000, 74).” Also, the proviso regarding the “absence of interactions” has been viewed as problematic. Kitcher worries that “virtually all (all?) actual processes are always interacting with other processes (Kitcher, 1989, 463, emphasis original).” All of the macroscopic examples of processes are undergoing incessant interactions of one kind or another (e.g. a baseball’s interactions with air molecules). Even in controlled settings involving microscopic phenomena, it isn’t easy to rule out interactions.\(^\text{76}\) Also, some classical physical theories imply bodies may move under the action of a field: whether this can be interpreted as a continuous series of interactions consistent with Salmon’s definitions isn’t clear.\(^\text{77}\) Considering the examples of macroscopic and microscopic processes also prompts the question of how processes are constituted. As discussed above, Salmon envisions causal process theory as underpinning constitutive as well as etiological

\(^{76}\) Salmon’s formulations don’t even clearly rule out that touching or crossing an empty space-time region could be considered an intersection (see Dowe, 2000, p. 74 and Kitcher, 1989, p. 466).

\(^{77}\) Salmon later claims this is not a problem, but does not offer a detailed discussion (Salmon, 1997, pp. 464-5).
causal explanations. But how do smaller causal processes compose larger ones? Salmon does not offer an account.

Other questions arise from the vagueness of the notions of “structure” and “characteristic.” What kinds of characteristics are permitted? As Dowe notes (Dowe, 2000, 76), the construction of problematic cases involving pseudo-processes that meet Salmon’s criteria often rely on this ambiguity (including his own example of the shadow of the Sydney Opera House “having the property of being closer to the Harbour Bridge than to the Opera House.” The suggestion here is that the definition of causal processes should be restricted by being more specific about “characteristics”: what are the properties which causal processes possess? Salmon does not supply an account of properties.

Even if these concerns about the formulation are overlooked, there is a worry that the particular marks used to identify a causal process may not be relevant to the causal explanation of a phenomenon. Hitchcock presents a simple example involving billiard balls: a player imparts both a blue chalk mark and a quantity of linear momentum to a cue ball when striking it (the cue ball subsequently succeeds in sinking another ball in a

---

78 Elliott Sober had made a similar point in his review of Salmon—see Sober (1987), pp. 253-254.
79 After presenting his shadow cases, Kitcher considers whether putting more restrictions on the kinds of “genuine” properties causal processes can be said to possess would help distinguish them from “gerrymandered” cases, but concludes there will be difficulties in doing this in a principled way (a problem he thinks parallels the issues raised by Goodman’s new riddle of induction—see Kitcher, 1989, p. 466). At one point Salmon argues that a notion he had developed in conjunction with his Statistical Relevance theory of explanation would be of use here. The concept was that of an “objectively codefined class” which is “explicated in terms of physically possible detectors attached to appropriate kinds of computers that receive carefully specified types of information” about properties of processes (Salmon, 1994, p. 300 and see Salmon, 1984, p. 68). While his (brief) description of this criterion would rule out some characteristics (including relational properties such as that possessed by the Opera House shadow), it does not appear it would rule out counterexamples such as the shadow case discussed in section 3.2 without further restrictions on the notion of property (the distorted shape of the shadow would seem detectable—cf. discussion in Dowe, 2000, pp. 77-78). It also isn’t clear it avoids a circular appeal to an implicit causal notion in its use of “detection”.
pocket). The bare-bones causal theory itself does not give us the resources to decide which feature of the cue ball is relevant to the explanation of the sinking of the ball.\textsuperscript{80} In more complicated situations, the problem may become more acute. It is difficult for the notion of a mark to convey the richness of changes possible in a complex system.

4. Conserved Quantity Theories

When discussing the mutual modification of processes involved in a causal interaction, Salmon says that “there is some correlation in the changes that occur in them (Salmon, 1984, 169).” Going on, he notes “in many cases—perhaps all—energy and/or momentum transfer occurs, and the correlations between the modifications are direct consequences of the respective conservation laws (Salmon, 1984, 169-170).” As Salmon himself notes in a later paper (Salmon 1994) these comments are suggestive of another approach to defining causal processes and interactions—one which would explicitly employ conserved quantities.\textsuperscript{81}

4.1 Dowe’s Theory and Salmon’s Revision

In his (1992), Phil Dowe criticizes Salmon’s theory and proposes a new process account of causation using the notion of conserved quantities. In doing so, Dowe looks to address the Humean/empiricist concerns about Salmon’s approach and also address the

\textsuperscript{80} See Hitchcock (1995), p.310. This concern can be extended to Salmon’s successor theory which utilizes conserved quantities discussed below (see Hitchcock, 1995, p.316).

\textsuperscript{81} In a footnote to his discussion of conservation laws in the 1984 book, Salmon references Fair’s (1979) proposal to use energy transfer as the basis for an account of physical causation. Dowe (1992) also acknowledges the work of Fair (and also Brian Skyrms) when introducing his conserved quantity theory.
vagueness of Salmon’s formulations. For Dowe, a conserved quantity is “any quantity universally conserved according to current scientific theories (Dowe, 1992, 210).” Examples are mass-energy, linear momentum, angular momentum, and charge. Dowe offers these new definitions:

1) A causal interaction is an intersection of world lines which involves exchange of a conserved quantity.

2) A causal process is a world line of an object which possesses a conserved quantity.

Dowe clarifies that the concept of “exchange” intended here is distinct from the idea of “transfer:” it is “weaker than [transfer]; it means merely that there are corresponding changes in the values of the physical quantities (Dowe, 2000, 110).” With regard to the reference to an “object”, Dowe say that: “an object can be anything found in the ontology of science (such as particles, waves, or fields), or common sense (Dowe, 1992, 210, emphasis original).” Dowe sees it as an advantage that these definitions do not utilize counterfactuals, and he believes any notion of “hidden powers” is avoided. The definitions also offer a more precise basis for distinguishing causal processes from pseudo-processes. To begin with, “shadows...do not possess conserved quantities (Dowe, 2000, 112).”

---

82 Dowe’s fullest treatment was given in his later book, Physical Causation (Dowe, 2000). The book includes a detailed exploration of how causal interactions and processes underpin the analysis of cause and effect in a wide variety of scenarios. A particularly controversial aspect of this work was Dowe’s attempt to analyze omissions and preventions in light of his theory (see Dowe, 2000, Ch. 6).

83 This version is from Dowe (2000), p. 90. It differs from the 1992 version only in that Dowe replaces the term “manifests” with “possesses.”
In his 1994 paper, “Causality without Counterfactuals”, Salmon abandons the mark-transmission theory:

When the mark criterion was clearly in trouble because of counterfactual involvement, it should have been obvious that the mark method ought to be regarded only as a useful experimental method for tracing or identifying causal processes…but that it should not be used to explicate the very concept of a causal process. Dowe took the crucial step. He pointed out that causal processes transmit conserved quantities; and by virtue of this fact, they are causal. I had come close to this point by mentioning the applicability of conservation laws to causal interactions, but did not take the crucial additional step. (Salmon, 1994, 303)

Salmon goes on to develop a new positive account similar but not identical to Dowe’s. Salmon offers three definitions (the first of which matches the first of Dowe’s).\(^{84}\)

1) A causal interaction is an intersection of world-lines that involves exchange of a conserved quantity.

2) A causal process is the world-line of an object that transmits a non-zero amount of a conserved quantity at each moment of its history (each spacetime point of its trajectory).

3) A process transmits a conserved quantity between A and B (A ≠ B) if and only if it possesses [a fixed amount of] this quantity at A and at B and at every stage of the process between A and B without any interactions in the open interval (A,B) that involve an exchange of that particular conserved quantity.

\(^{84}\) These were refined subsequent to the 1994 paper. The definitions here are from Salmon (1997).
The way the two theories handle the persistence (or propagation) aspect of causation is the biggest difference. In his 1994 discussion of Dowe’s theory, Salmon complains that the formulation has “abandoned one of the most fundamental ideas about causal processes, namely, that they transmit something (e.g., marks, information, causal influence, energy, electric charge, momentum) (Salmon, 1994, 306, emphasis original).” Dowe speaks only of objects manifesting or possessing a conserved quantity. For this reason, Salmon uses “transmits” in his second definition, and then adds his third definition, which embodies the “at-at” conception of transmission.

As in the earlier theory, the at-at conception invites the objection that the proviso of no interactions is unrealistic, especially for macroscopic processes. Salmon responds first by defending the idea that it is a workable notion at the microscopic level:

A gas molecule constitutes a causal process between its collisions with other molecules or the walls of its container. When it collides with another molecule, it becomes another causal process which endures until the next collision. A typical value for the mean free path is $10^{-7}$ m, which, though small, is much greater than the size of the molecule. (Salmon, 1994, 308-309)

For larger systems, on the other hand, “definition 3 should be considered an idealization (Salmon, 1994, 309).” Pragmatic considerations determine whether to consider a larger system as a single process, or as a complex network of processes and interactions. In his (1997), Salmon argued that these pragmatic issues of applying the theory should not be taken as a criticism of his formulation.

Dowe uses a different approach. He does see the dual-aspect (propagation/production) view of causation as a key advantage of process theories. However, Dowe is troubled by the practical issues Salmon faces in his account of
transmission. In Dowe’s approach the burden of explicating the propagation aspect of causation is placed on his account of objects. While Salmon uses the term “object” in his updated formulation, it is little more than a placeholder for whatever existent might be associated with the transmission of quantities. For Dowe “there is an implicit restriction on what counts as an object—it must display identity over time (Dowe, 2000, 100).”85 Whether an object is one encountered in a scientific context or known via “common sense”, it must maintain its identity through time regardless of its interactions or changing amounts of conserved quantities possessed. This creates a theoretical burden to explicate this concept of identity (discussed in Dowe, 2000, pp. 102-109). One option that is not available in the present context is to use the notion of causation to connect temporal parts of a four-dimensional object. Dowe concedes that these issues are difficult (and connected to contentious debates among metaphysicians) and leaves its ultimate resolution open. He believes it is sufficient for the formulation of his causal theory that we have a good working notion of object identity through time in practice.

4.2 Reliance on Scientific Laws/Generalizations

The move to the conserved quantity (CQ) approach represents a substantial conceptual change for Salmon. A primary motivation for his first account of causation was the inability of the covering law approach to provide an adequate framework for scientific explanation. Under the ontic conception, events are explained by locating them in the network of causal processes and interactions. Regularities in the causal network

---

85 The importance of objects for Dowe can also be seen by noting that in his formulation, causation could also be described in terms of them: objects that possess conserved quantities would be causal objects, while those that do not could be labeled pseudo-objects. A causal process would be then be derivatively defined as the worldline of a causal object (see discussion in Dowe, 2008). Also, because of this formulation, Dowe is willing to say that “an earlier segment of a causal process may be a cause of a later segment; and some feature of a causal process may be the cause of another feature (Dowe, 1992, 211).” Salmon, taking causal processes as basic units, would disagree with such statements.
form the basis for theoretical generalizations or laws. The CQ account, however, inverts this picture by taking a set of physical laws (or at least true generalizations) and making them the basis of causation.

It is true that Salmon had included in his 1984 book the idea that the changes resulting from causal interactions might be described by conservation laws. In fact, at times he writes that physical laws “govern” causal interactions. In context of the original theory, however, this talk is loose: he sometimes says that laws govern and that they describe causal regularities in the same passage.\textsuperscript{86} On the other hand, in formulating his mark-transmission account, Salmon clearly emphasizes that the primary notion involved is singular causation:

A causal process is an individual entity, and such entities transmit causal influence. An individual process can sustain a causal connection between an individual cause and an individual effect. Statements about such relations need not be construed as disguised generalizations. (Salmon, 1984, 182)

The last statement is in tension with the CQ theory. In his discussion of this point, Dowe asserts that his theory still can be thought of a singularist in a sense: “whether something is a causal process depends only on local facts about the process, namely, the object’s possession of a conserved quantity (Dowe, 2000, 96).” Still, the theory is committed to the truth that conserved quantities exist and they are described by universal conservation

laws.\textsuperscript{87} Dowe concedes that the theory would be refuted “if it turns out that there actually are no conservation laws (Dowe, 2000, 97).”

In his 1994 paper, Salmon lodges some objections to Dowe’s account with regard to how conserved quantities are characterized. First, he objects to Dowe’s reliance on current scientific theories, and second, to his invoking of conservation laws. With regard to the first point, he says that given the fallibility of our theories we should only look to the current ones to tell us “what quantities we can reasonably regard as conserved […] whether or not they are conserved is another question (Salmon, 1994, 309).” Dowe, in his (2000), did modify his account to clarify that current theories were to be viewed only as our best guides to the identity of universal conservation laws and their associated quantities. On the second point, Salmon warns that if a theory of causation is based on conservation laws, it will inherit all the problems involved with providing a theoretical account of laws. Salmon thinks the theory should avoid the question of drawing a distinction between the existence of accidentally true generalizations and lawful generalizations. Dowe disagrees because he thinks that if there is a merely accidentally universally conserved quantity, then that “should not enter into the analysis of causation (Dowe, 2000, 95).” It is unclear why this is so, if Dowe is willing to let his theory be contingent on empirical investigation.\textsuperscript{88} If he does insist on laws, then it is arguably a particular failing of his discussion that he does not provide an account of laws.

But even on Salmon’s law-free interpretation, his revised theory does depend on the assumed existence of certain true generalizations, departing from his earlier approach.

\textsuperscript{87} Dowe assumes that quantities must be universally conserved and cannot just be defined as those constant within closed system. A proposal regarding how the notions of conservation laws, conserved quantities and closed systems should be best construed for the purposes of the causal theory can be found in Choi (2003).

\textsuperscript{88} See McDaniel (2002), p. 261 for discussion of this point.
This raises two concerns. First, given this commitment, it is not clear that the CQ theories are free of concern about counterfactuals and non-Humean causal concepts. The notion of a truly conserved quantity may be implicitly invoking these notions. To assert that a quantity associated with a particular process is a conserved quantity can be seen as making a statement about the ability of the process to enter into certain sorts of interactions, whether or not they have actually been observed to occur.\textsuperscript{89} By refusing to countenance reference to conservation “laws”, Salmon attempts to resist bringing in a modally potent notion of conserved quantities. But why should we assume any quantity is truly universally conserved based on regularities we have observed? Dowe and Salmon do not justify this assumption. Note that the present objection is specific to the special context of fundamental theorizing about causation. Certainly it is warranted to accept many scientific generalizations as well-confirmed for most purposes. But Dowe and Salmon bear a higher burden: confirmation theory is itself highly contentious, and by ignoring these difficulties and assuming the existence of true generalizations, they provide an inadequate philosophical foundation for a causal theory. It is worth noting here that both Salmon and Dowe reject regularity theories of causation,\textsuperscript{90} and yet the conserved quantity theory indirectly rests on a set of regularities.

The second, related, problem is that the theory is vulnerable to any scientific development which brings the generalizations into question. This is not an idle concern. Even with just current theories in mind, general relativity greatly complicates the notion of universal conservation of energy: only in very special cases and with various caveats

\textsuperscript{89} Handfield (2010) poses this question.
does the concept appear to hold (see discussion in Rueger, 1998). Even if this immediate concern is overcome, the general worry persists that these apparent generalizations may be unmasked as merely approximately true. This raises the question: would causation truly be an idle concept in the absence of universally conserved quantities? The conserved quantity theory is not a conceptual analysis: it relies on a derivative of scientific theorizing. While Dowe (who calls his work an “empirical analysis”) may accept this, it is not an approach which can then serve in turn as a general foundation for a philosophical account of scientific explanations that do not invoke conserved quantities.

4.3 Explanatory Adequacy and Other Criticisms

In addition to this conceptual shortcoming, CQ theories suffer from the limited practical relevance of conserved quantities to a wide variety of causal contexts. Many scientific explanations require reference to characteristics of causal processes and interactions not readily reducible to physics. As a result, commentators who otherwise admire the causal process account don’t see it as useful: “Although...Salmon’s analysis may be all there is to certain fundamental types of interactions in physics, his analysis is silent as to the character of the productivity in the activities investigated by many other sciences (Machamer, Darden, & Craver, 2000, 7).” Related to this concern, Salmon has still not provided an account that would shed light on how larger processes are constituted from more basic ones. Dowe also doesn’t consider this question, but relies on

---

91 Dowe responds by claiming that as far as we know the appropriate conditions hold in the actual world (Dowe, 2000, p. 97). This response arguably doesn’t do justice to the complexity of the issue. Failure of the symmetries to hold for the purposes of defining causality locally may result from the character of distant and unobservable reaches of the universe. Also see Hoefer (2000) for an argument that there is no principle of energy-momentum conservation in General Relativity.

92 For Dowe’s discussion of empirical analysis and its difference from conceptual analysis, see Chapter 1 of Dowe (2000). For criticism see Hausman (2002), and Koons (2003).

93 See, e.g., the discussion in Woodward (2003), pp. 354-6.
the relatively undeveloped notion of an object when discussing phenomena at various scales. Also, even within appropriate physical contexts, the CQ theory is vulnerable to an objection to the mark-transmission theory discussed above (in section 3.3): while we may identify a conserved quantity exchanged in an interaction, an adequate explanation may depend on some different exchanged quantity.94

An unmet challenge for the conserved quantity theory concerns the explanation of indeterministic causation. The theory is meant to encompass both deterministic and irreducibly indeterministic processes, but Dowe does not delve deeply into this topic:

This is a feature that needs development. Whether that development will be Humean depends, of course, on how it is done. I expect that the probabilistic element in the theory must enter as a propensity...However, contra Salmon and as indicated above, I take it that propensities should be regarded as referring to the operations of objective, indeterministic causal processes, such that the propensity takes on values between 0 and 1 only where there is genuine indeterminism. These propensities would supervene on indeterministic facts about the world, but it remains to be shown how or whether they avoid hidden powers. (Dowe, 2000, 113-114).

Salmon’s discussion of his conserved quantity account does not revisit his earlier discussion of propensities.

---

94 It is also worth noting that despite its ambition to improve on the mark-transmission account in this regard, the CQ theory still faces counterexamples regarding pseudo-processes. See Hitchcock (1995) and Choi (2002).
4.4 Contemporary Status of Process Theories

Within the tradition of attempts to provide an analysis of causation, process theories have taken their place as one of several intriguing but flawed programs developed in the last few decades. Contemporary surveys will typically describe the conserved quantity version of the approach, while Salmon’s original account is discussed in a historical context, given the author’s ultimate abandonment of the theory. In their recent critical exploration of the philosophy of causation, L.A. Paul and Ned Hall discuss the CQ theory (along with transference accounts) as one of several categories of “rival approaches” (Paul & Hall, 2013). Among the flaws that Paul and Hall find most salient in the CQ approach (among those discussed above) are its limitations in providing a basis for explanations and also the narrowness of its conceptual reach:

Even if these views correctly describe the actual world, surely there could be worlds with laws that don’t single out anything as a “conserved” quantity—more generally, that do not describe the transfer of anything physically fundamental. (Paul & Hall, 2013, 55)

The authors thus prefer theorizing which has the goal of a broader analysis. They also hold out hope that such an analysis can also nevertheless succeed in reducing causal concepts to non-causal ones (see Paul & Hall, 2013, pp. 253-256). This last ambition, of course, has resisted the best efforts of many philosophers.

---

95 The recent volume entitled *The Oxford Handbook of Causation* includes a chapter on process theories (authored by Dowe) under the section heading of “Standard Approaches to Causation” alongside four other such chapters (Beebee, Hitchcock, & Menzies, 2009).
5. Causal versus Scientific Theorizing and the Implications for Process Theory

In his 1984 book, Salmon also includes a chapter on the topic of scientific realism. In it, he endorses a realist stance toward unobserved theoretical entities. The argument he uses relies on prior acceptance of his causal theory. In other words, given the existence of the network of causal processes and interactions, Salmon can make a case for a variety of scientific realism. This discussion is helpful because it highlights from another perspective the challenge Salmon faces in theorizing about causation itself. Salmon offers an intuitive picture of the scientific enterprise that focuses on the role of experimental interventions in uncovering the causal structure of the world. However, while Salmon doesn’t acknowledge the implication, his picture also places human investigators squarely inside the causal network. As a consequence, his frequent appeals to experimentation cannot provide fully independent justification for attributing his causal concepts to natural phenomena. Still, the attractiveness of Salmon’s philosophical framework suggests a strategy for revising his approach in a way that accepts this limitation on causal theorizing. This is a preferable alternative to the path Salmon actually pursued.

5.1 Salmon on Unobserved Theoretical Entities

Salmon points out that if a task of science is to provide explanations, then one must take a stance on the status of unobservable theoretical entities appealed to in these explanations. Salmon begins his discussion by assuming what he calls “physicalism”:

---

96 This is to be distinguished from phenomenalism, which holds that the only real entities are human sensations (see Salmon, 1984, p. 5).
[Physicalism] holds that perception (of a fairly direct sort) provides us with reliable—though not incorrigible—knowledge of ordinary middle-sized physical objects and events. We have, I believe, sound reasons for taking such entities as flowers, rocks, and sneezes to be real. (Salmon, 1984, 5)

Salmon also accepts that we can infer the continued existence of observable entities when they are unobserved by us, saying that we legitimately “endow our [everyday] world with a great deal of spatiotemporal continuity (206).” It is clear we also readily presume the existence of unobserved interactions by reference to observed continuities. These attitudes carry over into our scientific investigations:

In conducting the Compton scattering experiment, for example, one checks for a correlation between photons scattered at a certain angle and electrons ejected with a particular kinetic energy. The observed coincidences provide an entirely satisfactory basis for inferring that unobserved collisions between incident photons and (for all practical purposes) stationary electrons have occurred. (Salmon, 1984, 210-11, emphasis original)

We extend our realist assumptions about entities and interactions in the macroscopic world by analogy to the microscopic one.

As a case study for understanding what drives acceptance of a realist stance, Salmon looks carefully at the events which led most scientists to accept the existence of atoms and molecules by the early part of the 20th century. His initial argument here is that the wide variety of independent experimental findings which had been gathered
supported the attribution of a shared cause in the form of the discrete micro-particles.\footnote{In an earlier paper, Salmon originally interpreted the case as one of inference to a common cause. Van Fraassen (1980) criticized this: at best one can only infer similar causes, not the very same cause (p.123). Salmon then revised his interpretation to provide for physically distinct causes that nevertheless arose from common “background conditions” (1984, p. 226). A discussion of this topic with similarities to Salmon’s is in Cartwright (1983), pp. 83-86.} But another important ingredient is crucial to the case. Each of the experiments is readily characterized in terms of particular causal processes and interactions connecting micro-particles with the experimental outcomes:

There is little difficulty, I think, in seeing that the required causal processes do connect the existence and behavior of the micro-particles with the experimental results that furnish the basis for the calculation of Avogadro’s number. In the case of X-ray diffraction, for example, electromagnetic radiation of known wavelength is emitted from an X-ray source, it travels a spatiotemporally continuous path from the source to the crystal, where it interacts with a grating in a way that is well established in optics. The diffracted radiation then travels from the crystal to a photographic plate, where another causal interaction occurs, yielding an interference pattern when the plate is developed. (Salmon, 1984, 223)

Salmon concludes that all of the experimental approaches to finding Avogadro’s number “patently involve similar sorts of causal processes and interactions (223).” Salmon then makes the following remark:

Indeed, it is by virtue of the causal properties of atoms, molecules, and ions that the various experimental ascertainties of [Avogadro’s number] are in principle possible. (Salmon, 1984, 223)
Note that here Salmon does not show his typical caution about attributing causal properties (reminiscent of “hidden powers”) to natural phenomena.

Salmon goes on to discuss how and whether realism about theoretical entities can be made consistent with a form of empiricism. In the particular framework Salmon chooses, the key question is whether empirical evidence of the sort discussed in his examples justifies inferences to statements about unobservables. Attempts to derive such statements deductively from statements about observables have founndered, so Salmon wants to explore whether another kind of inference can work. He concludes that an inductive inference can succeed if it includes two added ingredients: a causal component and an analogical one. First, the theoretical aspects of the phenomenon must be amenable to description in terms of causal processes and interactions. Second, via analogy, the processes and interactions involved in the relevant experiments on unobservables can be seen as extrapolations from causal systems that we can observe. So, the inferences depend on the existence of a unified causal network:

If the account of causal processes, causal interactions, and causal forks offered [here] is anywhere near correct, then causal mechanisms frequently involve unobservable entities. (Salmon, 1984, 238)

And keeping in mind Salmon’s criticism of Russell for focusing on causally-based inferences as such, we should regard these causal mechanisms as objectively real. Leaving

---

98 He refers to this as “statement empiricism,” which is distinguished from “concept empiricism” (Salmon, 1984, p. 230).
99 Salmon says he agrees with Cartwright (1983) regarding her similar argument that “inference from effect to cause is legitimate” in such situations (Salmon, 1984, p. 233). Salmon also cites Ian Hacking’s (1981) analysis of microscopes, quoting at length from Hacking’s discussion of the use of an engineered microscopic grid (pp. 235-6). Indeed Salmon’s thinking shows a strong affinity to Hacking’s advocacy of “entity” realism set forth in the latter’s (1983).
aside the epistemological question of whether his formulation of a special kind of inference is sound, we can note that Salmon’s theoretical scientific realism clearly depends on a prior commitment to realism about the causal network that he has described. But the question remains, how can this latter commitment be justified?

5.2 Scientists in the Causal Network

As seen, Salmon thinks realism may be extended to unobservable entities given his causal interpretation of the experimental investigation involved in his examples. While his discussion is laden with a traditional reference to “observation” and its role in acquiring knowledge, a distinctive feature is his active conception of how scientists probe nature to ascertain its causal structure. Likewise, when Salmon presents his mark-transmission theory, experimentation also plays an important role. As discussed above, he argues that we can conduct experiments that distinguish causal processes from pseudo-processes, and causal interactions from mere intersections (see 149-150, and 172-174). This is also not a matter of passive observation. In discussing one putative counterexample involving shadows, he notes that mere “sitting around watching the shadows” is insufficient, but an experimental intervention can get the case right (173). Generally, the counterexamples to Salmon involving pseudo-processes which meet his MT and CI definitions presume passive observation rather than allowing for the use of creative experimental tests that Salmon envisions.

Here, however, we need to recall Kitcher’s criticism of Salmon’s appeal to experimentation.\(^{100}\) The complaint is that the experimenter is relying on background causal knowledge, and this itself needs justification. Setting aside the epistemological

\(^{100}\) Kitcher (1989), pp. 475-77.
formulation of Kitcher’s critique, it does appear that’s Salmon’s attempt to rely on experiments threatens to make the approach anthropocentric or subjective rather than objective. Experimentation is a human activity, after all. On this point, Salmon responds as follows:

Although we may often take an active role in producing a mark in order to ascertain whether a process is causal...it should be obvious that human agency plays no essential part in the characterization of causal processes or interactions. We have every reason to believe that the world abounded in causal processes and causal interactions long before there were any human agents to perform experiments. (Salmon, 1984, 174)

But if Salmon can avoid the charge of anthropocentrism, there is still an apparent challenge of circularity. Salmon would presumably agree that human beings meet his criteria for causal processes just as well as any macroscopic object, and human interventions are also causal interactions. Then experiments can be seen as a kind of causal interaction taking place within the network. There is no way, then, to determine whether a process is causal from some isolated vantage point. The framework which provides justification for scientific realism cannot be used to justify the attribution of causation to natural phenomena in the same way.

5.3 An Alternative Path Forward for Process Theory

The way in which this dilemma played out historically was discussed above: Salmon bowed to criticism that the counterfactuals used in his formulations could not be

---

101 Salmon’s theory would then be subject to criticisms similar to those lodged against certain agency and manipulationist theories of causation (see discussion in Woodward, 2009, pp. 25-28).
given adequate justification. In the paper where he makes this concession, Salmon explains that the key insight the formulations were mean to embody is that “a process is causal if it is capable of transmitting a mark, whether or not it is actually transmitting one (Salmon, 1998, 303, emphasis original).” Given the problems faced by the account, Salmon’s own diagnosis is that the mark method should not have been used to explicate the concept of the causal process:

The fact that [the causal process] has the capacity to transmit a mark is merely a symptom of the fact that it is actually transmitting something else. (Salmon, 1998, 303, emphasis original)

Referring to various comments made in his earlier book, Salmon says that the “something” referred to here was labelled at times “information”, “structure”, and “causal influence” (see 154-157). He goes on to propose that conserved quantities are what is transmitted.

The turn to the conserved quantity approach takes the theory in a very different direction than implied by the broader philosophical commitments which accompany the presentation of the mark-transmission theory. As discussed in section four above, the move was inconsistent with the ontic approach to scientific explanation which originally motivated Salmon’s development of the theory. Similarly, we now see its inconsistency with Salmon’s discussion of the role causation plays in supporting scientific realism. Instead of tracing the network of causal processes and interactions involved in experimentation in order to underpin inferences about theoretical entities, the presence

---

102 In a later paper, Salmon shows a way to formulate the mark-transmission theory without explicit counterfactuals but with a provision stating “A causal process is a process that can transmit a mark (Salmon, 1998, 299).”
of conserved quantities—theoretical constructs—are inserted back into the theory of causation. It would seem much more in the spirit of his earlier view to propose that causal processes transmit something like “causal influence”, and that it is the role of science to explicate the presence and nature of this influence. One can try to make this notion more precise, but to use particular products of science in the formulation is inconsistent with Salmon’s intuitive and appealing understanding of the relationship of scientists to the world they investigate.

The cost of accepting this conclusion, however, is to give up on the goal of a fully reductive theory of causation. We cannot completely characterize the nature of causal processes in non-causal terms. However, this may be a sensible limitation to accept. While I won’t attempt to argue that all attempts at reduction must fail, I do suggest that the failure of Salmon’s attempt calls into question the kind of empiricist-style reduction invoked above by Kitcher. Salmon’s focus on active experimentation to test for causal processes puts into sharp relief the circularity which ensues if humans are also causal processes. But the result would seem to generalize to the traditional empiricist approach. To recall, this is the notion that one attempts to give “necessary and sufficient conditions for the obtaining of causal relations...by deploying only concepts whose satisfaction is observationally ascertainable (Kitcher, 1989, 460).” But this formulation presumes not only that causation cannot be directly observed but also that human observation itself stands outside the causal network. If all human activity (including sense perception and cognition) is part of the natural world, and if we decide the basic structure of that world is

---

103 It seems clear that merely moving to alternative terminology (invoking, say, “information”) cannot offer a path out of the dilemma.
aptly described as “causal”, then it seems unlikely that we can also coherently analyze causation in non-causal terms as Kitcher’s formulation demands.¹⁰⁴

Turning back to process theory, a different way to develop the original account would take its cue from the various times Salmon attributed powers to causal processes, in the guise of abilities/capabilities (e.g. p. 147), propensities (203), dispositions (204) and causal properties (223). This alternative approach would stipulate that causal processes do feature a capacity for (or perhaps a collection of dispositions or propensities toward) participation in particular modifying interactions. This approach may form the basis of a non-reductive but illuminating theory: one which overcomes various shortcomings of the mark-transmission formulation while also providing a superior foundation for an account of scientific explanation.

¹⁰⁴ For reasons of space I do not discuss Kantian approaches to responding to the empiricist dilemma.
Chapter 4: The Causal Nexus and its Role in Explanation

Scientific explanations seek to create understanding in part by employing the concept of causation. Actually, as discussed in Chapter 2, there is strong evidence for the existence of two independent concepts of causation (difference-making and production), and I argued there that explanatory relations based on both concepts have a role to play. However, after assessing criteria for their applicability, I further argued that production (“bringing about change via a connection”) takes precedence in the physical settings found in the natural sciences. Chapter 2 concluded by noting that there is a need for a theory that both fleshes out the notion of production and details the role it plays in natural science explanations. In recent decades, the most detailed attempt is Salmon’s (1984) causal process theory and his related view of explanation. The causal theory includes two components: propagation (transmission of causal influence by entities called processes), and production (changes in processes due to their causal interactions). Proper explanations in Salmon’s view are simply those which exhibit the processes and interactions responsible for a phenomenon. In Chapter 3, I critiqued Salmon’s original process theory as well as later versions, which founder primarily due to an insistence on certain empiricist restrictions on theorizing about causation. I suggested there that a different, non-reductive approach to process theory can be more successful. Importantly, however, Salmon’s related conception of explanation is also inadequate. For this reason, the task of constructing a revised causal process theory must be shaped by an updated account of how such a theory underpins explanation.

Below, I describe and defend a new approach. In section one, I briefly review the debate over the nature of explanation, beginning with Salmon’s taxonomy of ontic and
epistemic conceptions. After endorsing Illari’s (2013) broad perspective on the debate over these conceptions, I detail how a causal theory should provide the basis for the “ontic constraint” on good causal explanation. A comparison to Weisberg’s (2013) account of scientific modeling inspires a view of the actual causal structure of the world as the target onto which the causal elements of a scientific explanation should map. I next argue that a singular causal network is the appropriate target even for explanations aspiring to wide scope of application. Section two begins describing the causal theory by showing how elementary causal processes can be defined in terms of dispositions toward various particular interactions (“dispositional causal process theory”). All interactions involve mutual manifestation of dispositions of two or more processes. Simple physical examples are used to demonstrate how components of a scientific explanation map onto corresponding elements in this conception of the causal nexus. In section three, I clarify the causal theory by discussing the role of causal properties: rather than ontological posits, the properties referred to in causal scientific explanations represent idealized regularities in the dispositional profiles of causal processes. Section four describes composite causal processes and their role in providing a target for constitutive explanations. Patterns of interaction among processes provide the raw material for a hierarchy of composites. Section five compares aspects of the theory to the work of Cartwright and Wimsatt and considers several objections. Finally, section six argues for the superiority of the approach over that of “new mechanist” theorists.

1. Goals for a Causal Theory in Service of Explanation
In developing a new version of causal process theory, it is necessary to clarify its objectives. First, in keeping with the discussion in Chapter 2, the scope of the task is limited to developing a theory of production (or “productive causation”). As in Salmon’s theory, the concept of productive causation—*bringing about change via a connection*—is to be fleshed out in terms of propagating processes and their interactions. Second, in light of Chapter 3’s critique of prior accounts, a non-reductive approach will be taken: in particular, the account will develop Salmon’s occasional suggestion (particularly in his discussion of indeterministic phenomena) that causal processes transmit dispositions (or propensities) toward certain interactions. Finally, there is an additional limitation on the ambitions of the present account as a theory of causation: its only goal is to support explanations in the natural sciences. As such, it will not attempt to provide an analysis of other kinds of causal claims. Also, the theory will avoid metaphysical commitments that are unnecessary to meet the objective. In the course of the discussion below, I will argue that the approach should be consistent with a number of views regarding “deeper” or more fundamental ontology.

With regard to this relationship between the causal theory and scientific explanation, there is a significant difference versus what was intended by Salmon. Given Salmon’s so-called ontic conception of explanation, a causal account was to directly supply the content of proper explanations. In the present case, the theory is not meant to precisely spell out what an explanation should be. Rather, as detailed below, its normative role is to supply the *ontic constraint* on causal explanations. Specifically, this means that
while explanations may take a wide variety of forms, the representational elements employed should map onto the causal structure the theory describes.\textsuperscript{105}

1.1 The Causal Nexus as Ontic Constraint and Target

Under Salmon’s ontic conception, to explain something is to exhibit how it fits into the causal structure of the world.\textsuperscript{106} This was offered in pointed contrast to the account of explanation offered by Hempel’s deductive-nomological (“DN” or “covering-law”) approach. The DN account is categorized by Salmon as “epistemic” because the explanation takes the form of an argument (necessarily featuring laws of nature) demonstrating that the phenomenon of interest was to be expected.\textsuperscript{107} Critics have argued, however, that Salmon’s conception of explanation falls short in a number of ways. First, it isn’t clear if it can account for the variety of things that come under the category: an explanation may take a number of forms (e.g. text and/or diagram and/or model), and also can refer to either a communication or a cognitive act.\textsuperscript{108} More to the point, it is argued that it fails to pick out the distinctive role of explanation in the epistemic achievement of understanding:

[Salmon] is right that in mechanistic explanation a scientist appeals to causal relations and mechanisms operative in nature, which are taken to generate or

\textsuperscript{105} In addition to its normative role, the theory also has a modest descriptive dimension: the account is intended to be broadly compatible with causal explanatory accounts offered by scientists. This compatibility is relatively straightforward to establish in cases where the explanation features the causal properties of interacting entities. In other cases the connection will be more indirect. The kinds of examples that are the focus of recent literature on mechanistic explanation are of special interest, and will be discussed below (section 6).

\textsuperscript{106} See Salmon (1984), p. 19. In his (1989), Salmon clarifies that there are two acceptable ways to interpret the ontic conception: either explanations exist concretely in the world, or an explanation consists of a list of sentences or propositions that “report” the relevant worldly facts (p. 86).


\textsuperscript{108} See discussion in Craver (2014). Also see Potochnik (2016, 2017), who emphasizes the communicative and cognitive aims of explanation.
produce the phenomena being explained. However, it is important to note that offering an explanation is still an epistemic activity and that the mechanism in nature does not perform any explanatory work. (Bechtel, 2006, 33-34)

On the other hand, Salmon’s insight has been defended, most prominently by Craver, who argues that “the philosophical theory of explanation depends fundamentally on an ontic conception of explanation, that is, on a view about the kinds of structures of the world that count as genuinely explanatory (Craver, 2014, 51).”

Illari (2013) argues that the debate between defenders of the ontic and epistemic conceptions is unproductive given that neither alone seems to offer a full understanding of what it means for something to be a causal explanation. Instead, she thinks the debate should shift to the ground of theorizing about the normative constraints on good explanatory practice: good mechanistic explanations, for instance, need to meet both ontic and epistemic constraints. This aligns with the idea that there are two dimensions involved when an audience achieves scientific understanding via an explanation (see Chapter 1). The first involves a kind of connection between the audience and the actual worldly structure represented in the explanation, while the second is about the audience’s cognitive grasp of that structure via tailored communicative methods and representational forms. Acknowledging the importance of both dimensions, one might then debate whether there is a sense that either the ontic or epistemic set of constraints is more fundamental than the other or can otherwise claim priority.109 Illari thinks there is

---

109 The terminology of “ontic” and “epistemic” constraints (inherited from the literature) is not optimal, since understanding is an epistemic achievement primarily in virtue of meeting the “ontic” constraint, while the “epistemic” constraint has more to do with the cognitive aspects of understanding. Below, I focus on the nature and role of the ontic (worldly) constraint, but will continue to refer to “epistemic” constraints (or virtues/norms) as a catch-all for the other considerations that facilitate understanding in the cognizing audience.
no good case for asserting such a priority, and argues for a picture where good explanatory practice simultaneously satisfies both kinds of constraints (Illari, 2013, 250).

This reframing of the debate in terms of normative constraints is apt. With regard to priority, however, a case can be made that the ontic constraint has at least a methodological kind of priority in the case of causal explanations. Here, the ontic constraint is provided by the actual causal structure responsible for a phenomenon, and good explanatory practice would seem to begin with an assessment of this structure. To the extent one is motivated to deviate from faithfully representing the actual causal structure in an explanation, it would be because the epistemic value of doing so offsets the loss of ontic fidelity. A useful comparison can be made to the account of target-directed models offered by Weisberg (2013). In target-directed modeling, the modeler has a particular real-world system in view, and the goal is to “generate predictions and explanations about this specific target.”

According to Weisberg, models are composed of a structure and the scientist’s interpretation of that structure (24). The interpretation, which Weisberg calls a construal, includes the notion of assignment—an assignment specifies how elements of a target system map onto parts of the model (39). The interpretation also includes a determination regarding intended scope, that is, the choice of just which aspects of the target system to represent. There are a great variety of options here, but one approach focuses on an abstracted subset of properties that includes those “of primary interest, and every factor that is causally linked to the properties of primary interest (92)”. Finally the

\[\textit{\textsuperscript{110}} \text{ It should be kept in mind when speaking of a “causal structure responsible for a phenomenon” that this can have two dimensions. The etiological aspect concerns identifying antecedent causes of the phenomenon, while the constitutive aspect concerns the causal network comprising a phenomenon.} \]

\[\textit{\textsuperscript{111}} \text{ The close relationship between models and explanations is explored by Rohwer & Rice (2016).} \]
interpretation includes fidelity criteria. In the case of criteria for representational fidelity, “typically, these…specify how closely a model’s internal structure must match the causal structure of the real-world phenomenon to be considered an adequate representation (39).” Modelers, however, also employ idealization, which is the “intentional introduction of distortions into scientific representations (98).” Weisberg describes several varieties of idealization and their purposes. For target-directed models, idealizations are primarily motivated by pragmatic considerations or epistemic ones, as in the case of “minimalist” idealization, which seeks to include in the model only “core causal factors which give rise to a phenomenon” (100). The idea is that by featuring the most important causal factors, one can best serve explanatory goals. In Weisberg’s discussion, we can see how the practice of modelling often involves balancing the motivation to faithfully represent the actual target phenomena on the one hand, and epistemic considerations on the other. We can see here an analogue of the discussion of the roles of ontic and epistemic constraints in forming causal explanations. Connecting this to the question of priority, we might say that while the end result of modelling reflects this balancing (as well as many other factors), there is a sense in which the actual phenomena, the target system, is what the work is directed toward, and thus is the starting point for the practitioner. For causal explanations, which may sometimes employ

---

112 The other type of fidelity criteria concern dynamical fidelity: this has to do with how well the model produces an output similar to the real world system (see Weisberg, 2013, p. 41).
114 It can be debated how faithfully the elements of a good explanation must represent the target. As discussed further below, what is important in the present approach is that elements correspond to features of the world’s causal nexus, even if these elements are highly distorted by idealization. This is similar to Potochnik’s idea that idealizations can “play a positive representational role” and can “stand in for significant causal influences (Potochnik, 2017, 52).”
models, the ontic constraint—the actual causal structure giving rise to the phenomenon—is the target.\footnote{One caveat in this discussion concerns whether it is presumptuous to assume that the constraint must be conceived of in terms of actual causal structure. Philosophers wary of realism about causation may want to interpret the ontic constraint as another form of epistemic constraint (perhaps properly labeled “meta-epistemic” given its structuring role). Perhaps more importantly, scientists may also have varying levels of commitment to the causal character of the target phenomenon vs. the epistemic value of, e.g., representing it as if it were a causal mechanism. This and related issues are discussed in Matthewson & Calcott (2011) and Levy (2012). For present purposes I will assume practitioners do assume phenomena have actual causes and in light of this are seeking specifically causal explanations.}

1.2 The Case of General Explanation

In some cases, however, assigning methodological priority to the ontic constraint may present an overly simplistic picture of the formulation of a causal explanation. This would be true if some kinds of (good) causal explanations have features with no counterpart in the ontic constraint. A very important case to consider here concerns causal explanations that seek to claim a wide scope of application. As discussed further below, explanations usually depict broad kinds or types of entities, properties, and interactions and frequently employ idealization (or risk making false generalizations) in pursuit of scope. On the other hand, like Salmon’s original process theory, the account introduced in this chapter takes the form of a singular causal network. What is the relationship between the generalizing aspect of explanatory practice and the ontic constraint?

Sheredos (2016) recently argues that ontic constraints play no direct role with regard to important norms involving scope: that is, those that govern the formulation of general or systematic explanations.\footnote{Generality has to do with “the invocation of categorical claims regarding classes” of the elements of explanation, while systematicity has to do with “specification of some principle of extrapolation for applying an explanation to multiple cases (Sheredos, 2016, 922)”. I focus on the notion of generality in the text (and also greatly simplify the detailed discussion in the article).} Whatever claim to priority the ontic constraint has
in the case of explaining a particular phenomenon, this does not extend to these other kinds of explanation. For instance, in striving to account for apparent regularities in the causal structure we may form general explanations that subsume token parts of the structure under a class or type. Here one starts with a general explanatory framework and applies it to instances. With regard to the actual causal structure, unless one wants to commit to the reality of universals, such a generalization goes beyond the ontic structure of the world (if the latter is defined as a network of particular happenings). This pursuit of scope, then, must be related to satisfaction of non-ontic norms, presumably relating to the intuition that greater scope means greater explanatory force. Scientific practice would seem to support treating this search for general explanations as one driven by these epistemic considerations. Looking again at the case of modelling, representational fidelity criteria may be purposefully relaxed in order to develop models that can be applied to many actual and/or non-actual targets. Here, idealization clearly serves an epistemic goal—one that is different from that in the case of minimalist idealization discussed above: it is the attainment of explanatory power that comes with wide application. There is no sense here that the more general model must track something in the world that is responsible for uniting the many (actual and/or possible) particular targets to which it is applied.

So, it is plausible to conclude that the striving for general explanations should be seen as serving epistemic ends, rather than as an attempt to map onto real causal features

---

118 See Weisberg (2013), 109-110, and the account of the trade-off between precision and generality in Mathewson & Weisberg (2009). Note that in these discussions, “intended scope” and “generality” mean different things. The former normally refers to the aspects of a real-world system being modelled, while the latter has to do with the number of targets to which the model may be applied.
119 For a related viewpoint, see Glennan (2017). He argues that mechanistic models are typically general in nature, but considers real world mechanisms to be particulars.
of the world. While a philosopher can make additional metaphysical posits that would supply the missing structure, the theory offered in this chapter will refrain from committing to the existence of universals or other deeper ontological elements that underwrite regularities. The theory takes the form of a singular causal network, and thus does not provide a direct target onto which the general elements simply map. Given this, can the theory still be viewed as providing an ontic constraint (as it does in the case of explanations of particular phenomena)? There is a sense in which it provides a clear if indirect constraint on general explanations. Sheredos suggests explanatory practice involves a dynamic shifting back and forth between ontic and epistemic norms (943). One part of the process of forming general explanations plausibly features testing these against singular candidate instances, and in this way the ontic constraint will still shape the explanation.120 This of course complicates the picture of methodological priority sketched above, but it remains appropriate to think of the ontic constraint as playing the role of ultimate target: all of our causal scientific explanations will be checked against the actual causal structure of the world.

1.3 Summary of the Causal Theory’s Role

A good causal explanation in the natural sciences is one that meets both ontic and epistemic constraints. The ontic constraint is the actual causal structure (or causal nexus) that gives rise to the phenomenon of interest. This structure is also, I argue, the target of the explanation: it is what the explanation is about. For the explanation to be about the

---

120 On this point, it is interesting to look again at the earlier Salmon and Hempel conceptions. This picture of the relationship between general explanations and causal structure is a departure from Salmon’s ontic conception—he thought concrete regularities in the world’s causal structure were explanatory (although his causal theory only defined a singular network). On the other hand, the testability of the statements in a (general) explanation is a requirement of Hempel’s conception (along with the “expectation” criterion mentioned above – see Hempel, 1966, pp. 47-49).
actual causal structure, its representational elements must map onto that structure (allowing for the fact that the mapping of general explanations has an iterative dimension). In what follows, I present a theoretical picture of the world’s (productive) causal nexus, and detail this mapping relationship with causal explanations.

2. Elementary Causal Processes

The causal nexus includes elementary causal processes, which are assumed to be irreducible, and composite processes, which are constituted from the elementary ones. Elementary processes can be assumed to correspond to the elementary particles described in physical theories: these are entities for which scientists do not seek to offer a constitutive explanation. On the other hand, this correspondence should be taken as provisional given the prospect of advances in physics. The ontic structure described here is intended to be compatible with known physics, but is not intended to resolve active debates about what kind of fundamental metaphysical account is most consistent with modern physics. It should be compatible with multiple such “deeper” accounts.

2.1 Elementary Causal Processes – Definitions

Dispositions and their manifestations are employed to define elementary causal processes and their interactions. At the elementary level, the dispositions that characterize causal processes are linked to unique manifestations that then serve to characterize particular causal interactions. Following an idea from C.B. Martin, these manifestations only occur as mutual manifestations involving dispositions characterizing two or more processes—the respective causal processes involved are “reciprocal disposition
partners”.\textsuperscript{121} This is in keeping with Salmon’s idea that causal interactions involve a mutual modification of the processes involved. In addition, while a process featuring a single disposition need not be ruled out, processes typically are characterized by a cluster of dispositions that make it capable of interactions with a wide range of partners. I will use the term compresence to describe a (symmetric and transitive) relation that ties a number of dispositions together. The manifestation of a disposition, in turn, takes the form of an alteration in the make-up of the cluster (i.e. the gain/loss of particular dispositions): this is what it means for a causal process to undergo change. As shorthand, I will call this a change in the dispositional profile of the process.

Definitions:

1. A causal process is something that transmits a cluster of dispositions between space-time points.

2. A cluster of dispositions is transmitted between space-time points A and B (A≠B) if and only if the dispositions are compresent at A and at B and during the interval between A and B.

\textsuperscript{121} See Martin (2008), especially Ch. 5. The mutual manifestation idea stands in contrast to a traditional account where an object is said to possess a disposition for a characteristic manifestation in condition C (but may not manifest or else may have a different manifestation in condition C’). Mutual manifestation (which can involve many partners) takes the place of a reference to conditions. I should note that Martin usually describes a single disposition as capable of myriad mutual manifestations with different pairings of reciprocal partners: a disposition, considered in this way, he sometimes calls a “readiness line” or a “power net” (p. 47). At other times, however, he describes a cluster of dispositions with particular (mutual) manifestations: “quarks have countless dispositions toward countless (nonactual) manifestations (p.50, see also p. 54)”. Other theorists who utilize the idea of mutual manifestation (or a similar notion) include Heil (2003) and Mumford & Anjum (2011). These theorists generally take single dispositions to be capable of multiple manifestations exhibiting a shared character, and don’t distinguish between elementary and macro-level dispositions.
3. A *causal interaction* is an intersection of causal processes at a space-time point where there is mutual manifestation of dispositions, and where the associated modifications to the processes are transmitted beyond the locus of the intersection.

The network of elementary processes and their interactions provide the structure upon which the elements of good productive causal explanations will map.\(^{122}\)

Before moving to examples, some differences with prior causal process theories can be briefly noted (in addition to the employment of dispositions and their manifestations as primitive elements). Both Salmon (1984) and Dowe (2000) used a single set of definitions to cover causal processes of all kinds, including macroscopic ones, whereas here elementary and non-elementary processes are treated separately. Also, Salmon’s concern about distinguishing causal processes from other phenomena that appear to display a consistency of characteristics (so-called pseudo-processes) does not arise in the treatment here. At the elementary level, all processes are causal processes, and these comprise everything else (including us). At the macro-level, our observations of what appear to be pseudo-processes can in principle be re-described by attending to the complex causal interactions actually involved with, say, a visual perception of a shadow.

On the other hand, the distinction made by Salmon and Dowe between causal interactions and “mere” space-time intersections is kept. The possibility that two causal processes (corresponding for example to light pulses) may occupy the same space-time

---

\(^{122}\) By design, there is an etiological but not a constitutive dimension to explanations at the elementary level. Also, note these definitions only encompass what Salmon (1984) referred to as X-type interactions. He also discussed two other types: Y-type (where a process divides into two or more processes), and \(\lambda\)-type, where two or more processes fuse. In the present framework, these can only occur at the level of composite processes.
location without interacting is thus maintained. Salmon’s notion of transmission is also retained (with minor alterations), and a causal process is simply “something” that can transmit dispositions. For present purposes, this is preferable to adopting a more detailed theory of persisting objects (or some other kind of particular) such as that involved in Dowe’s theory (see further discussion of these issues in Section 5).

2.2 Simple Physical Examples

In the simplest cases of scientific explanation, the phenomenon to be explained will be either an entity and its causal properties (corresponding to a propagating process and its dispositions) or a change in an entity’s properties (corresponding to an interaction event), and the explanation will employ elements that correspond to antecedent elements in the causal network. To illustrate, consider a collision scenario described in classical physics: assume the initial conditions feature one particle with a given mass moving at constant velocity toward a stationary particle. Sometime after colliding, each particle has a new velocity, and one seeks to explain this final condition. In the current framework, each particle corresponds to a causal process that, prior to the interaction, is transmitting a disposition toward this interaction (as well as many others). The interaction leads to a changed dispositional profile of the objects toward various further singular interactions. For the scientific account utilizing physical theory, an explanation can be given that cites the initial properties of each particle and the conservation of momentum and also conservation of kinetic energy if the particles are assumed to have characteristics needed for an elastic collision. What makes the explanation a causal one (satisfying the ontic constraint) is the fact that these properties and their changes in the collision can be mapped onto the dispositions of causal processes and their manifestations in a causal interaction. Note that a number of idealizations may be employed in the scientific
explanation, such as the absence of other forces, no loss of energy to heat, etc. There will be a trade-off between capturing an important regularity in the disposition/manifestation profile of processes in the causal network via the physical description (with wide application elsewhere), and accurately assessing an actual case involving particular causal processes. As discussed further in the next section, a more idealized description may make for a more powerful explanation via increased generality.

Other cases can be thought of similarly. Consider a scenario featuring two charged particles that carry a charge of the same sign. Assume at time $t_1$ these are stationary at positions $p_1$ and $q_1$, respectively. At a later time $t_2$ they have repelled each other and moved to new positions $p_2$ and $q_2$, and we seek a causal explanation of this new condition. On a first pass, these particles correspond to causal processes that at $t_1$ are transmitting particular dispositions to mutually repel each other. And as in the prior case, the manifestations we associate observationally with movement of the particles are interpreted as changes in the disposition profile of the processes represented by the particles: i.e. at $t_2$, each particle has an altered disposition to interact with other processes at various locations compared to before. One additional element is needed here, however. Given that our definitions specify interactions at particular space-time locations, this scenario needs to be described in terms of each processes’ interaction with an additional connecting process or processes. These may be represented in physics by a field (as in classical theory) or by an exchange of photons (as in quantum electrodynamics).\(^\text{123}\) Note that ascribing the physical property of charge to the particles will attribute a general pattern of dispositions toward such interactions. Charge in a simple classical example

\(^{123}\) As noted in Chapter 3, causal process theory is incompatible with true action-at-a-distance (with no mediating particles or fields).
would typically be inter-defined in conjunction with Coloumb’s law. The degree to which the theoretical description accurately depicts the results of an actual causal interaction will again depend on the nature of idealizing assumptions made in its definition, and the presence or absence of other causal influences.\textsuperscript{124}

To bring out these distinctions further, consider a physical scenario where two particles of like charge are in some proximity but no change is manifested due to the presence of another force. In classical mechanics, this situation may be described (using vector addition) in terms of a null resultant force: the force associated with charge offset by another force of equal strength. However, in the causal account of the case, the processes represented by the particles do not possess dispositions toward a causal interaction with each other—they are not disposition partners. Here, formulating an “explanation” for the static scenario using theoretical properties is incongruous with the underlying causal network: there is no productive causation at work. However, this deviation can be seen as compellingly warranted by the epistemic virtues that accompany the consistent use of broadly applicable theories.

3. The Role of Causal Properties

Given our familiarity with the idea of component forces, this last discussion seems counterintuitive. However, this is a reflection of the goals presented in section one.

\textsuperscript{124} The case of the electro-magnetic field raises the question of whether the notion of transmission still makes sense in light of theories which imply continuous interactions. The first option is to understand transmission as including the possibility of truly continuous interaction as a limiting case. However, to the extent this appears problematic, one can hold to the discrete model by interpreting continuous interactions as a theoretical idealization.
Physical explanations employ theoretical property terms (mass, etc.) that are intended to have wide, indeed universal, scope of application. As discussed above, the present framework posits that the elements of general explanations are to map onto a network of singular causal processes and their interactions. To clarify this relationship, it is important that the notion of a disposition employed in the definitions above be distinguished from that of a causal property. While causal properties play a role in some philosophical theories (discussed below), here they are not elements of the causal nexus. I will reserve “causal properties” as a label for the various theoretical terms scientists employ in causal explanations. For instance, when explanations feature terms such as mass, velocity or charge in a causal role, these are causal properties. What is the relationship between these theoretical causal properties and the causal nexus? In the present interpretation, these property terms in an explanation map onto a regularity in the dispositional profile associated with a causal process. Those causal processes that are taken to share certain properties are the target for labelling as theoretical entities or objects (e.g. electrons). Because causal properties are associated with regularities, they may be inter-defined with (causal) laws which also describe rules for interaction. But it is the causal nexus (made up of singular processes and interactions) that is the actual target which causal properties and laws seek to describe or represent.125

This treatment of the respective roles played by dispositions (a.k.a. “powers’) and causal properties differs from that offered by philosophers who explore these issues in the context of metaphysics and/or scientific realism. There, causal properties are part of the fundamental ontology, although the precise relationship between properties and powers

125 Cf. Cartwright (2009): “Causal laws describe what singular causal processes will or can occur under specified circumstances whenever the prescribed causes are instantiated (129).” A more detailed comparison with the work of Cartwright is in section 5 below.
is described in more than one way. The properties are either seen as ontologically prior to powers in some fashion or else properties are defined in terms of powers. Shoemaker (1980) first distinguishes between “powers...and the properties in virtue of which things have the powers they have (Shoemaker, 1980/2003, 211, emphasis added), but then fleshes this out in a more neutral way by asserting that “properties are clusters of conditional powers (Shoemaker, 1980/2003, 213, emphasis added).” Mumford (2004) endorses a neutral view, whereby “there is nothing more to a property than its powers and...the powers fix the identity of the property (Mumford, 2004, 171)”. Chakrabartty (2007) says that “a causal property is one that confers dispositions on the particulars that have it to behave in certain ways when in the presence or absence of other particulars with causal properties of their own...the property of mass, for instance, confers on bodies that have it certain dispositions to be accelerated under applied forces (Chakrabartty, 2007, 108, emphasis added).” He later acknowledges some ambiguity in this formulation between the idea that properties are distinct and prior to associated powers or else identical to them, but says either interpretation is acceptable (137-141). The important point, for Chakrabartty in particular, is that taking the unobservable causal properties found in our best physical theories to be ontologically basic serves to support his advocacy of a form of scientific realism. Note that on all of these views, it is still the case that observed phenomena are due to dispositions/powers and their manifestations in interactions between particulars.

While causal properties may well be part of the fundamental ontology of the world, and perhaps our best physical theories successfully pick them out, there is no

---

126 In a later paper he rejected this formulation in favor of the view that “it is essential to a property that it bestow a particular cluster of particular powers (Shoemaker 1998/2003, 412, emphasis added).”
reason to make this assumption for present purposes. By treating them as theoretical
terms whose goal is to track regularities in the causal structure of the world, we can
account for their role in causal explanations without taking a stance on their degree of
fidelity to truth. Consider the fact that the theoretical definitions of fundamental causal
properties (and associated laws) typically involve idealization. Chakravartty argues that
the pragmatic use of idealized properties and associated laws in practice has no bearing
on the question of the ontological reality of other, true, properties and laws. Further, in
the case of properties and laws which are idealized only in the sense of omitting the
presence of some causal influences, as in a ceteris paribus law statement, the causal
properties/laws may be picked out correctly, even if actual scenarios are unrealistic
(Chakravartty, 2007, 141-150). These claims won’t be disputed here, but go beyond
what needs to be assumed in our framework: the actual causal network serves as the target
of investigation and as such it provides an ontic constraint for the representations
employed in explanations. The degree to which the scientific descriptions employed are
true is a further topic of debate.

4. Composite Causal Processes

Salmon outlines a distinction between etiological explanations and constitutive
explanations. Etiological explanations trace the relevant preceding processes and
interactions leading up to a phenomenon. A constitutive explanation, on the other hand,
is one that cites the interactions and processes that compose the phenomenon:

127 Related to this point Chakravartty (2007) argues for a distinction between abstraction and idealization,
rather than viewing the former as a kind of idealization (pp. 221-224).
A constitutive explanation is thoroughly causal, but it does not explain particular facts or general regularities in terms of causal antecedents. The explanation shows, instead, that the fact-to-be-explained is constituted by underlying causal mechanisms. (Salmon, 1984, 270)\textsuperscript{128}

Salmon does not offer a detailed discussion of constitution. Rather, referencing a diagram (reproduced as Fig. 4.1), he shows how one can divide a causal network into etiological and constitutive elements:

\textbf{Fig. 4.1: Etiological and constitutive aspects of explanation}

If we want to show why $E$ occurred, we fill in the causally relevant processes and interactions that occupy the past light cone of $E$. This is the etiological aspect of

\textsuperscript{128} Of course, explanations may have both etiological and constitutive elements (see Salmon, 1984, 270-1).
our explanation; it exhibits $E$ as embedded in its causal nexus. If we want to show why $E$ manifests certain characteristics, we place inside the volume occupied by $E$ the internal causal mechanisms that account for $E$’s nature. This is the constitutive aspect of our explanation; it lays bare the causal structure of $E$. (Salmon, 1984, 275, emphasis original)

This passage doesn’t provide a recipe for marking off the volume that defines the constitutive aspect, however. It suggests two different ways of “zooming” in on the causal network, but not what would demarcate scales or levels.

In addition to addressing this question of scaling, the constitutive dimension of the theory should also provide the ontic resources for the variety of causal properties identified in scientific explanations. As discussed in Chapter 3, the conserved quantity version of process theory had limited resources in this regard. When the causal properties that characterize elementary processes also happen to scale up, as when we describe macroscopic bodies in terms of their mass or net charge, then we can continue to apply the framework. But the theory should also provide the basis for mapping other sorts of properties, such as those relevant to explanations in biological sciences. Salmon’s sketch above needs to be enhanced: we need to be able to distinguish causal processes and interactions at various scales, and show how they underpin a wider scope of theoretical properties and entities.

4.1 Composite Causal Processes – Definitions

To define a composite causal process, we will employ the idea that a coherent structure at the higher scale arises from a series of repeated interactions at a lower scale (Fig 4.2). We should pick out both higher level causal processes and interactions by
attending to patterns at the lower level.\textsuperscript{129} In Herbert Simon’s discussion of complex systems, he notes that complexity often “takes the form of hierarchy (Simon, 1962, 468)” and notes the role interactions play in this context:

In hierarchic systems we can distinguish between interactions among subsystems, on the one hand, and the interactions within subsystems—that is, among the parts of those subsystems—on the other. (Simon, 1996, p.197, emphasis original)

Simon describes hypothetical systems where there are no interactions between parts as “decomposable,” and then goes on to discuss “nearly decomposable systems, in which the interactions among the subsystems are weak but not negligible (Simon, 1996, 197, emphasis original).”\textsuperscript{130}

\textsuperscript{129} Andersen (2017) recently has proposed an alternative approach to identifying patterns in the causal nexus. She first assumes that the Salmon/Dowe conserved quantity theory describes the causal nexus at the microphysical level (and she criticizes the theory for lacking the resources to describe higher level causes). Then, she adds the notion of a pattern ontology: identification criteria can be used to pick out (a vast number of) possible patterns, any of which can serve as the basis for causal relata. Using a phase space representation of the nexus one can define counter-factually robust volumes to serve as the relata – Andersen sees a benefit in the potential use of information-theoretic tools to define causal relations between such volumes. The approach taken in the present theory differs in that the elementary level causal nexus has within it the resources to build a non-arbitrary hierarchy of composites that correspond to the entities investigated in the natural sciences.

\textsuperscript{130} This idea was first introduced in Simon & Ando (1961), which also presented modelling techniques for estimating the time scales over which the independence of the parts can be approximated.
Fig. 4.2: Two causal processes form a composite via repeated interaction

The present proposal will rely on the idea that differential interaction rates give rise to a hierarchy of processes. In physical descriptions this differential will typically be described in terms of the varying strength of forces, but in the causal nexus itself the raw material is interaction frequency. To complete the definition of a composite process requires just one additional element. The interacting sub-processes must constitute a cluster of dispositions that characterize how the composite process will itself enter into interactions.

Definitions

1. A composite causal process consists of two or more sub-processes (the constituting group) that interact with a greater frequency than each does with other processes, and that together transmit a composite cluster of dispositions between space-time regions.

2. A cluster of dispositions is transmitted between space-time regions A and B (A≠B) if it is compresent at A and is not significantly modified by interactions
with processes outside the constituting group during the interval between A and B.

3. A composite causal interaction is an intersection of composite causal processes at a space-time region where there is a mutual manifestation of dispositions, and where the associated (significant) modifications to the processes are transmitted beyond the locus of the intersection.

A few comments are appropriate regarding these definitions and their differences from the ones offered for elementary processes. First, the sub-processes may be elementary or themselves composite, allowing for the creation of a hierarchy. Also, unlike the definitions of elementary causal processes, these definitions imply a degree of indeterminacy and the existence of borderline cases. The “no interaction” criterion for transmission cannot be strict, hence the appeal to the notion of “significant” modification (this is the reason for using the subscript “C” in this context). It may be that to transmit dispositions, the differential between intra-process and extra-process interactions must be particularly high, or else there may need to for a particular pattern in the repeated interactions. Otherwise an external interaction might be disruptive. Using different language, we might say that composite processes are subject to perturbation by external processes, and their dispositions must be robust under this perturbation for transmission to occur.

On this point, it is important to stress that the dispositions characterizing a composite process are not constituted by a static collection of the dispositions characterizing its sub-processes. The dispositional profiles of the sub-processes in the constituting group are changing due to their frequent interactions. The dispositions of composite processes are constituted by the spatio-temporally extended pattern of
interactions and changing disposition profiles among its sub-processes. These composite-level dispositions are for interactions at a larger scale: an interaction involving two or more composite processes takes place in a space-time region that is large relative to the respective sub-processes and their interaction rates (Fig 4.3). As a result, any causal properties which map onto to the dispositions at different levels cannot be said to have an inter-level (“vertical”) relationship of either composition or supervenience at an arbitrary point in time.\footnote{This discussion of the constitution relationship may be compared to suggestive remarks offered by Martin (2004), p. 35. Also, an argument for dispositions based on their role in physical part/whole relationships is offered by Hüttemann (2013), section 2. Hüttemann goes on to offer an account of causation that has something in common with the present proposal in its combining of dispositions with a notion of process. However, the differences are substantial: Hüttemann’s key idea is that “a cause is an actual disturbing factor (antidote) to the default process that a system is disposed to display” (p. 116). This incorporates the difference-making concept of cause and also requires epistemic/pragmatic factors to define default processes.}

**Fig. 4.3: Interaction of composite processes**

Compared to elementary processes and their straightforward X-type interactions, the lifecycle of a composite process will have richer features. A composite process comes...
into being if the right pattern is instantiated. It may be divided or destroyed if modifications are not just significant but disruptive to its constituting processes.

4.2 Examples

In the physical sciences, the familiar hierarchy of entities includes atoms, molecules, and condensed matter. Each of these can map onto composite causal processes, given that their properties plausibly depend on repeated patterns of interaction among a constituting group of sub-systems. To highlight this distinguishing feature, Simon considers a boundary case: that of gases. Ideal gases, which assume interactions between molecules are negligible, are, for Simon, “decomposable” rather than “nearly decomposable” systems (Simon, 1997, 197). Similarly, an ideal gas doesn’t have a clearly defined constituting group: the molecules do not have a characteristic pattern of interacting with each other at any greater frequency than they do with the external system (the container). A non-ideal gas, with weak but non-negligible interactions between constituent molecules, may map onto a composite causal process.

In biology, the complexity of systems may sometimes defy the easy identification of boundaries based on the notion of a composite process. Also, as in the examples considered in section two above, the epistemic constraints on explanation may warrant deviations from the causal theory. In these cases the interests of scientists would result in positing theoretical entities that do not conform to a quantitative accounting of intra-process and inter-process interactions. In the case of a cell, we have a paradigm case where the correspondence with a composite causal process is relatively clear. In contrast, at the level of organisms and groups of organisms there are a number of difficult cases that have given rise to a rich debate over the criteria for biological individuality. Still, the causal account of constitution is a useful starting point:
The individuality of organisms involves a distinction between self and other—between inside and outside. This distinction is defined by characteristic causal relations. Parts of the same organism influence each other in ways that differ from the way that outside entities influence the organism’s parts. (Sober, 1993, 150)

A discussion of the way things influence each other, of course, goes beyond a mere accounting of interactions, and is an entry point where theoretical concerns will introduce a distortion away from the notion of the composite causal process. From a biological perspective, sub-processes and interactions related to reproductive capacities may, for example, receive disproportionate attention in defining individuality.132

It is also the case that constitutive explanations in science will rarely attempt to explain the entire entity in full. This would mean explaining the entire dispositional profile in terms of the dispositions and causal interactions of sub-processes. As discussed further below (sections 5.2 and 6), it is more common for a scientific explanation to target a single characteristic regularity in the dispositional profile of a composite process.

5. Comparisons and Objections

With the addition of the framework for composite processes, the theory offered in this chapter has the tools to meet the desiderata set out at the end of Chapter 2. To further

132 The biological individuality debate is reviewed by Clarke (2011), who in listing 13 approaches to the problem, notes that she deliberately omits a criterion that would define an individual “in terms of the nature of its parts, or the kind or amount of interaction that takes place among them (316).” This is because this approach, in her analysis, is too general to settle difficult cases. What is needed is more precise specification that gives a “theoretical justification for including some kind of interaction but not others (316).”
clarify the theory of the causal nexus and its relationship to causal explanations in the natural sciences, this section will compare aspects of the approach to the work of two philosophers with similar theoretical frameworks, and will also consider and respond to a number of likely objections.

5.1 Comparison to Cartwright

The discussion of physical examples in section two features an understanding of causation and its relationship to theoretical properties and laws that has affinities with Nancy Cartwright’s work. In Cartwright (1983), she argues that physical laws may achieve a high level of unifying scope and associated explanatory power, but this comes at a cost in terms of idealizing the actual phenomena (Cartwright, 1983, Ch. 3). This view fits well with the framework described above, where general explanations are motivated by epistemic rather than ontic normative constraints. She further argues (1989) that capacities are the actual target of scientific investigation. Compared to the present discussion, Cartwright sometimes describes capacities in a way that suggests what I call a dispositional profile, noting that “capacities, as I use the term, are not restricted to any single kind of manifestation” and “objects with a given capacity can behave very differently in different circumstances (Cartwright, 1999, 59).” On the other hand, while clearly not tied to particular manifestations, capacities seem to pick out one characteristic pattern of behavior. As a result, they would not typically connote the full cluster of dispositions associated with a causal process. In many of her discussions, where she uses gravity and charge as examples, capacities appear to more closely track what Shoemaker and Chakravartty would label causal properties.

For instance, in her (2009), Cartwright describes a three-fold distinction between the obtaining of a capacity, its exercise, and the “manifest (occurent) results” (p. 151).
Using gravity as an example, a massive object has the capacity to attract other massive bodies, the attraction is the exercise of the capacity, and the manifest result is the movement closer of the bodies. This last result, however, may be confounded by other factors. In other words the effects typically associated with a capacity may fail to occur because its manifestation takes the form of a causal influence or force, and not an actual interaction. For comparison, in the present framework only dispositions toward singular interactions are “exercised”, and in these cases the exercise of the disposition just is the (mutual) manifestation. This is a two-fold distinction.

Now, Cartwright’s treatment is consistent with her frequent emphasis on the general, wide-scope nature of a capacity, as when she (quoting Ryle) distinguishes the capacity for “grocing” from the characteristic specific things a grocer does, such as “weighing coffee, wrapping cheese, stocking shelves” (Cartwright, 2009, 153). The question is whether it is necessary to view a capacity as a “higher-level” disposition itself, with its manifestation taking the form of a causal influence, such as gravitational force. Corry (2009), in an article that in part responds to this Cartwright discussion, notes that such a view of capacities would seem inconsistent with Cartwright’s frequent dismissal of the reality of component forces in her discussions of physical laws.\footnote{See, e.g. Cartwright (1983), pp. 59-71. For Corry’s discussion, see Corry (2009), pp. 179-184. He goes on to argue that scientific analysis, of the sort that has led to the understanding of a broad range of phenomena in terms of the behaviors of components, implicitly assumes the 3-fold distinction, and this gives us good reason to believe in the existence of such capacities.} In the present framework, causal processes possess a rich cluster of dispositions (a dispositional profile), and the scientific ascription of gravitational attraction corresponds to a pattern of regularities in this profile. The question then can be reframed: does one need to postulate a “higher level” capacity/disposition to support the existence of these regularities? If this is what Cartwright is advocating, then the response is the same as that regarding the
adoption of an ontology of causal properties discussed in section three: they may exist, but it is unnecessary to assume this for our purposes. The dispositional “action” can remain at the level of underwriting singular causation.\textsuperscript{134}

5.2 Comparison with Wimsatt

The picture of a hierarchy of composite causal processes outlined in section four invites comparison to philosophical accounts of complex systems and their analysis. Wimsatt, for instance, examines the relationship between an interaction-based approach to analysis and alternative ways to decompose systems, such as assessment of spatio-temporal boundaries. First, he notes that decomposition of a complex system into parts can typically be performed in more than one way depending on a theoretical perspective, and these may have different implications for assessing spatial boundaries—this gives rise to “descriptive complexity” (2007, 181-2). Then, drawing from Simon and others, Wimsatt discusses the idea of decomposition based on interactions, i.e., breaking down a system into subsystems based on the relative strength of intra vs extra-subsystem interactions. The degree to which this decomposition maps onto the theoretical decompositions gives rise to “interactional complexity” (184-6).

In the case of the causal processes defined above, composites built directly from elementary processes should feature a fairly close match between the interaction-based relations among sub-processes and their spatio-temporal proximity. Then, as scales increase and phenomena get more complicated, this linkage can be expected to weaken given variation in the temporal extension of processes between significant interactions. While the introduction of the notion of theoretical perspectives introduces an epistemic

\textsuperscript{134} Below (section 6), I will make use of the term “capacity” in its sense as a collection of dispositions of some shared character (or, equivalently, an aspect of the dispositional profile of a causal process).
component to Wimsatt’s template, his focus on how spatio-temporal and interaction-based analyses can differ is not inconsistent with the hierarchical picture implied by the present view: without doubt the causal nexus can come to feature less direct linkages between interaction rates and spatio-temporal organization, giving rise to complex forms.

In a related discussion, Wimsatt makes it clear that he views causation as the ultimate basis for the various kinds of complex systems:

> Ontologically, one could take the primary working matter of the world to be causal relationships, which are connected to one another in a variety of ways—and together make up patterns of causal networks…Under some conditions, these networks are organized into larger patterns that comprise levels of organization, and under somewhat different conditions they yield the kinds of systematic slices across which I have called perspectives. Under some conditions they are so richly connected that neither perspectives nor levels seem to capture their organization, and for this condition, I have coined the term causal thickets. (Wimsatt, 2007, 200, emphasis original)

Wimsatt explains that levels of organization are “compositional levels”, characterized by hierarchical part-whole relations (201). They are “constituted by families of entities usually of comparable size and dynamical properties, which characteristically interact primarily with one another, and which, taken together, give an apparent rough closure over a range of phenomena and regularities” (204, emphasis original). Wimsatt’s notion of composition includes not just the idea of parts, but of parts engaged in certain patterns of causal interactions, consistent with the approach to composite causal processes described above.
Wimsatt discusses at some length how notions of reduction and emergence should be understood given this understanding of composition. He offers a definition of reductive explanation that shows a similarity to the causal process view of constitutive explanation:

A reductive explanation of a behavior or a property of a system is one that shows it to be mechanistically explicable in terms of the properties of and interactions among the parts of the system. (Wimsatt, 207, 275)

The ability to provide such an understanding is consistent with a form of “emergence,” in the sense that the properties of the whole are intuitively “more than the sum of its parts (276-7).” The key idea here, again, is that composition includes the interactions between the parts. For comparison, Wimsatt introduces the notion of an aggregate, where the properties of the whole are “mere” aggregates of the properties of its parts. For this to happen, “the system property would have to depend on the parts’ properties in a very strongly atomistic manner, under all physically possible decompositions.” (279-280). He analyzes this condition in detail and concludes it is nearly never met (absent idealization) outside of the conserved quantities in physical theories. In the context of causal process theories, the fact that the conserved quantities are such a small subset of the properties employed in natural science explanations prompts the need for more adequate resources for constitutive explanation (compared to the work of Dowe and the later Salmon). The present proposal can fulfill this need: causal properties map onto dispositions of composite processes, and these do not depend atomistically on lower level dispositions, but are built from a richer base of changing dispositions/interaction patterns among constituents.
5.3 Objections

The theory of the causal nexus presented so far touches on a wide variety of debates, and raises many additional questions and potential objections. Here I focus on concerns about processes and interactions as defined above.

A familiar objection is the complaint that invoking a disposition to explain a manifest phenomenon is explanatorily empty, as in Molière’s quip about citing opium’s “virtus dormitiva” to explain why it induces sleep.135 But in using the present framework one can accept the conclusion that citing a single disposition for a particular interaction has little explanatory power. Most effective etiological explanations will cite a larger number of causal antecedents that work in concert. Also, as discussed, another tool for understanding comes in the form of constitutive explanation: explaining a phenomenon by citing the processes and interactions that compose it (e.g. an account of the properties of opium in terms of its chemical constituents). Finally, it must be recalled that while the causal nexus provides the ontic constraint, the final explanatory product will include features that increase understanding but depart from a precise mapping of the causal nexus. In particular, greater explanatory power will often be pursued by discovering and describing patterns and regularities in the network (one might describe the scope of the soporific effect of opium across types of organisms).

Another concern that arises in discussions of dispositions (or “powers”) has to do with their ontological status. In the analytic tradition, many philosophers have thought that dispositional ascriptions were reducible to more empiricist-friendly occurrent properties. A thorough debate on this question, however, has shown that no

135 From the play Le malade imaginaire, which premiered in 1673 (see Molière, 1985, p. 478).
straightforward reduction is forthcoming.\textsuperscript{136} Even if dispositions are accepted into a theory, however, there is another active debate regarding whether they need to be grounded in or related in some way to other, non-dispositional (categorical or qualitative) properties.\textsuperscript{137} The definitions above make no reference to such properties: manifestations take the form of new or changed clusters of dispositions. The only other sort of property whose existence might be implied by the definitions is spatiotemporal in nature. An ontology that only features dispositions has been viewed skeptically: Armstrong notes that “causality becomes a mere passing around of powers from particulars to further particulars (Armstrong, 2005, 314).” I think the sole reliance on dispositions can be defended, but there is no reason to do so for present purposes.\textsuperscript{138} As in the discussion of causal properties, the theory can fulfill its function without denying that a more complete metaphysical picture will entail additional features. Philosophers who hold various positions in metaphysical debates about the nature of properties and particulars should be able to accept the elements of the causal nexus describe above as a limited provisional ontology in the service of causal explanation.\textsuperscript{139}

Another controversial element in the causal process approach is the notion of propagation. The processes themselves create connections in the network via their

\textsuperscript{136} For a concise review, see Cross (2012).
\textsuperscript{137} C.B. Martin’s approach was to propose that properties were simultaneously dispositional and qualitative: see Martin (2008), pp. 63-69. The position was also been adopted by Heil (2003) and G. Strawson (2008) among others.
\textsuperscript{138} For a defense, see Mumford (2009). Mumford considers the objection that since powers/dispositions are potencies, there also needs to be a distinct concrete notion of “act” in the world. The response is that it is open to the pan-dispositionalist to argue that powers are not mere possibilities, but are concretely real even if unactualized. The objector may also then argue that if powers manifest in terms of other powers this leads to a regress. Here the response is that there may be a holistic mutual dependence, but not a vicious regress (see Mumford, 2009, 98-102). For a collection of articles exploring the debate over pan-dispositionalism, see Marmodoro (2010).
\textsuperscript{139} Such compatibility may be lacking if one adopts a structuralist metaphysics that sees laws and related symmetries as fundamental rather than objects and properties. See French (2014) for a recent defense of such an approach.
temporal extension. The present approach deploys Salmon’s notion of transmission: the upshot of this proposal (the “at-at” theory of propagation) is that temporal extension is basic to a causal process in the absence of interactions. It therefore makes no sense to talk of arbitrarily dividing the process and asking how its temporal parts are connected.

Chakravartty discusses in some detail the concern that realist theories of causation may encounter trouble if time is assumed to be continuous and infinitely divisible. In a theory where causation involves a connection between contiguous events, the notion of contiguity itself becomes problematic: if time is continuous and events can be arbitrarily brief, there is always room for additional intervening events between two relata. Also, another difficulty presents itself if one assumes events are temporally extended: how does the early part of an event connect to a later part (if not causally)? In part due to these concerns, Chakravartty offers a version of a process theory featuring properties which confer continuously manifesting dispositions: objects with such properties “are thus in a continuous state of causal interaction (Chakravartty, 2007, 108).” The “events” that are the relata of causal ascriptions are then taken to be aspects of a continuous causal flux picked out for pragmatic or epistemic reasons.

This approach seems inferior for several reasons. First, it is problematic to collapse the distinction between dispositions and manifestations in this way: a

---

140 Salmon (1984), p. 147.
141 Cf. Ehring (1997), who offers an account of causation in which the persistence of tropes plays the central role.
142 See discussion in Chakravartty (2007), Ch. 4.
143 Mumford & Anjum (2011) also offer an account of causal processes without the discrete propagation/transmission feature found in Salmon/Dowe and the present account. In their picture, dispositions mutually manifest in a temporally extended way, giving rise to new dispositions and potentially new processes (see Mumford & Anjum, 2011, Ch. 5).
continuously manifesting disposition doesn’t appear to be a disposition at all. A more important objection concerns the relegation of the distinction between stability and change to merely epistemic or interest-relative considerations. Theorizing about the distinction in this way leads to a highly revisionary picture of nature. In contrast, the present approach provides the ontic resources for both etiological and constitutive explanations of phenomena: stability and change are interwoven, but the distinction has a basis in the characteristics of elementary causal processes and their interactions. Finally, the idea that there is an irreducibly discrete aspect to nature at the fundamental level should be more acceptable given the advent of quantum theory.

It should be noted that Dowe also criticized the idea of transmission (Dowe, 2000, 115-122). Importantly, he questioned whether the formulation really avoids controversial issues of object identity. One can ask how exactly a causal process can be picked out as the same one on either side of an interaction. In other words, how substantially can a dispositional profile of one process be altered? It isn’t clear that this issue can be addressed without controversial metaphysical commitments, but doing so does not seem necessary for the theory to play its role: pragmatic and epistemic factors will govern the treatment of boundary cases.

On another point, Dowe questioned an implicit assumption in Salmon’s depiction of transmission: that causal processes are spatio-temporally continuous entities. In the last formulation of his conserved-quantity account Salmon tied the idea of transmission between points “A” and “B” to the possession of a quantity at every point in the space-time interval between A and B. It isn’t clear whether such an assumption can hold up in

---

144 Such a disposition seems indistinguishable from a categorical property (cf. Hüttemann, 2013, p. 107).
145 Dowe’s own account adopts a notion of object identity, but as discussed in Chapter 3, this ultimately involves taking object persistence (across changes due to interactions) as basic.
light of quantum theory. The present formulation may need to be interpreted in a less precise way regarding what it means for dispositions to be compresent “during the interval between A and B”: this will be revisited below in the context of quantum theory (Chapter 5).

6. Mechanistic Explanation

Scientists offering explanations of phenomena often describe mechanisms, and several philosophical accounts of mechanistic explanation have been offered in recent years. In this section, I argue that the approach to causal explanation provided in this chapter is superior to these recent philosophical approaches to mechanistic explanation. This is for two related reasons. The present approach provides an underlying causal theory that these accounts fail to offer. Second, by thereby providing a clear account of the ontic constraint, it clarifies how mechanistic explanations meet standards for good explanation.

6.1 Mechanisms and Productive Causation

The idea of a mechanism builds on the concept of productive causation (“bringing about change via a connection”). Rather than a single instance of productive change, a mechanism traces multiple related connections and changes. The causal process theory introduced in this chapter thus provides the basis for an account of mechanisms. And, indeed Salmon described his earlier framework as a lying within the tradition of “the
mechanical philosophy.”146 There have been several more recent accounts of mechanistic explanation, often referred to as “new” mechanistic approaches. One important way in which they are different from prior mechanical theories is in their focus on the contemporary use of mechanistic explanation in biology. Another distinctive move in these accounts is to divorce the idea of mechanism from a basis in a theory of causation. This innovation, however, is a step backward in pursuit of the larger question of how natural science explanations increase understanding.

In their 2000 article, Machamer, Darden and Craver (MDC) provide the following definition: “Mechanisms are entities and activities organized such that they are productive of regular changes from start or set-up to finish or termination conditions (MDC, 2000, 3).” They add that “to give a description of a mechanism for a phenomenon is to explain that phenomenon, i.e., to explain how it was produced (MDC, 2000, 3).” The organization of entities and activities involves the specification of their spatiotemporal features (orientations of entities/durations of activities) and relationships. With regard to the role of the concept of causation, the word “produce” is used repeatedly in the paper’s account of mechanism, and the authors particularly highlight “activities” as the “producers of change (MDC, 2000, 3).” Despite the centrality of productive causation in their discussion, they do not explicate the concept. Citing Anscombe, the authors argue that “cause” is a generic term that “only becomes meaningful” when a more specific activity is specified.147 In a subsequent article, Machamer confirms that “one does not need a theory of cause (Machamer, 2004, 27).” MDC critique the Salmon and Dowe

147 See Anscombe (1971). Also, Bogen (2008) examines this approach in the context of mechanistic explanation.
process theories as an account of productive causation by asserting that their approach “does not exhaust what these scientists know about productive activities (MDC, 2000, 7).”

So what are these various activities? The authors appeal to examples where scientists employ their own theoretical terms to describe mechanisms. For instance, in the description of chemical transmission at synapses, activities “include biosynthesis, transport, depolarization, insertion, storage, recycling, priming, diffusion, and modulation (MDC, 2000, 8).” It is clear that the mechanistic explanations deployed by scientists will use a variety of terms to describe both entities and their interactions, but using the framework defended in section one of this chapter we can see that such an explanation is seeking to meet both an ontic constraint and an epistemic one. The ontic constraint reflects the underlying productive causal structure giving rise to the phenomenon, while the epistemic constraint mandates strategies that seek to achieve an explanatory power tailored to achieving understanding in a particular (scientific) audience. This involves decisions regarding scope, grain, and the use of idealization, as well as the employment of an existing theoretical apparatus and/or other theoretical terms that are tailored to the audience.

MDC’s approach, then, mixes a philosophical gloss on the idea of mechanism (the high-level framework of entities and their properties, activities, etc.) with deference to science to supply the details. This hybrid approach falls short because it obscures the fact that there is one underlying productive causal nexus in the natural world. The relationship between the concept of productive causation and the causal elements in scientific representations of mechanisms is not one of generic/specific levels. Rather, the scientifically defined terms are to be mapped onto the elements of a unified theory of productive causation in the way described earlier in this chapter. In this way, we can
understand the relationship of each mechanistic explanation to the ontic constraint that governs good causal explanations.

6.2 Mapping Mechanisms

The approach described in this chapter can effectively subsume the frameworks of the new mechanists. To begin, I will show how the features of the MDC (2000) account correspond to elements in the causal nexus. The first step is clear: theoretical entities and properties described by scientists will map onto (composite) causal processes and their dispositions toward interactions (to be more precise, property terms describe regularities in the dispositional profiles of processes). There is no reason to think entities and properties play any different role in the mechanistic framework than that discussed above in the general context of causal explanation. But what about “activities”? For MDC:

Activities are producers of change. Entities are the things that engage in activities. Activities usually require that entities have specific types of properties. The neurotransmitter and receptor, two entities, bind, an activity, by virtue of their structural properties and charge distributions. (MDC, 2000, 3)

At first glance, what the authors call an activity seems to straightforwardly pick out a particular kind of causal interaction. MDC’s preference for “activities” rather than “interactions” derives from a view that the latter implies a picture that is “impoverished,” failing to do justice to the richness of the productive changes various entities are engaged in. According to MDC, the notion of interaction “emphasizes spatio-temporal intersections and changes in properties without characterizing the productivity by which those changes are effected at those intersections (5).” In the present account, however, we can see that interactions between composite processes, while involving the mutual
modifications of the dispositional profiles of the processes involved, can also be analyzed further in all their complexity if one moves to the level of the various sub-processes at work.¹⁴⁸

The authors’ choice is also motivated by a disagreement with Glennan’s (1996) account of mechanism which does feature “parts” and their “interactions;” this is because Glennan’s notion of interaction ultimately relies on a problematic invoking of “causal laws” (see MDC, 2000, 4). The authors express a worry that “reliance on laws and interactions seems to us to leave out the productive nature of activities (4).”¹⁴⁹ The present account, of course, provides a theory of production in terms of the mutual manifestation of dispositions. This provides the essential link between the idea of interaction and the underlying causal concept of production.

Overall, if we look at an etiological version of mechanistic explanation, where the explanandum is a final result or “termination conditions”, there is a straightforward correspondence of the elements of mechanisms (as described in the MDC framework) with the features of the causal nexus. Entities, properties, and activities map onto causal processes, their dispositional profiles, and their interactions. For a sequence to comprise a mechanism in the MDC sense, one more criterion is added, which is a regularity

¹⁴⁸ While MDC (2000) also discuss hierarchies in mechanisms, their approach differs in that the entities and activities involved, including those at the “bottoming-out” level of description, will again be specified by the scientist/scientific field with no attempt to connect to a hierarchy in a causal theory (pp. 13-14).
¹⁴⁹ In subsequent work, Glennan revised his views. First, he replaced reliance on laws with a requirement that interactions be characterized by Woodward-style “direct, invariant, change-relating generalizations (Glennan, 2002, S344).” More recently, Glennan (2017) discusses several options for theorizing about fundamental-level mechanisms, including employment of laws but also the possibility of appealing to causal powers. But he also questions the idea that one needs to posit any fundamental level. He says the idea of a mechanism is classical, and quantum phenomena of superposition and entanglement make the idea of fundamental-level mechanisms problematic. His final view is that mechanisms emerge (docohere) from the quantum realm in a local, piecemeal fashion and that this supports the pluralistic conception of compound mechanisms and their activities (see Glennan, 2017, pp. 184-193).
requirement: “Mechanisms are regular in that they work always or for the most part in the same way under the same conditions (MDC, 2000, 3).” Here, we see another way that epistemic constraints concerning generality enter into accounts of mechanistic explanation: in addition to using similarities across cases to underwrite the type-identification of theoretical entities, properties and activities, an additional requirement of reliable (temporal) repeatability is needed for an etiological sequence to comprise a “mechanism” as defined.

The case of constitutive explanation presents a subtler challenge in assessing the relationship between the present approach and the “new” mechanistic frameworks. Here, rather than explaining an end result, a mechanism is described as constituting a “phenomenon”. The approach to constitutive explanation described in section 4 would seem to serve as a basis for explaining how an entity and its properties come to be (corresponding to a composite causal process and its dispositional profile). But what does it mean to constitute a phenomenon?

Craver (2007) offers a more detailed treatment of constitutive mechanistic explanation. He begins with a description where the explanandum phenomenon is understood as a “property or behavior explained by the mechanism”, and where the explanans consists of the organized component entities and activities that comprise the mechanism (6-7). Later in the book, he elaborates on this picture (using the example of the action potential), drawing on Robert Cummins’ account of functional analysis as a foil. In Cummins’ work, the explanandum is a capacity or disposition of a system, and

150 See Andersen (2012) for a recent examination of the regularity criterion.
151 “Mechanisms are sought to explain how a phenomenon comes about or how some significant process works (MDC, 2000, 2).”
152 See, e.g. Cummins (1975) and (1983).
it is explained by analyzing it in terms of the organized exercise of the system’s sub-capacities (110). Craver explains that in the new mechanistic framework these sub-capacities would be characterized by component entities and their activities along with the way these are spatially and temporally organized (122-138).153

The comparison to Cummins helps to illuminate how Craver’s account compares to the present framework. Rather than taking an entity as the explanandum (e.g. a neuron), where the appropriate analogue would naturally be a (whole) composite causal process, we are only interested in a characteristic capacity associated with the entity (e.g. the capacity to produce an action potential). While Craver himself continues to prefer describing the explanandum phenomenon as a “behavior” (139), taking it to be a capacity for a behavior allows for a clearer connection with the ontic causal nexus. Such a capacity would typically map onto a disposition or, more likely, a sub-set of the dispositional profile of a composite process.154 Rather than the explanans featuring the component entities and activities that map onto all of the sub-processes and their interactions which constitute a composite process (perhaps pruned for epistemic/pragmatic reasons), it would include only a subset that is relevant to the case.

But how is the subset to be identified? In the case of the complex biological phenomena of interest, Craver argues that neither a spatial nor an interactionist criterion

153 In Cummin’s account, the sub-capacities could be the capacities of system components, but he also allowed for analysis into sub-capacities which did not correspond to those of component parts. See Cummins (1983), pp. 28-9.
154 Craver (2007) explains that an explanandum phenomenon may be complex and “multi-faceted (125, emphasis original).” Also “the explanation should encompass “multiple features of a phenomenon, including its precipitating conditions, manifestations, inhibiting conditions, modulating conditions, and non-standard conditions (139).” In the present interpretation, this means the explanation should account for a collection of related dispositions toward mutual manifestations.
for decomposing systems will do (discussing the approaches of Simon and Wimsatt among others—Craver, 2007, p. 141-4):

The spatial and interactive boundaries of mechanisms depend on the epistemologically prior delineation of relevance boundaries. Spatial boundaries are those that circumscribe all the relevant entities and activities. Temporal boundaries are those that include all the relevant activities. An account of constitutive mechanistic explanation must include an account of constitutive relevance. (Craver, 2007, 144, emphasis original)

It makes sense that given the nature of the explanandum, the relevant features included in an explanans may map onto a collection of sub-processes and interactions that differ in boundary conditions from those used to identify the constituents of a whole composite process. In line with Wimsatt’s discussion (in his account of levels and perspectives discussed in section five above), the case of mechanistic constitution involves theoretical considerations (and associated epistemic norms) that may compel the formation of explanations whose elements fail to neatly map onto the hierarchical backdrop of interacting composite causal processes.

In more recent literature, Ylikoski (2013) undertakes an examination of constitutive explanation, and concludes that causal capacities are their targets, in broad affinity with the conclusions reached above. Kaiser & Krickel (2016), however, reach a different conclusion. They pose two objections to taking dispositions or capacities to be the targets of constitutive mechanistic explanation. First, they say that “explanations of

---

155 Inspired by consideration of the kinds of experimental techniques used to evaluate compositional relationships in biological phenomena, Craver (2007) goes on to offer a manipulation/intervention-based account of constitutive relevance in the spirit of Woodward’s causal theory (pp. 152-160).

156 Bechtel (2015) emphasizes the epistemic basis of mechanism boundaries.
dispositions refer again to dispositions (not to their manifestations”), but mechanistic explanations as usually understood refer to “objects and non-dispositional occurrences (Kaiser & Krickel, 2016, 19, emphasis original).” On the present theory, this objection has no bite. It is agreed that explanations (of dispositions/capacities) will refer to objects, but these objects will in turn correspond to processes that also feature dispositions (some of which will be manifesting in the relevant context). In addition, the explanans will also feature interactions (mutual manifestations) among objects/sub-processes—these correspond to the authors’ “non-dispositional occurrences”. Kaiser & Krickel’s second objection is that examples from the sciences do not typically target dispositions, but rather “processual manifestations” of dispositions:

Consider the mechanism of muscle contraction. The phenomenon that biologists want to explain by citing this mechanism is not the disposition of a muscle fibre to contract when a neuronal stimulus occurs. Rather, they want to explain the entire process of how a muscle fibre contracts. (Kaiser & Krickel, 2016, 20, emphasis original)

The authors go on to propose that rather than dispositions or capacities, the best way to describe the targets of constitutive mechanistic explanation is in terms of “object-involving occurrences” (23-31). The authors’ description of how these targets are often described is unobjectionable, and it is in the spirit of Craver’s including “behaviors” in the target. However, the conflict here can be readily diagnosed. In these cases where the description of the target includes reference to the interaction events involved in

---

157 Here, the authors use “objects” and “occurrences” in place of “entities” and “activities”, respectively. They argue that these terms are more precise (see Kaiser & Krickel, 2016, pp. 10-11). The distinction is not crucial for the present discussion.

158 The authors also criticize using the notion of behavior as the explanandum for constitutive mechanistic explanation (see Kaiser & Krickel, 2016, pp. 20-23).
manifesting a capacity (rather than just the capacity itself), then on the present view the explanation has both a constitutive and an etiological dimension. The manifestations are explained by bringing in some reference to other objects involved, the i.e. disposition partners that are involved in mutual manifestation of the capacity. This extended analysis comprises the etiological aspect. While the authors may be right that the targets of “constitutive mechanistic explanation” are often or typically framed in this way, this convention of terminology in the mechanistic literature should not be used to characterize the nature of constitutive explanation itself, which has broader application.

To summarize, the work of “new mechanists” in explicating both etiological and constitutive mechanistic explanation can be better understood in the light of the present theory of productive causation and its role in explanation. New mechanists typically see their approach as competing with and superior to the causal nexus approach due to Salmon. But rather than being in competition, the account presented in this chapter provides the unifying causal backdrop for the natural science explanations that are the focus of the new mechanists’ work. The theory of the causal nexus provides the proper ontic constraint for explanation while the particular way entities and interactions (or activities) are described by scientists will also reflect the priorities that derive from the epistemic constraint. This account effectively subsumes the “new” mechanistic framework within a superior one.

6.3 Other Critiques of “New Mechanism” Accounts

Several points made by other critics of the new mechanist approach to explanation are worth noting, as they shed additional light on the differences with the

---

159 See Glennan (2010), pp. 253-258.
present approach. L. R. Franklin-Hall (2016) criticizes new mechanists on several fronts, although the common thread is that they fail to provide an adequate account of those “constraints capable of distinguishing the good mechanistic models—those that provide adequate explanations—from those that fall explanatorily short (44).” Franklin-Hall discusses three kinds of errors a mechanistic model may embody and argues the new mechanist frameworks don’t offer standards to reliably identify them. The first is a causation error, and the problem is that the frameworks lack a theoretical causal standard:

The central feature of this account of causation—the activity—is, from a philosophical perspective, brute. Scientists identify activities, but they have nothing generally in common: short of listing those taken seriously by scientists at a given time, we can’t say anything about what they are (Franklin-Hall, 2016, 53)

Franklin-Hall’s discussion of standards does not employ the scheme of ontic and epistemic constraints adopted in section one. But the need she identifies here is for a causal theory to provide the ontic constraint on the explanation. The second error is a “carving” error (46). This has to do with identifying entities that serve as components or parts of a mechanism. Here again, while several suggestions have been put forward, Franklin-Hall argues that the new mechanist frameworks lack an adequate standard to identify good explanatory practice. Here, I would argue that the notion of a composite causal process can help to distinguish when an entity cited in a scientific explanation meets the ontic constraint, although epistemic challenges and pragmatic considerations will enter into distinguishing entities from etiological sequences in a mechanism. The notion of the epistemic constraint arguably plays an even greater role in Franklin-Hall’s third category (which is related to carving): this involves standards for employing the
proper level of analysis—or the avoiding of “zooming” error (47). Here the idea is that the mechanistic explanation should take place at a level that is just below the target phenomenon, so as to provide the “right” amount of explanatory detail. Identifying the right level is difficult, even if one does have a basic theory of hierarchy such as that provided by the notion of composite causal processes. But the challenge is much greater in the absence of such a theory of the ontic constraint. In summary, the present framework goes a long way toward supplying the kind of explanatory standard that Franklin-Hall finds lacking in the new mechanist accounts.

To conclude, I want to consider a different sort of critique. John Dupré (2013) argues that the mechanistic approach to understanding biological phenomena is inadequate. Among many concerns, he has a central complaint:

All these accounts start with an inventory of entities…But the entities that form the hierarchy of biological ontology are not stable. They are, rather, stabilized over a wide variety of timescales, and the processes of stabilization are a fundamental part of the explanation of the activities of living systems. Living things are the explananda in biological sciences as least as much as they are the explanantia. What are stable and robust in biology are not things, but processes. (Dupré, 2013, 30)

Dupré thinks it is more appropriate to describe biology in terms of processes, not things, and describes a hierarchy where the stabilization of a higher-level process depends on lower-level processes. He complains that there is no good ontological theory of processes,
and has pursued developing such a theory in recent years.\textsuperscript{160} But I suggest that the present approach to causal process theory (featuring the notion of composite causal processes) provides the resources needed to address this concern. The recognition of the role of internal processes in sustaining a higher-level process is incorporated, as is the fact that the constitution and maintenance of the higher-level system is relative to a time-scale.\textsuperscript{161}

Dupré is concerned that “the reification of stages of processes can have serious, sometimes epistemically harmful consequences (Dupré, 2013, 32).” Here it is the epistemic norms on explanation that must come into play. Good explanations will presumably sometimes involve mapping static notions of objects/entities onto temporally-extended causal processes to aid understanding, but in doing so the practitioner needs to respect the ontic character of the target system, which is the processual causal nexus. To the extent Dupré advocates a thorough-going process ontology, then the present account may not be precisely what he has in mind, but when it comes to providing a framework for biological explanation, it provides a way to meet his critique.

\textsuperscript{160} Dupré mentions the earlier Salmon/Dowe process theories, but makes the familiar complaint that “these are firmly grounded in problems in physics, and in ways that greatly limit their applicability to issues in the life sciences (Dupré, 2013, 35).”

\textsuperscript{161} Dupré includes an insightful point regarding how temporal extension is basic to biological processes, and how this isn’t accommodated well by metaphysicians who assume objects or their temporal stages can be divided into infinitesimally small slices (see Dupré, 2013, p. 31).
Chapter 5: The Compatibility of Causal Process Theory and Quantum Mechanical Explanation

Wesley Salmon worried that his causal process theory might be inconsistent with quantum mechanics (QM). His main concern centered on whether the references to space-time locations in his definitions were in tension with the non-local character of quantum phenomena.\textsuperscript{162} The potential problem is actually broader than this. The advocate for using a causal process account to underpin a conception of scientific explanation must generally address whether and how quantum physical descriptions map onto processes and their interactions. I argue below that the modified theory of processes presented in Chapter 4 is well suited to meet this challenge. First, the use of dispositions to characterize processes creates a more plausible target for mapping the properties of quantum systems. There is no requirement that these dispositions be localized in space (or space-time) between interactions. In addition, causal interactions (mutual manifestations) are well-suited to correspond to quantum measurements. Finally, the structure of the causal theory aligns well with at least some interpretations of QM. On the other hand, some features of QM prompt a need to clarify the corresponding elements in the causal theory, particularly when it comes to accommodating irreducibly probabilistic measurement outcomes and the distinctive phenomenon of entanglement. While the examination of these issues complicates the case, I argue it need not undermine the applicability of the process account in a quantum context.

\textsuperscript{162} In some formulations of the notion of transmission, Salmon specifically required that a mark (or later, a conserved quantity) must be present at each space-time point in a space-time interval. For reference, two versions of Salmon’s definitions are provided in the appendix.
Section 1 introduces the potential pitfalls in using the causal process approach in the context of QM. The discussion begins with Salmon’s concerns, and then, following a brief outline of (non-relativistic) QM, summarizes the challenge facing the process theorist. Section 2 offers reasons why the revised framework presented in Chapter 4 above is better positioned to meet this challenge than was Salmon’s theory. Section 3 discusses the approach’s compatibility with the orthodox (Von Neumann) interpretation of QM and examines issues surrounding measurement and entangling interactions. Section 4 is a brief look at the compatibility of the causal theory with some other interpretations of QM. While an in-depth survey is beyond the scope of this chapter, I discuss how the conceptual/ontological aspects of process theory fit into this landscape. Finally, brief concluding remarks are offered in section 5.

1. Quantum Worries for the Process Theorist

Salmon was motivated to develop a causal theory in order to provide a foundation for scientific explanations. Any incompatibility between the conceptions developed for the theory and those employed by scientists would be very problematic. This would be especially true for the physical sciences, where using an ontic explanatory framework which maps phenomena onto a network of causal processes and interactions would seem prima facie suitable.

Relativity theory, in fact, played an important role in inspiring Salmon’s causal process approach. The space-time interval between events is an invariant quantity in special relativity, and this is typically seen as giving a good reason to view events and their
space-time relations as the theory’s fundamental physical quantities. However, there is another way to interpret special relativity:

At any point in space-time, we can construct the Minkowski light cone—a two-sheeted cone whose surface is generated by the paths of all possible light pulses that converge upon the point (past light cone) and the paths of all possible light pulses that could be emitted from the point (future light cone). (Salmon, 1984, 140)

Light pulses are causal processes in Salmon’s sense. And so there is a basis for thinking about special relativity with processes as fundamental elements (and this conception might be extended to general relativity163).

Material objects, in the classical picture, seem to pose no conceptual problems in being interpreted as causal processes in Salmon’s sense. They follow well-defined space-time paths and enter into interaction events with each other.164 However, our foundational physical theory is quantum mechanics. Does it make sense to represent quantum physical phenomena in terms of causal processes and their interactions? In an extended passage of his 1984 book, Salmon considers whether his account can be utilized in the quantum context. Salmon thinks that the apparently indeterministic character of quantum mechanical phenomena is not an obstacle (at least, Salmon intends his theory to

---

163 In fact, from a historical perspective, the advent of general relativity removed what was an apparent obstacle for broadly “mechanical” causal theories like Salmon’s. The “action-at-a-distance” that characterizes the Newtonian theory of gravity appears to be incompatible with the approach (see Salmon, 1984, pp. 209-210, 242).
164 Actually, as noted in Chapter 3, Salmon did face a concern regarding how his theory applied to macroscopic objects. For Salmon, a causal process transmits its structure, or remains uniform with regard to a characteristic, in the absence of interactions. The proviso regarding the “absence of interactions” raises questions, since macroscopic examples of processes (like baseballs) are undergoing incessant interactions of one kind or another (e.g. interactions with air molecules). This issue is addressed in the present approach by making a distinction between elementary and composite processes.
encompass indeterministic cases—see discussion in section 2 below). Rather, it is his theory’s assumption that natural systems can be described “in terms of spatiotemporally continuous causal processes and local causal interactions” (Salmon, 1984, 245) that is the potential problem.  

Salmon first discusses the traditional two-slit experiment and the idea of wave-particle duality often invoked in describing the phenomenon. Assume electrons are sent one-at-a-time through a screen with two slits and on to a detector. They appear to exhibit particle-like behavior in that the detected “hits” are consistent with passing through an open slit. However, when both slits are open the electrons also seem to exhibit a wave-like character. While each hit is localized, the series of detected particles displays the interference pattern of a wave. Salmon thinks his process account fits both classical particle and wave phenomena, but when the two-slit experiment is conducted with a quantum particle he worries that there “cannot be a single consistent description that explains what happens in terms of spatiotemporally continuous causal processes and local interactions” (Salmon, 1984, 245).

Salmon’s concern is understandable, although his gloss on this familiar example perhaps understates the extent of the difficulty of interpreting what QM implies about the location of particles. Compared to classical mechanics, where the position of a particle with a given mass can be predicted based on initial conditions and application of Newton’s laws (and other Newtonian assumptions), quantum mechanics offers a very different picture. Assuming for simplicity a particle of a given mass is constrained to be

---

165 While I focus on the discussion in Salmon (1984), it should be noted that the later conserved quantity theories of Salmon and Dowe (2000) do not address this concern: both of these begin by defining causal processes in terms of world-lines of objects (see discussion in Chapter 3, section 5).
located in one-dimensional space (on the $x$ axis),\textsuperscript{166} the practitioner determines the particle’s (position-space) wave function $\Psi(x, t)$ by solving the (time-dependent) Schrödinger equation:\textsuperscript{167}

\[
\frac{i\hbar}{\partial t} \frac{\partial \Psi}{\partial t} = -\frac{\hbar^2}{2m} \frac{\partial^2 \Psi}{\partial x^2} + V\Psi
\]

Here, one is not solving for a particular position, but for a function of position and time. The only connection to a particular location is via the additional, statistical, role played by the wave function, whereby the probability of finding the particle in a particular location (between say, points a and b) at time $t$ is given by:

\[
\int_a^b |\Psi(x, t)|^2 dx
\]

However it is interpreted, the formalism of QM earns its keep by giving accurate probabilistic predictions in this fashion. Returning to the wave function, however, it should not be inferred from this simple example that the particle position is straightforwardly “spread out” over actual space prior to measurement. It is a function from points in a configuration space to a (complex) number where the configuration space represents all the possible determinate position values: for multiple (N-) particle

\textsuperscript{166} This is common in introductory discussions: see Griffiths (2005), Ch.1.
\textsuperscript{167} The general version of the time-dependent equation is $i\hbar \frac{\partial \Psi}{\partial t} = \hat{H}\Psi$, where $\hat{H}$ is the Hamiltonian operator appropriate for the system.
systems this space will have $3N$ dimensions. The wave function alone does not in any straightforward way characterize quantum particles (prior to measurement) in terms of a location in our familiar space (or space-time).

Mathematically, the wave function can be represented as a vector. The most general representation of a state of a quantum system is as a (normalized) vector in a (complex valued) Hilbert space. The wave function presentation corresponds to a vector expressed in a particular basis chosen based on the observable of interest (position in the case above). For a brief discussion of the vector representation, it will be helpful to use a simpler example involving the property of spin for particles such as the electron. The counterintuitive aspects of other examples discussed derive from the same essential phenomenon of quantum superposition displayed in this case.\textsuperscript{168} For a quantum particle, spin is an intrinsic form of angular momentum. Along any axis ($x$, $y$, and $z$), there are only two possible determinate values, which can be labelled informally “spin up” and “spin down” (the Hilbert space will have only two dimensions). However, if an electron is prepared with a determinate spin value measured, say, in the $z$ direction (call this observable “$z$-spin”\textsuperscript{169}), then it will not have a determinate value in the $x$ direction. It will be in a \textit{superposition} state with regard to $x$-spin.\textsuperscript{170} In the vector representation (using

\textsuperscript{168} This is the strategy followed by Albert in his exploration of QM (1992), and in a recent concise introductory overview provided by Ney (2013). Griffiths says that the spin case provides “the simplest and cleanest context for thinking through the conceptual paradoxes of quantum mechanics (Griffiths, 2005, 177).” An introductory QM text that begins with an analysis of spin systems is McIntyre (2012).

\textsuperscript{169} Observables are mathematically represented by (Hermitian) operators, which perform transformations on the state vectors. Eigenvectors of an observable are those vectors left unchanged by the operator, except for multiplicative numbers called eigenvalues. Eigenvalues of the operator are the only possible results of a measurement of the associated observable. In the example, $|\text{-spin up}\rangle$ and $|\text{-spin down}\rangle$ are the two eigenvectors associated with the $z$-spin observable and what the text labels “spin up” and “spin down” are the eigenvalues (actually $\pm\hbar/2$ for a spin-$1/2$ particle like the electron). If $z$-spin is measured and the system is in either of these eigenvector states, the associated eigenvalue will be found with certainty.

\textsuperscript{170} $Z$-spin and $x$-spin are so-called incompatible observables which are subject to the uncertainty principle.
notation derived from Dirac), the state of the system in the $x$-spin basis may be represented as a sum of vectors as follows:

$$ |z - \text{spin up} > = \left( \frac{1}{\sqrt{2}} \right) |x - \text{spin up} > + \left( \frac{1}{\sqrt{2}} \right) |x - \text{spin down} > $$

A quantum superposition is a sum (or difference) of vectors in Hilbert space: this represents a state of affairs for which we have no easy intuition. Note that the wave function presentation could be used here as well. As before, it is a function from points in a configuration space representing the possible determinate values of the observable to a (complex) number: the relationship to the vector representation is that these numbers correspond to the vector components in the appropriate basis.

In the state depicted, the electron’s spin has neither a determinate value of up nor of down in the $x$ direction. However, if $x$-direction spin is measured in an experimental setting, then it will be found to be either spin up or spin down. As discussed above, but now in a simpler discrete setting, taking the absolute value of the vector components (or amplitudes) associated with each value and squaring yields a probability for the measurement outcomes. In this case, we can predict the outcome of measuring the particle as either spin up or spin down to be $|{(1/ \sqrt{2})}|^2 = 50\%$.\footnote{Finally, while the dynamics of this simple example are not very interesting, it is important to note that the evolution in time of the system between preparation and measurement can also be expressed by the appropriate version of the Schrödinger equation.} A central conundrum about QM results from the fact that we never observe a superposition, but instead always observe a determinate outcome with the calculated probabilities. The system no longer
evolves according to the Schrödinger equation. Instead, we have a “collapse” to a particular value.\textsuperscript{172} The spin example reminds us that it is not just position which is not well defined prior to measurement: the problem of superposition applies to all observable properties of the particle.

Returning to Salmon, he proceeds to consider what he sees as the most problematic paradoxical quantum effect, which is that of space-like separated but entangled particles in EPR-style scenarios.\textsuperscript{173} Entanglement describes a situation where two or more particles that have interacted with each other cannot be described by states on their own, but are represented in QM by what is termed a non-separable state. In the case of two such particles (labeled 1 & 2), the state may be as follows:

\[
|\psi\rangle = \frac{1}{\sqrt{2}}|\text{spin up}_1\rangle|\text{spin down}_2\rangle - \frac{1}{\sqrt{2}}|\text{spin down}_1\rangle|\text{spin up}_2\rangle
\]

Measurements of a given observable performed on either system 1 or 2 in this case will not give independent outcomes, but instead will be correlated with each other. Like the case of superposition, this phenomenon of entanglement is a situation for which there is no classical analogue. Using the example of spin, if two such particles (such as the

\textsuperscript{172} Here it would generally be more precise to say that the composite system consisting of the electron and the detecting instrument ceases its dynamic evolution and the two are \textit{jointly} observed in a determinate state where either \{the electron is spin up and the instrument registers “spin-up”\} or \{the electron is spin down and the instrument registers “spin-down”\} with 50\% probability (see Ney, 2013, pp. 22-24). But where exactly the line between systems should be drawn in the context of quantum measurement is not defined by the theory itself: see discussion in section 3 below. Note that immediately following the measurement, the state of the system will be represented by the vector associated with the measurement outcome, and this state will resume a dynamical evolution in accordance with the Schrödinger equation.

\textsuperscript{173} “EPR” here refers to Einstein, Podolsky, and Rosen, who described an example of this type in an article authored in 1935.
Experiments on entangled particles have been carried out since the 1970’s with increasing refinement. A recent and compelling result is presented in Hensen, et al. (2015).

positr(on/electron pair considered by Salmon below) are entangled, quantum theory would say there is no fact of the matter regarding the spin states for the individual particles prior to measurement (this is the case regardless of basis). Nevertheless measurement of the value for one of the particles will also determine the value of the other:

This system remains a single system that can be described in terms of a single state function even after the parts have been spatially separated from one another. Thus a measurement of the spin of the positron is an interaction with the entire system, and the state of the entire system is modified as a result. The state describes the electron as well as the positron. If we look at the EPR experiment in these terms, we have a special case of the general problem of the ‘collapse of the wave function’ (or ‘reduction of the wave packet’). Just as the localized detection of an electron described in terms of a wave that is spread out in space involves an instantaneous change of state over a large region, so, also, the spin measurement on the positron has an effect upon the state of the system that is also spread over a large spatial region. It is this aspect of quantum theory that is so deeply puzzling from a causal standpoint. (Salmon, 1984, 251, emphasis original)

Salmon’s discussion is flawed, in that one cannot picture the quantum state of a two-particle system as a wave spread out in actual space in the absence of measurement, but it is well-confirmed that in these systems the result of the measurement in one location will determine the result of a second measurement at a distant location.174

174 Experiments on entangled particles have been carried out since the 1970’s with increasing refinement. A recent and compelling result is presented in Hensen, et al. (2015).
So, can physical phenomena described by quantum mechanics be understood in a way that is compatible with conceptions of causal process theory? Recall that in grappling with the implications of quantum phenomena, Salmon is adamant that instrumentalist moves must be rejected. He also argues that an “epistemic” conception of explanation (like Hempel’s) that would employ laws without reference to “ontic” processes or mechanisms will also be unsatisfactory in the quantum domain. He briefly considers the idea that QM should force a revision to his causal conceptions, but finally leaves the issue open: he asks whether the EPR result requires a “special mechanism”, or rather whether it is “one of the fundamental mechanisms by which nature operates?” (Salmon, 1984, 258, emphasis original). Importantly, entanglement cannot be exploited to send faster-than-light signals, and therefore positing a “special” superluminal causal mechanism is not mandated. Salmon’s latter italicized comment seems to convey that he was hopeful his causal theory could stretch to encompass quantum phenomena without a drastic change to its structure.

To summarize, given the basic framework of QM, causal process theory faces a problem of compatibility. It appears that in order to meet the challenge, there cannot be any stipulation that elementary causal processes follow definite space-time paths between interactions, and the theory should also allow that modifications to processes introduced by local causal interactions may have non-local consequences. In addition to meeting these constraints, however, there is a broader goal of providing a target for mapping the causal elements of a scientific explanation featuring QM. One might pose the challenge

---

175 At one point Salmon tentatively speculates that relativistic quantum field theories, notably quantum electrodynamics as depicted by Feynman, might comport somewhat better with the causal process conception, but the case for this is not made very clearly (see Salmon, 1984, pp. 255-256). The present discussion will be limited to non-relativistic QM. As a result, the discussion below includes some inconsistencies in that it combines references to space-time with discussion of time qua external parameter in the QM context.
If we turn from the Hilbert space formalism to the wave function representation, then it is true that the position of a one-particle system is a function on a 3D space, but as discussed above this is not the case for multi-particle systems, where the required dimensions total 3N.

2. Propensities as a Bridge from QM to the Revised Process Theory

One of the criticisms of Salmon’s mark-transmission theory was that it was fairly vague about the make-up of causal processes: they are characterized usually as possessing a “structure” or (as in his formal definitions – see the appendix) having “characteristics.” Taking the example of an electron, how would the representations deployed in QM map onto these notions? A first rough pass might simply propose that an electron corresponds to a causal process whose “characteristics” are represented by a quantum state $|\psi\rangle$, and with its propagation described using the Schrödinger equation. The only adjustment we need to make is to dispense with our classical preconceptions about what possessing a structure or characteristics amounts to. The fact that the propagating electron might have no determinate value for an observable property is no obstacle to being considered a causal process. For example, with regard to position, the electron according to QM is not localized in familiar 3D space absent certain measurements. Its state is represented in the theory by a vector which is formally defined only in an abstract higher-dimensional space.\(^{176}\) The idea, the would be that the electron corresponds to a causal process, which

\(^{176}\) If we turn from the Hilbert space formalism to the wave function representation, then it is true that the position of a one-particle system is a function on a 3D space, but as discussed above this is not the case for multi-particle systems, where the required dimensions total 3N.
while possessing “structure” or having “characteristics,” likewise does not need to be seen as localized in 3D space between causal interactions. The only adjustment to the notion of transmission needed is the removal of any requirement that the process has a definite space-time location.

But on further consideration additional changes are needed. In the mark theory, Salmon’s causal processes are actually defined partly in terms of his causal interactions: that is, a causal process must be capable of transmitting a mark, which is a change in a characteristic. We must also ask, then, what corresponds to a causal interaction in QM. Actually, there is more than one candidate. As discussed, two quantum particles can become entangled in the absence of measurement: this creates certain correlations that can be represented as a new composite state evolving in accordance with the Schrödinger equation. But for now I will set this phenomenon aside (see section 3 below) to focus on the distinctive issue of measurement. Does the measurement of a quantum system correspond to a causal interaction in Salmon’s sense? The problem with giving an affirmative answer is that measurement appears to involve a change that cannot be viewed as an alteration of a pre-existing definite characteristic. In a measurement, a wave function (which encodes many possible outcomes) collapses and the particle is observed to have a determinate value for the measured property. It doesn’t appear that the notions used in Salmon’s framework accommodate this distinctive discontinuous change.

The concepts employed in the revised account offered in Chapter 4 appear better suited to capture this kind of change. Recall that in this account a causal process is characterized by a cluster of dispositions toward particular interactions (its dispositional profile), while causal interactions themselves are (mutual) manifestations of these dispositions. These manifestations take the form of a change in the dispositional profile of
the processes involved (the disposition partners). Here are the definitions (for elementary processes):

- A *causal process* is something that transmits a cluster of dispositions between space-time points.

- A cluster of dispositions is *transmitted* between space-time points A and B (A ≠ B) if and only if the dispositions are compresent at A and at B and during the interval between A and B.

- A *causal interaction* is an intersection of causal processes at a space-time point where there is mutual manifestation of dispositions, and where the associated modifications to the processes are transmitted beyond the locus of the intersection.

In line with the discussion above, a clarification needed so that the second provision (that dispositions are compresent “during the interval between A and B”) is not taken to imply the dispositions must be located at the points comprising the space-time interval in question. Rather, the phrase conveys the requirement that the cluster of dispositions characterizing the process remains unaltered between A and B. In the absence of interactions (which do take place at points in space-time and which modify processes) there is no requirement that processes have positions.

Turning now to the mapping of QM descriptions, let us preliminarily assume quantum particles such as electrons correspond to causal processes. Then, we can posit that the state vector prior to measurement represents the initial dispositional profile, the measurement itself corresponds to a causal interaction, and the collapsed state corresponds to the manifested change in profile that results. Note that, as always,
interactions involve mutual manifestation involving two or more disposition partners: the change in the process represented by the electron is matched by the corresponding alteration in the measuring device.\textsuperscript{177}

There is one thing lacking from the revised causal process account offered above, however, the addition of which would improve the correspondence: there is no explicit statement that manifestations may have a probabilistic character. While there are suggested interpretations and modifications of QM that seek to preserve a fundamental determinism (see section 4), a stipulation can be added to the definitions to accommodate indeterminism:

- Stipulation regarding probabilistic manifestations: a disposition may have more than one possible manifestation with the same partner(s), each with a certain probability of occurrence.

This sort of disposition may be called a propensity.

It should be acknowledged that something like this understanding of causal processes is anticipated by Salmon (1984) in the context of a discussion of probabilistic causation. Salmon explores scenarios that appear to involve a transmission of probabilistic causal influence. He considers several examples, including ones involving quantum phenomena:

\textsuperscript{177}Also, it is important to keep in mind that for more elaborate experimental settings the causal interaction is not necessarily just between a process corresponding to the quantum system and one corresponding to a detector plate or instrument pointer: the interaction may be with the entire (effectively isolated) apparatus.
If radiation of suitable frequency impinges upon a hydrogen atom, there is a certain probability that the radiation will be absorbed and the atom will exist for a time thereafter in an excited state. (Salmon, 1984, 203)

He pictures in these cases a kind of causal process “that carries with it probability distributions for various types of interactions” (203). He says “it seems to be altogether appropriate to refer to these probabilities as propensities” (203, emphasis original):¹⁷⁸

A given alpha particle, impinging upon a gold foil, has propensities of given magnitudes for no interaction, for small deflection, and for large deflection. (Salmon, 1984, 203)

He goes on to conclude:

I am inclined to think that this idea of propensity, as a probabilistic disposition, is valuable. It is just such dispositions that seem to me to lie at the foundation of probabilistic causality. (Salmon, 1984, 204)

In another article, Salmon includes a brief discussion that explicitly ties propensities to quantum mechanical phenomena:

My suggestion is that the quantum mechanical wave is a wave of propensity—a propensity to interact in certain ways given appropriate conditions. (Salmon, 1988, 15)

Despite these considerations, Salmon does not include propensities (or dispositions) in his official formulations. As discussed in Chapter 3, he wanted his official view to adhere

¹⁷⁸ Salmon goes on to say that his comments must be distinguished from an endorsement of the propensity interpretation of the probability calculus, which is a different claim (Salmon, 1984, pp. 204-5).
to an empiricist tradition that avoids any reliance on anything resembling causal powers. Cartwright (1989) thinks Salmon’s notion of propagating causal processes is best viewed in terms of their possessing a capacity. She uses this characterization of Salmon’s view as part of an analysis of quantum mechanical phenomena (pp. 243-250).

Setting aside this concern, we can see that a process account featuring dispositions/propensities has an advantage in the quantum context.

Importantly, there is also an independent tradition of appealing to propensities or kindred notions in thinking about how the formalism of quantum mechanics might be interpreted. Suárez (2007) surveys and critiques several such historical accounts, including the suggestive though vague discussion offered by Werner Heisenberg in his *Physics and Philosophy* (Heisenberg, 1958). Suárez offers his own positive proposal (Suárez 2004a, 2007) called the “selective-propensity” interpretation. He sums up the basic idea in the following way. If one asks the question of what it means, with respect to an observable, for a quantum system to be in a state like that described in equation 5.3, the answer is as follows:

\[
\text{It means that the system possesses the disposition, tendency, or propensity, to exhibit a particular value of \([\text{observable}] \, Q\) if \(Q\) is measured on a system in \([\text{superposed}]\) state } \psi. \quad (\text{Suárez, 2007, 419, emphasis original})
\]

On this interpretation an electron possesses a number of propensity-properties, including those linked to manifesting values of position, momentum, and spin. The “selective” of the selective-propensity interpretation refers to the idea that particular measurements

---

179 Cartwright (1989) thinks Salmon’s notion of propagating causal processes is best viewed in terms of their possessing a capacity. She uses this characterization of Salmon’s view as part of an analysis of quantum mechanical phenomena (pp. 243-250).

180 The other views discussed in some depth by Suárez are those of Margena (1954), and Maxwell (1988). Redhead (1987) also includes a brief discussion of the propensity interpretation (pp. 48-9), while Dorato (2007) surveys the application of notion of dispositions in several interpretations of QM (see also the discussion in sections 4.3 and 4.4 below). Suárez (2004b) discusses Popper’s unsuccessful attempt to show that a propensity interpretation of probability would resolve the mysteries of quantum phenomena like the two-slit experiment.
interact selectively with just one of these properties. In our example, the electron possesses a propensity to exhibit spin up or spin down (with 50% probability) if $x$-direction spin is measured using an appropriate apparatus. The electron possesses other propensities relating to manifestations that would result from other possible measurement interactions.

This account has an affinity with the causal framework outlined above, particularly since the notion of selective measurement parallels the latter’s use of the idea of mutual manifestation: the disposition manifested by a causal process depends on its disposition partner(s)—in this case the measurement apparatus. The main difference is that Suárez proposes that propensities are dispositional properties possessed by the system and that there exists one of these corresponding to each QM observable. In contrast, the present causal theory is limited to positing that QM states are (idealized) representations that map onto to a collection of singular dispositions/propensities (for particular mutual interactions) that characterize a causal process. The additional step of positing the existence of causal properties that directly correspond to observables in the physical theory is not taken (the general reasons for this approach are discussed in Chapter 4). However, the Suárez proposal of appealing to propensities to theorize about quantum properties is broadly in accord with the causal process approach discussed here.

To summarize, the goal of this chapter is to assess whether causal process theory can provide the ontic backdrop for scientific explanations even when the latter feature quantum physical descriptions. A preliminary positive assessment runs as follows.

---

181 There are other details to the Suárez account that I am passing over having to do with how the propensity-properties possessed by the system relate to the QM formalism. These properties are not represented by the so-called pure state of the system (as depicted, e.g., in 4.3). Rather a property is represented by a mixed state featuring a weighted sum over the eigenstates of the observable in question (see discussion in Suárez 2004a).
Assume for example an explanation of the following case. An experimental set-up begins with an ensemble of electrons in an unknown spin state. The electrons are directed into a Stern-Gerlach apparatus which filters the electrons according to their detected spin along the z-axis. The electrons recorded with spin-up are directed to another apparatus that measures spin along the x-axis. Approximately half of these are found to be spin-up and half spin-down. Why is this result achieved? In the scientific explanation, the electrons are quantum systems with states described by vectors according to the theory. A given electron in the initial condition has an unknown state $|\psi>$, evolving in terms with the Schrödinger equation. The first (state preparation) apparatus collapses the state onto one of the two possible eigenvalues of the z-spin observable. The electrons resume evolution: because of the linearity of the Schrödinger equation, the state remains the same, either $|z - \text{spin up}>$ or $|z - \text{spin down}>$. The electrons measured spin up are then directed to the second measurement apparatus. According to QM, the z-spin up state is in a superposed state in the x-spin basis. The measurements collapse the wave function, and the statistical theory of quantum measurement (the Born rule) entails that they will be found with 50% probability to be spin up or spin down. This explanation is a causal one because of following mapping: the electrons correspond to propagating causal processes; their states map onto a characteristic cluster of propensities toward (mutual) manifestations; and the measurements correspond to the manifestation of particular propensities in causal interactions.

3. Comparison to the Orthodox Interpretation and the Challenge of Entanglement
Causal process theory preliminarily seems consistent with an explanation featuring a textbook example of a simple experimental outcome. Indeed, the conceptual features of the theory have an affinity with what is sometimes called the orthodox or textbook interpretation of QM. Using Salmon’s terminology, the causal account has two distinct aspects, *propagation* (of causal processes between space-time locations) and *production* (change inducing interactions). Likewise, an early treatment of QM provided by Von Neumann describes that theory in terms of two components. Taking them in reverse order, what Von Neumann labeled process 2 is the evolution of the state according to the Schrödinger equation. Process 1 is that of measurement, which replaces the state with the weighted set of states (a mixture) associated with the possible values of the observable, one of which is realized. Unlike some early interpretations, this account does not treat quantum states in an instrumental or merely epistemic fashion, nor does it posit a necessary bifurcation between quantum phenomena and a classically described world. It thus avoids commitments that would conflict with the realist stance of causal process theory (where the causal account provides the ontic constraint for an explanation).

The interpretation has a number of perceived shortcomings. For physicists in seek of a fundamental theory of universal scope, the orthodox interpretation fails to offer a consistent physical account covering both microscopic systems and the experimental setups used in measurement. While Von Neumann’s process 2 is described in the clear

---

183 Von Neumann’s account was labeled “orthodox” by Wigner (1963). A recent explication of this interpretation (in a broader context featuring the views of Dirac and others) is in Jaeger (2009), pp. 117-120. Jaeger calls it the “basic” interpretation. I will refer to it as either as orthodox or as Von Neumann’s. While many discussions give the so-called Copenhagen interpretation a leading role (sometimes also referred to as “orthodox”), this term covers a number of different views, some of which have anti-realist flavors.
language of mathematical physics, process 1 has (at present) no accepted dynamical representation. It would seem to be instantaneous, or so fast as to be indistinguishable from instantaneous, and it appears wholly discontinuous with process 2. This is widely viewed as an unattractive state of affairs, and other realist interpretations/modifications of QM (discussed in section 4 below) seek to avoid the conclusion that these two distinct processes exist in nature. Still, it is a realist interpretation broadly compatible with the ontic picture described by causal process theory. As discussed in the prior section, the quantum system corresponds to a causal process, while its state represents the dispositional profile associated with the process. Evolution according to process 2 (between, say, preparation and measurement) corresponds to the transmission of the dispositional profile, while process 1 represents the manifestation/change in profile associated with causal interaction.184 Note again that measurements, like causal interactions, occur in space-time, while it is not required that the transmission of dispositions between interactions be localizable.

Another limitation of the orthodox interpretation is that there is no precise specification of when and where process 1 occurs. A measurement is typically depicted in terms of the interaction of the quantum system with a macroscopic measuring device. However, as Von Neumann discusses, an experimental situation can also be viewed in another way, where the measuring device is considered part of a composite system along with the quantum particle, which is then subject to a measurement interaction with the human observer. The precise placement of the process 1 interaction is arbitrary.185 If fact,

184 A philosopher who has endorsed a two-process understanding of QM is Brian Ellis in his Metaphysics of Scientific Realism (2009; see pp. 66-69 and 73-84).
185 The Von Neumann interpretation is often associated with the view that human consciousness plays a role in collapse of the wave function. But this is not a necessary component of the interpretation. As Von Neumann explains in Mathematical Foundations of Quantum Mechanics, it is not that subjective
others have suggested that the indeterminacy involved is even broader than this. Wigner describes a scenario where he asks a friend to conduct the measurement on the quantum particle on his behalf: As far as QM is concerned, the friend can be considered to be in a non-collapsed composite state with the quantum system (and the measuring device) from Wigner’s perspective outside the lab.\textsuperscript{186} Would the state collapse only when he communicates with his friend? The orthodox interpretation has no objective criteria for answering such questions.\textsuperscript{187} The GRW program discussed below is an attempt to formulate an account that addresses this issue. However, before examining GRW it is worth summarizing how these questions surrounding measurement impact the proposed correspondence between QM and process theory.

In and of itself, the uncertainty regarding when and where measurements occur does not lead to a compatibility problem. Of course, we would like to have a more precise account of measurement, but as long as the concepts used in scientific explanations can be mapped onto those in the causal theory, then the purpose of the latter is fulfilled. The real problem in the present discussion has to do with how QM describes interactions between systems that are inferred to have occurred in the \textit{absence} of measurement. As in the EPR-type scenario discussed in section 1, quantum-mechanically described systems may establish correlations with each other, creating new larger systems. We can picture

\textsuperscript{186} See Wigner (1961), pp. 289-98.

\textsuperscript{187} There is an interesting interpretation of QM that looks to treat all interactions on an equal footing. This is the relational interpretation of QM (Laudisa & Rovelli, 2013). The idea is that when two systems interact, the interaction is measurement-like “for” the systems directly involved, but only results in correlations when considered relative to a further system. The price to be paid is the loss of the notion that state descriptions and interactions have a fully objective character.
these correlations, represented by appropriate composite quantum states, as corresponding to correlations between the dispositional profiles of two or more causal processes. However, the definitions provided for the causal theory in Chapter 4 say that causal interactions take place at space-time points, and imply that these interactions are the sole source of change. If measurements are taken to correspond to causal interactions, and there is only one type of causal interaction, then these non-measurement, correlation-establishing, quantum interactions pose a potential source of conflict.

As discussed in Chapter 4, many features of the scientific descriptions involve idealization. In explanations, these deviations from fidelity to the ontic causal structure underlying the target phenomena are justified by epistemic and pragmatic benefits. In the case of quantum mechanics, however, entanglement is not something which can be viewed as a product of idealization. It seems forced onto physics by a careful examination of microscopic systems, and it is telling that there appear to be no major (realist) approaches to the interpretation of QM in which it does not feature in some way. To provide an adequate ontic target for mapping quantum phenomena, then, an allowance must be made for to provide for a correspondence with non-local correlations/entanglement. Specifically, if causal interactions are limited to “localizing” interactions (corresponding to collapses in the physical theory), then a separate provision must allow for the establishment of correlations with other processes. These correlations impact the probabilities associated with causal interactions (corresponding to measurement outcomes) in the following way: the outcomes now also depend (mutually) on causal interactions involving the correlated partners. So, we must add a second stipulation as follows:
• A *causal process* is something that transmits a cluster of dispositions between space-time points.

• A cluster of dispositions is *transmitted* between space-time points A and B \((A \neq B)\) if and only if the dispositions are compresent at A and at B and at each stage of the process between A and B.

• A *causal interaction* is an intersection of causal processes at a space-time point where there is mutual manifestation of dispositions, and where the associated modifications to the processes are transmitted beyond the locus of the intersection.

• Stipulation regarding probabilistic manifestations: a disposition may have more than one possible manifestation with the same partner(s), each with a certain probability of occurrence.

• Stipulation regarding correlated manifestations: during transmission, a group of two or more causal processes may establish correlations such that the probabilities associated with manifestations for each also depend on causal interactions involving other members(s) of the group.

The need for these kinds of provisos to accommodate quantum phenomena shouldn’t be surprising, and still allows a case to be made for a broad compatibility of causal process theory with explanations featuring QM descriptions.

There is one more feature of interest in textbook quantum mechanics which is worth a brief discussion alongside entanglement: the indistinguishability of elementary particles of the same type. Such particles, which can often be investigated separately in practice (i.e. described using separate wave functions), are indistinguishable in principle
according to QM when part of a single system (described using a single wave function). Here if the position and spin values of two particles are swapped, there will be no empirical consequences: in particular, $|\psi|^2$ is unchanged. There are two ways this indistinguishability can be reflected in terms of the composite wave function: it can be unchanged following the swap or it can change sign ($\psi \rightarrow -\psi$). In other words the wave function is said to be symmetric or anti-symmetric with regard to the particle interchange. These two possibilities describe the two families of particles, bosons and fermions, respectively (in alignment with their division into integer-spin and $\frac{1}{2}$-integer-spin types). The fact that electrons in particular are fermions, and therefore have wave functions which are anti-symmetric with regard to their interchange, has important consequences for our understanding of atoms and molecules. For instance, it follows that two electrons with the same spin values cannot occupy the same point in space (because in that case an interchange would leave the wave function unchanged). The question can be then asked: is this feature of quantum theory compatible with the present description of causal processes (including the two stipulations noted above)? In this case, there appears to be no reason why this kind of composite state cannot be viewed as a special case of a correlation between the dispositional profiles of two or more processes. There is no reason to stipulate that there are actually indistinguishable processes for the purposes of providing an ontic constraint on explanations that incorporate this feature (although, of course, such a metaphysical possibility is not ruled out).

4. Other Interpretations and the Question of Compatibility
Given the perceived shortcomings of the orthodox account, many other interpretations and modifications of QM have been proposed. Below is a brief look at three of these (limiting the scope to realist approaches) and an assessment of how they reflect on the current discussion. Following Putnam (2005), it is useful to divide approaches to thinking about QM according to whether they accept the reality of the wave function collapses associated with measurement or not. Examples of realist views that omit collapse (discussed further below) are the Everett interpretation (usually glossed as the many-worlds interpretation),\textsuperscript{188} and the program of Bohmian mechanics\textsuperscript{189} (which supplements the basic picture of QM with additional structure to privilege a determinate position for particles at all times).\textsuperscript{190} A research program that does incorporate collapse is the GRW approach (from Ghirardi, Rimini & Weber, 1986), which will be discussed first.

4.1 GRW and Dispositional Ontology

The GRW approach incorporates the collapse of the wave function, however, this is done in a way that is importantly different from the two-process orthodox interpretation. GRW is an attempt to develop a new dynamics of QM (non-linear and stochastic) that would effectively subsume both process 1 and 2. A wave function evolving in accordance with GRW dynamics will spontaneously collapse from time to time. The rate of collapse is linked to the size of the system by new constants of nature incorporated into the approach. The values of these constants are such that an isolated microscopic system may evolve in a superposed state for an extended period, while macro-sized composite systems will collapse extremely quickly (a “hit” to any one constituent spreads

\textsuperscript{188} Named for Hugh Everett III, who first advocated the view in his 1957 doctoral thesis.
\textsuperscript{189} Named after David Bohm. The original exposition is in Bohm (1952).
\textsuperscript{190} The family of “modal” interpretations also involves privileging definite values for a subset of properties (see Lombardi & Dieks, 2016).
to all of those in the composite). In the case of a macroscopic instrument pointer with two possible settings, a collapse into one of the possible positions can expected to ensue with the appropriate probabilities (see Ghirardi, 2016). Essentially, the values of the constants chosen by the theorists allow for predictions in line with experimental findings. An important aspect of modifying QM in this way is that all wave function collapses take place in the position basis: the outcome of a measurement we observe is always via the position of some macro-sized detection device which has established a correlation with the measured system.\textsuperscript{191}

In terms of correspondence with process theory, the fact that interactions are not necessary for collapse (despite the fact that as a practical matter they accompany it in practice) introduces a key difference compared to the two-process interpretation. The theory treats a “measurement interaction” as a non-measurement interaction followed by a spontaneous position-basis collapse. With this said, it is open to the causal theorist to view this as an idealization in the theory, and map an explanation featuring QM representations onto the causal nexus as before. However, there is a loss of conceptual alignment with this alteration.

Interestingly, Dorato & Esfeld (2010) argue that an ontology featuring dispositions provides a superior interpretation of the GRW theory.\textsuperscript{192} They propose that quantum systems possess a “disposition for spontaneous localization” (Dorato & Esfeld, 2010, 43, emphasis original).” Specifically, they propose to interpret the properties or states of quantum systems as such dispositions. The manifestation of the disposition (or “power” or “propensity”) is the collapse. Unlike in the present causal account, this would

\textsuperscript{191} This is the basis for a potential objection to the theory. See Albert (1992), pp. 100-111, and Ghirardi (2016), section 10.
\textsuperscript{192} This had also been suggested by Dorato (2007) and Suárez (2007).
not be a mutual manifestation: the authors explain that, in keeping with the GRW theory, there is no need for an external triggering event. Still, it is interesting to see that the authors provide another example of the move to utilize dispositions and their manifestations as providing the ontological basis for a QM theory.

At this point, it is worth mentioning that in addition to Putnam’s breakdown, recent philosophical discussions of (realist) interpretations of QM have highlighted another dimension along which these may differ. This is a distinction between approaches that posit ontology based closely on the wave function, versus those which propose the addition of a supplemental “primitive ontology.” Briefly, advocates of deploying primitive ontologies see insurmountable difficulties in explaining the evident features of the macroscopic world based on an attempt to read ontology solely off of the wave function. The additional ontology may involve, for instance, postulating particles in a 3-dimensional space, as in Bohmian mechanics, where the wave function then plays the role of determining the dynamics of the particles. In the case of GRW, two ideas have been put forward to supplement the (modified) wave function with additional primitive ontology. The first, due to Bell (1987), envisions that the GRW collapse events, which are localized in 3-dimensional space, are primitive ontological elements. This is the so-called flash ontology interpretation of GRW (typically labeled GRWf). Another proposal couples the GRW framework to a continuous matter density field in 3-D space (GRWm- introduced in Ghirardi, Grassi & Benatti, 1995).

With this as backdrop, Dorato & Esfeld explain that their interpretation can be considered an alternative ontology for GRW where dispositions are fundamental. One

---

193 For discussion of the primitive ontology approach, see Allori, Goldstein, Tumulka & Zanghi (2008) and Allori (2013). Belot (2012) analyzes options for how the wave function should be interpreted in the approach. For advocacy of (pure) wave function ontology, see Albert (1996, 2013) and Ney (forthcoming).
interesting suggestion they make is to view Bell’s flashes as the manifestations of the dispositions. They don’t place much emphasis on this idea, however, saying that dispositions are “more fundamental than flashes because flashes are the manifestations of the dispositions to collapse (Dorato & Esfeld, 2010, 47, emphasis original).” Here I would argue that the conceptual tie between dispositions and manifestations argues for treating both as equally fundamental.\textsuperscript{194} I note also that the resulting combination of non-local dispositions and local manifestations mirrors how the issue of locality is treated in the causal process framework and appears to be an attractive way to draw a connection between QM and the events (flashes) in space-time that comprise the manifest world.

With ontology in focus, one can view the discussion from section 3 above in these terms: causal process theory, featuring non-localized dispositions and local causal interaction events, might provide an attractive starting point for developing an ontology consistent with the orthodox interpretation. It also can provide a basis for one consistent with GRW theory, if, as discussed, one treats the lack of necessity of interactions in the collapse process as an idealization incorporated into the model.

4.2 Bohmian Mechanics and the Reducibility of Dispositions

So far, our discussion of QM has taken collapse seriously as a natural process. However, given the concerns discussed in section 3, some interpreters avoid doing so. As mentioned, in Bohmian mechanics, a primitive ontology of particles with trajectories in 3D space is posited, and the wave function’s role is often described as that of guiding the

\textsuperscript{194} I note that in a later article, Esfeld (2014) takes a different view of the relationship between dispositions and flashes than is proposed in the paper discussed here. Instead of being possessed by microsystems, the dispositions are viewed as possessed by the global configuration of flashes. In a similar fashion, Egg & Esfeld (2015) argues for attribution of dispositions to the (entire) matter density field in the GRWm theory.
trajectories.\textsuperscript{195} Here, there is no collapse of the wave function: at all times the particle positions are determined. The probabilistic appearance of measurement outcomes is a due to our ignorance given the inaccessibility of the complete underlying dynamics: if the wave function represents a real entity that guides the particles, it exists not in our familiar space, but in the high-dimensional configuration space.\textsuperscript{196}

At a first pass, one might attempt to map the Bohmian framework onto causal process theory, beginning with the suggestion that the particles correspond to causal processes. One problem with this is that the non-position properties we associate with quantum particles (such as spin) would not map onto the dispositional profiles of the processes these particles would represent. Experiments involving spin have the results that they do not because the property is possessed by the particles, but because the overall configuration of particles is guided such that those making up the measured system and the measuring apparatus (including, e.g., instrument pointers) adopt the positions we associate with a measurement outcome. Indeed, the proper scope for understanding the guidance of particles is at the level of the universe. On this point, Bell notes that “in principle the correct application of the theory is to the world as a whole,” while the “singling out of a ‘system’ is a practical thing defined by circumstances, and is not already in the fundamental formulation of the theory (Bell, 1987, 114).”\textsuperscript{197} The notion of a measurement “interaction” between, say, a single particle and a macroscopic device (comprised of many particles) thus has a different meaning than in the orthodox

\textsuperscript{195} Bell (1987) includes a helpful discussion of the theory. Also see Dürr, Goldstein & Zanghi (1992) and the overview provided by Goldstein (2017).

\textsuperscript{196} To replicate the probabilistic outcomes of QM, certain assumptions about initial conditions must be made (See Dürr, Goldstein & Zanghi, 1992).

\textsuperscript{197} Goldstein & Zanghi similarly write that “from a fundamental point of view, the only genuine Bohmian system in a Bohmian universe—the only system you can be sure is Bohmian—is the universe itself, in its entirety (Goldstein & Zanghi, 2013, 94).”
interpretation. The combined configuration must be broken down in order for the trajectory of the particle to be considered separately from the apparatus. Advocates of the view can indeed make this move to formulate a Bohm-style interpretation of what effectively occurs in measurement scenarios such as those involving spin (for a recent discussion, see Bricmont, 2016, pp. 140-50). With this in hand, it seems that one could look to map the scenario onto the causal process account after all. However, adherents to the Bohmian theory would properly view such a causal account as reducible to the more fundamental scheme of combined particle configuration and guidance wave. Pagonis & Clifton (1995) discuss the idea that if spin is considered to be a dispositional property, then in Bohm’s theory it is reducible to categorical properties (positions). Dorato (2007) refers to this work in coming up with a classification of QM interpretations that can be viewed as compatible with the ascription of either irreducible or reducible dispositions. He concludes that collapse interpretations imply that quantum properties are irreducible dispositions (since there are no definite values in the absence of collapse), while Bohmian mechanics implies reducibility.

On the other hand, one might still interpret Bohmian mechanics using a (irreducibly) dispositional ontology at least for the purposes of explaining individual particle motion. Suárez (2015) advocates an interpretation where the wave function is not a governing object or law, but a mathematical tool for describing or representing the dispositions of particles toward particular motions at each point in 3D space. These individual particle dispositions would have, of course, a non-local character, and depend

---

198 The procedure for assigning an “effective” or “conditional” wave-function to a subsystem corresponding to the measured system is discussed in Dürr, Goldstein & Zanghi (1992), section 5 (also summarized in Goldstein, 2017, section 8).

199 Specifically, Dorato (2007) associates irreducible dispositions with Bohr’s interpretation, the relational interpretation (see footnote 187), and GRW.
on the properties of all the other particles (again, the wave function applies to the entirety of the 3D system of particles). Because of this, the dispositions Suárez associates with individual particles have a somewhat different character than those characterizing fundamental causal processes in the present theory, where entanglement of dispositions is possible, but not mandated.\textsuperscript{200}

4.3 The Many-Worlds Approach and Causal Explanation

Whether a deployment of the causal account makes any sense against the backdrop of the Everett (or many-worlds) interpretation of QM is less clear. In this interpretation there are no collapses, but also there is no further modification or supplement to orthodox QM.\textsuperscript{201} The idea is to take the wave function/Schrödinger equation formalism as a complete description of the universe. It follows that everything that exists is subject to superposition and entanglement, and, despite appearances to the contrary, all of the possible physical scenarios associated with the quantum state exist in an even-handed way. In the most common way of thinking about the thesis, these possibilities correspond to a plentitude of branching quasi-classical “worlds” comprising the universe. What is striking about this interpretation is not only its departure from our everyday view of the natural world, but also how revisionary it is from the perspective of the science: quantum mechanics, the physical theory which probabilistically but accurately predicts experimental outcomes, is re-interpreted in a fashion that denies there are particular experimental outcomes. All possible outcomes exist.

\textsuperscript{200} The holistic nature of the guidance provided by the universal wave function motivates Esfeld, Lazarovici, Hubert, & Dürr (2014) to argue that in attributing a dispositional ontology to Bohmian mechanics, the disposition is properly seen as a property of all the particles together at a given time. Dorato (2015), however, offers reasons to prefer a position closer to that of Suárez (2015).

\textsuperscript{201} Recently, the idea of adding a primitive (mass density) ontology to the scheme has also been proposed by Allori, Goldstein, Tumulka & Zanghi (2011).
There are a number of ways that this idea is developed featuring various strategies to account for why things illusorily appear to us as they do. A leading role is typically assigned to the theory of decoherence, which describes how the distinctive interference effects associated with superposition can be suppressed in a way consistent with the Schrödinger evolution. Turning to the issue of compatibility with the causal theory, perhaps to the extent a given “world” is effectively isolated from the rest, the causal account might be associated with a sequence of branching events that comprise that world in something like the usual fashion. However, this correspondence wouldn’t accurately reflect the true nature of that world. Its phenomena would not be due to a network of processes and interactions in that world, but would be fixed by the evolution of the larger reality of which it is a part.

It appears that non-collapse interpretations have less of an affinity with causal process theory, with the many-worlds picture being especially incongruent. Compared to the present project, the many-worlds thesis represents a different approach to reasoning toward the ontic structure implied by scientific theory and practice. First, there is a difference in emphasis regarding the importance of local vs. global explanations. Wallace, arguing against taking an instrumentalist view of QM, says that “the purpose of scientific theories is not to predict the results of experiments: it is to describe, explain and understand the world (Wallace, 2012, 24).” Wallace does not highlight the goal of explaining (as well as predicting) particular experimental outcomes. An advantage of the

---

202 See Bacciagaluppi (2016). Other difficult issues remain, including how to justify that the quantum state should be thought of as representing a collection of worlds, and how probability can have any meaning if all outcomes are always realized. For an overview see Saunders (2010) and Vaidman (2016). For a thorough defense of the Everettian position, see Wallace (2012).

203 In Dorato’s assessment, dispositions would be eliminable when a local “world” is considered, although he also suggests that dispositions could be seen as playing a role at the level of the universe (Dorato, 2007, p. 264).
causal interpretation offered here is its support of this “local” explanatory endeavor. If successful, the many-worlds program will also explain, in its revisionary way, the appearance of these outcomes, but the focus of the interpretation is “global”: one takes QM-without-collapse to be a universally applicable account and then works to recover the local phenomena. Collapse theories, while also seeking universality, are generally less revisionary about the character of local facts compared to non-collapse approaches. For a causal theory that hopes to underpin a conception of natural science explanations (which generally deal with particular phenomena), supporting the local explanatory endeavor is vital. It is also the case that there is no reason, at least at this juncture, to think that the causal nexus cannot also underlie a compelling global explanation of the world’s fundamental nature.204

Another difference in focus relates to the status of physical models and theories as representations. Many-worlds advocates focus closely on the abstract mathematical content of the physical theory, assuming it faithfully tracks concrete reality.205 The causal theory, on the other hand, is explicitly a target for various scientific theories throughout the natural sciences. It is presumed that application of these will involve abstraction and idealization.

These differences are also related to opposing positions in a long-standing debate regarding the relationship of causation to fundamental physics. In a tradition usually traced to Russell (1912/13), the temporally symmetric dynamical equations of physics are thought to be incompatible with causal concepts. Defenders of a causal perspective on

204 Compared to other sciences, universality is typically taken as a core theoretical goal of physics. On the other hand, when it comes to QM (in all of its interpretations), there is still no incorporation of gravity. So, perhaps taking a particular stance on QM based primarily on the thesis of global applicability will prove premature.

205 For discussion of this point, see Maudlin (2010).
physics note that when these equations are actually used as a part of an explanation of an observed phenomenon, along with assumptions about initial conditions and specifications of boundary conditions, causal concepts are routinely invoked, as they are in other sciences. The pursuer of the Russelian line, like the many-worlds advocate, focuses on the mathematics and assumes its form conveys something fundamental about the nature of the universe. But this move gives insufficient weight to the considerations which motivate the development and use of these equations in the experimental and observational arenas. It is this latter context which makes it appropriate to map physical descriptions onto a theory of underlying causal structure.

5. Conclusion

Compared to classical physics, QM offers distinctive challenges for the advocate of causal process theory: these include non-locality, irreducibly probabilistic manifestations and (especially) the correlation of these probabilities via entanglement. If these features are accommodated, though, there is a strong compatibility between causal process theory and how orthodox QM theory is used to explain phenomena. On the other hand, we’ve seen that QM prompts a variety of proposals for interpretation, supplement or modification, and some of these approaches incorporate notions that may be harder to reconcile with the causal theory. Further, the problem of reconciling QM with relativity theory brings further complications not discussed here. However, the mapping of quantum physical representations to the causal nexus is well-motivated by the way the

\[\text{206 For a recent argument on these lines, see Frisch (2014).}\]
theory is used by scientists, and this perspective also provides an avenue toward a
consistent treatment of scientific explanations throughout physics, chemistry and biology.

Appendix: Salmon’s Definitions

A. The original formulations of the mark-transmission theory (from Salmon, 1984) are
as follows (note the use of counterfactuals):

1. Mark-Transmission Thesis

MT: Let P be a process that, in the absence of interactions with other processes, would
remain uniform with respect to a characteristic Q, which it would manifest consistently over
an interval that includes both of the space-time points A and B (A ≠ B). Then, a mark
(consisting of a modification of Q into Q’), which has been introduced into process P by
means of a single local interaction at point A, is transmitted to point B if [and only if]²⁰⁷ P
manifests the modification Q’ at B and at all stages of the process between A and B without
additional interventions. (Salmon, 1984, 148, emphasis original)

2. Principle of Causal Interaction

CI: Let P₁ and P₂ be two processes that intersect with one another at the space-time point S,
which belongs to the histories of both. Let Q be a characteristic that process P₁ would
exhibit throughout an interval (which includes subintervals on both sides of S in the history
of P₁) if the intersection with P₂ did not occur; let R be a characteristic that process P₂
would exhibit throughout an interval (which includes subintervals on both sides of S in the

²⁰⁷ The ‘only if’ was omitted by Salmon, but is implied by his discussion.
history of P2) if the intersection with P1 did not occur. Then, the intersection of P1 and P2 at S constitutes a causal interaction if:

(1) P1 exhibits the characteristic Q before S, but it exhibits a modified characteristic Q' throughout an interval immediately following S; and

(2) P2 exhibits the characteristic R before S, but it exhibits a modified characteristic R' throughout an interval immediately following S. (Salmon, 1984, 171, emphasis original)

B. An alternative formulation from (Salmon, 1994) avoids the use of counterfactuals (and corrects a circular aspect of the above definitions) but relies on the notion of a causal process possessing a capability to transmit a mark (note what the definition of “transmitted” says about space-time location):

Causal Processes and Mark Transmission

- A process is something that displays consistency of characteristics
- A mark is an alteration to a characteristic of the process that occurs in a single local intersection.
- A mark is transmitted over an interval when it appears at each space-time point of that interval, in the absence of causal interactions.
- A causal interaction is an intersection in which both processes are marked and the mark in each process is transmitted beyond the locus of the intersection.
- A causal process is a process that can transmit a mark.
Chapter 6: AIM Theory and the Interactive Conception of Chemical Bonding

According to the dispositional causal process theory developed above, a composite process is formed from a pattern of interaction between more elementary processes. Does the notion of chemical bonding between atoms in molecules correspond to this picture? To examine this question requires exploring the contemporary scientific view of bonding. This turns out to be a fraught topic, however. While quantum physics is the foundation for modern chemistry, the traditional chemical concept of bonding is not easily reconciled with a purely quantum mechanical (QM) approach to modelling molecular systems. After reviewing recent philosophical work on conceptions of bonding, I take a detailed look at one promising approach to reconciliation. The theory of Atoms in Molecules (AIM) developed by Richard F.W. Bader and colleagues offers a way to define bonding (and other chemically relevant features of a molecule) by relying on a topological analysis of the electron density distribution (\(\rho\)), which in turn can be calculated from a molecular wave function. The “bond paths” identified by the analysis are posited as indicators of a special pairwise physical relationship between atoms. While details of the theory remain subject to debate, the AIM approach embodies a distinctive interactive conception of bonding within a quantum physical framework—a conception that has an affinity with the notion of the composite causal process. One element of the AIM conception is particularly controversial, however. This is the idea that the bonding relationship can be viewed as localized in familiar, three-dimensional space. The basis for this feature is Bader’s claim that \(\rho\) represents the distribution of charge density in real space, which is an understanding that is opposed to the standard statistical role \(\rho\) plays in
the QM formalism. While AIM advocates point to the experimental estimation of $\rho$ to support their view, this does not appear to be conclusive. However, I show that an alternative way to interpret $\rho$ supports the interactive conception and also highlights its compatibility with the causal theory.

This chapter is organized as follows. Section one briefly reviews the challenges facing the notion of the chemical bond in light of modern QM modelling techniques. The focus is on the structural and energetic conceptions discussed by Hendry (2008) and Weisberg (2008). An overview of AIM theory is presented in section two, contrasting its approach with the other conceptions. In section three I argue that AIM provides a distinctive and attractive notion of bonding—an interactive conception that has a distinct affinity with the notion of a composite causal process. In section four, however, I argue that the issue of spatial localization poses a challenge for the conception. Central to AIM theory is the role played by the electron density distribution, and AIM’s interpretation of $\rho$ as describing or representing something existing in space is problematic. This is despite the support which AIM claims is provided by the ability to experimentally measure $\rho$ using X-ray diffraction. In section 5, an alternative interpretive strategy to support the spatial localization envisioned by the AIM conception is sketched. Section 6 concludes and briefly recounts the connections between this conception of bonding and the causal theory introduced in prior chapters.

1. Conceptions of the Chemical Bond in the Wake of Quantum Theory

The notion of the chemical bond played a key role in the development of modern chemistry and remains central to our understanding of molecular structure and chemical
reactions. At the same time, with the advent of quantum mechanics and its successful application to molecules, it has become difficult to reconcile the traditional idea of bonds with the underlying physical theory.

In the second half of the 19th century, several scientists developed models of molecular structure, particularly as an avenue to explain phenomena in organic chemistry. A key figure in the developing the theory of chemical bonds in the early 20th century was G.N. Lewis. Lewis (1916) distinguished two types of compounds, polar and non-polar. The former came to be described in terms of so-called ionic bonds: here electrostatic forces (which act in all directions) are responsible for the combination of oppositely charged ions. The other sort of compound (non-polar), Lewis reasoned, required the sharing of electrons. Specifically, the sharing of a pair of electrons between two atoms creates a covalent bond. In the theory, each element has a characteristic configuration of unpaired outer shell electrons: this is the raw material for creating covalent bonds and resulting molecules.

Modern (non-relativistic) quantum theory was developed in the 1920’s. Compared to classical physical conceptions, it offers a very different picture of electrons, atoms and molecules. In quantum theory the state of a system is described by a wave function. For instance, for a free particle moving in space, one cannot ascertain its position at a given time, but instead must determine its wave function $\Psi$, which is a function from the possible positions to a (complex) number. To interpret what this

---

208 Lewis recognized that most molecules have some hybrid character (Lewis, 1916, p.767).
209 Ironically, quantum theory appears to solve one problem with the Lewis model from the perspective of classical electrostatics. This is the question of why electrons would be localized in pairs given the repulsive force between them. The understanding of how quantum spin impacts multi-electron systems explains why opposite-spin electrons are found in proximity to each other. Unfortunately this “help” is overwhelmed by the other revisionary features of QM, as discussed in the main text.
means for possible measurements of the particle’s position, one calculates the probability of finding the particle in a given volume of space at a given time from $|\Psi|^2$, the square of the absolute value (or modulus) of the wave function.\textsuperscript{210} Absent a measurement interaction, the particle has no defined spatial location.

For an atom, the behavior of the system is described by the time-independent Schrödinger equation:\textsuperscript{211}

\[ \hat{H}\psi = E\psi \]

Here $\psi$ is the wave function, $E$ is the energy, and $\hat{H}$ is the Hamiltonian operator appropriate for the system.\textsuperscript{212} For an atom, the Hamiltonian will contain a kinetic energy term and a potential energy term that is based on the electrostatic attraction between the electrons and the nucleus (along with repulsion between electrons). By solving the equation, one finds the wave function and the energy: in fact, given the form of the equation, there are many solutions corresponding to many energy states (the lowest energy state is the ground state). In the case of the hydrogen atom (where the nucleus is assumed to be stationary at the origin of the coordinate system), the calculated wave functions (called orbitals) indicate the position state of the electron: again this is in terms of complex-valued amplitudes over the possible position configurations.\textsuperscript{213} Also as before

\textsuperscript{210} Equivalently, one can use the product of $\Psi$ and its complex conjugate $\Psi^*$.

\textsuperscript{211} The assumption required here is that the potential energy of the system does not change with time.

\textsuperscript{212} I follow the convention of using the lower-case $\psi$ for the time-independent equation, since it represents only the spatial dependence of $\Psi$ after applying the separation of variables method to the time-dependent Schrödinger equation.

\textsuperscript{213} The domain of the wave function for an $N$-particle system is a configuration space with $3N$ dimensions.
one can use the wave function (via the square of its absolute value) as the basis for calculating the probability of finding the electron in a given spatial volume around the nucleus. For multi-electron atoms an approximate description of possible electronic states is built up from successive hydrogen-like orbitals of increasing energy. In the context of multiple electrons, one can use the wave function as the basis for calculating the electron density distribution for the system: this gives the expected number of electrons one would find at a particular spatial location upon measurement (see section 4 for more details).

Given that electrons are not localized in quantum theory in the absence of a measurement, the idea that a molecule is formed by sharing particular electrons is problematic. Linus Pauling prominently sought to reconcile the two pictures. His approach was to interpret quantum theory as describing “resonance” structures, which were hybrid combinations of multiple possible classical configurations. Some critics, however, viewed this perspective as unhelpful in the search for a purely quantum foundation for chemistry. While precise solutions to the Schrödinger equation for

---

214 While the non-localizability of electrons will be the focus of the discussion, there is another feature of quantum mechanics that is inconsistent with the classical picture. Particles of the same type, which may at times be investigated separately in practice (i.e. described using separate wave functions), are indistinguishable in principle when part of a single system (described using a single wave function, as in the case of a multi-electron atom or molecule). Here, if the position and spin values of two particles are swapped, there will be no empirical consequences: in particular, $|\psi|^2$ is unchanged. There are two ways this indistinguishability can be reflected in terms of the composite wave function: it can be unchanged following the swap or it can change sign ($\psi\rightarrow -\psi$). In other words the wave function is said to be symmetric or anti-symmetric with regard to the particle interchange. These two possibilities describe the two families of particles, bosons and fermions, respectively (in alignment with their division into integer-spin and half-integer-spin types). The fact that electrons in particular are fermions, and therefore have wave functions that are anti-symmetric with regard to their interchange, is crucial for understanding the structure of atoms and molecules. This is because two electrons with the same spin values cannot occupy the same point in space (because in that case an interchange would leave the wave function unchanged). This consequence underlies the so-called Pauli exclusion principle, whereby only two electrons may occupy the same orbital, and they must have opposite spins.

215 See Pauling (1960).

216 See discussion in Hendry (2008), section 3.
molecules are generally intractable, techniques to use quantum theory to estimate a molecular wave function and thus calculate molecular energies and other properties were quickly developed.\footnote{One feature which figured in these efforts from the beginning is the Born-Oppenheimer approximation: all calculations start with the assumption that the nuclei are in a fixed configuration (which can be altered iteratively to find the best solution).} One approach to calculation (which retains a conceptual link to the Lewis model and resonance theory) uses what are called valence bond (VB) models. In this approach, one starts with the wave functions associated with individual atoms (two at a time) and creates hybrid orbitals from their overlap. As a result, VB-based calculations preserve a degree of localization in the resulting orbital structure, and the idea of overlapping orbitals provides an intuitive notion of a bond. While various models employ this idea, over time a competing approach has become dominant: this features the use of molecular orbital (MO) models.\footnote{For discussion of VB and MO models, see Weisberg (2008), sections 2 - 4. See also Woody (2012) for discussion of the interplay of chemical concepts and representations with the development of various quantum mechanical methods.} In this approach one constructs orbitals for all of the electrons in the molecule together (given fixed nuclear coordinates). These orbitals are not localized: they “cover” the entire molecule. This is the basis for most computational approaches today,\footnote{The focus here is on so-called ab initio MO calculations. Historically, various semi-empirical methods have been very important as well, given limitations on computational resources. Another family of models utilizing Density Functional Theory (DFT) is also frequently employed. These models estimate functions on the electron density distribution to extract information about energy and other molecular properties.} and the success of the MO methods for calculating molecular wave functions leads to questions about how the notion of a chemical bond should be viewed given the state of the science.

In assessing this question, Hendry (2008) describes two conceptions of the chemical bond. The first, the structural conception, seeks to “retain the explanatory insights afforded by classical structural formulas (917).” In order to adopt this conception in the context of quantum theory, Hendry suggests a functional approach that would
identify “physical realizers” of the role traditionally played by bonds. The requirements are that the realizers would be “material parts of the molecule that are responsible for spatially localized submolecular relationships between individual atomic centers (917).” Weisberg (2008) offers a slightly different “working definition” whereby the conception says “a covalent bond is a directional, submolecular relationship between individual atomic centers that is responsible for holding the atoms together (934).” The main challenges facing this conception are the indistinguishability and non-localized nature of electrons in a molecule.²²⁰ A possible solution Hendry discusses is the identification of the bond with “nonarbitrary” components of the electronic wave function and/or electron density distribution of the molecule (918).

The alternative Hendry outlines is called the energetic conception. Here, no part of the molecule responsible for bonding is identified. Instead facts about chemical bonding are facts about “energy changes between molecular or super-molecular states (919).” If a molecule in a bonded state has lower energy compared to its separated atoms, this represents the formation of bonds. For two atoms forming a diatomic molecule, one can plot total potential energy as a function of inter-nuclear separation and identify the minimum value associated with bonding. For a polyatomic molecule, a potential energy surface in higher dimensions can likewise be calculated from trial wave functions. The energetic conception is, as Hendry puts it, “more a theory of chemical bonding than a theory of bonds (919, emphasis added).” Weisberg (2008) argues that the idea that bonding involves energetic stabilization is a consistent, or robust, feature of the various molecular models. This favors the energetic conception. He also argues that across the

²²⁰ Hendry also discusses another objection to the conception, which is that valence structural descriptions appear to be inapplicable to some molecules. Hendry argues that this wouldn’t vitiate the applicability of the structural conception for “paradigm nonpolar substances (Hendry, 2008, 918).”
models he surveys, greater delocalization of electrons correlates with an increased match between calculated values for molecular properties and empirical estimates. This puts pressure on the structural conception, which depends on identifying the realizers of the bond role in localized regions between atomic centers.

2. Overview of the AIM Theory

The theory of atoms in molecules (AIM) developed by R.F.W. Bader and others offers an alternative approach to thinking about chemical bonding.\footnote{Also known as QTAIM (for “quantum theory of atoms in molecules”): See Bader, 1990, and expositions by Bader & Matta (2013), Gillespie & Popelier (2001, Ch. 6-7), Popelier (2000, 2014, 2016), Matta & Gillespie (2002) and Matta & Boyd (2007). The term quantum chemical topology (QCT) has been used recently to describe AIM along with additional ideas for applications of topological analysis that go beyond those originally encompassed by AIM (see Popelier, 2014, pp. 273-4).} The approach involves analyzing the topological features of the electron density distribution of a molecule. These features are then carefully examined and linked to a variety of chemical concepts. The electron density distribution, usually labeled \( \rho \) (or \( \rho(r) \) where \( r \) represents position and spin coordinates), is a product of the calculation of molecular wave functions (whatever model is employed), as discussed above. It can also be experimentally estimated, most often by means of the scattering of x-rays by molecular electrons in x-ray crystallography.\footnote{See further discussion in section 4 below.} In examining \( \rho \) for a given molecule, the most obvious characteristic is its concentration near atomic centers and low concentration elsewhere. However, a detailed look reveals more features. Since one can treat \( \rho \) as a scalar field in three-dimensional space, one can proceed to examine the gradient vector field associated with \( \rho \) (by applying the vector differential operator \( \nabla \)): this shows the direction in which the
density is increasing the most at a given point (and the magnitude of the increase). In this way, one can find features such as critical points associated with extrema (minima, maxima and saddle points), as well as gradient paths—trajectories that follow the line of steepest “ascent” at successive points.\footnote{Gradient paths are thus everywhere orthogonal to contour lines of equal electron density. Note also that further details can be found by examining the second differential operator or Laplacian $\nabla^2(\rho)$. The Laplacian is interpreted as indicating local concentration and depletion of charge density, and can be used to identify (imperfect) analogues of localized electron pairs (see Bader, 1990, Ch. 7, Popelier, 2000, Ch. 8, and Gillespie & Popelier, 2001, Ch. 7).}

In examining $\rho$ for a given molecule, a set of gradient paths originating at infinity will converge on maxima associated with each nucleus.\footnote{Technically, the maxima associated with nuclei are not true critical points due to discontinuities, but AIM proceeds to treat them as such as a practical matter (see Bader, 1990, p. 19).} According to AIM, the space traversed by all of these paths (called the atomic basin), along with the nuclear “attractor” itself, define an individual atom: “An atom, free or bound, is defined as the union of an attractor and its associated basin” (Bader, 1990, 28, emphasis original).\footnote{Bader insists this topological definition of an atom also identifies a true quantum physical atom (see discussion in section 3).} There is also a critical point (a saddle point) between nuclei: the set of gradient paths originating at infinity and converging on these points define a boundary, called the interatomic surface.\footnote{This surface is also referred to as a zero-flux surface, in that no gradient vectors of charge density cross it at any point (see Bader, 1990, pp. 28-29).} The atom is bounded inside the molecule by this surface and extends to infinity in the open directions away from the rest of the molecule: in practice the boundary in these directions may be defined using a pragmatic cut-off level of electron density.\footnote{0.001 a.u. is a practical estimate of the van der Waals envelope around the molecule (Popelier, 2000, p. 43).} Next, one can observe gradient paths that mark out lines of concentrated density linking two atomic centers to these same inter-nuclear critical points. Such a critical point is called a “bond critical point” and the paths that run from it to the paired nuclei are used...
to define what AIM calls “bond paths;” the full set of bond paths comprises what is called the molecular graph (Bader, 1990, 32-3).228

What is the relationship between AIM’s bond path and other notions of the chemical bond? In their exposition, Gillespie and Popelier caution that “a bond path is not identical to a bond in the sense used by Lewis (2001, 152).” A molecular graph will not be identical to a Lewis structure, for instance, because “double and triple bonds are represented by only one bond path (152).”229 Still, they assert that “the existence of a bond path between two atoms tells us that these atoms are bonded together (153).” Given this claim, one might ask if AIM theory is a way to “hold on to the structural conception of the bond understood functionally (Weisberg, Needham, and Hendry, 2016, sec. 4.3).”

On this point, it is important to immediately note that the AIM definition of a bond path includes a stipulation that draws on the energetic conception of bonding. The identification of a bond path includes not only the presence of the signature pattern of electron density, but also requires specifically that “the forces on the nuclei are balanced and the system possesses a minimum energy equilibrium internuclear separation (Bader, 1990, 33).” Otherwise the feature is referred to as an “atomic interaction line (32).” But there is no reason a structural approach to understanding bonding cannot include this energetic component, and the AIM approach does at first appear to include important elements of the structural conception (as defined in Weisberg, 2008): bond paths map a directional, submolecular relationship between atomic centers. With regard to the Weisberg’s last criterion, Bader at times seems to endorse the notion that this feature “is

228 The bond critical point is the minimum on the gradient paths connecting to the nuclei (the bond path), but is a maximum in the perpendicular directions along the interatomic surface.
229 Part of the AIM approach is to look closely at the characteristics of the BCP’s and neighboring topology to show how they correspond with various types of bonds.
responsible for holding the atoms together” (Weisberg, 2008, 934). He says “nuclei…are linked by a line through space along which electronic charge density, the glue of chemistry, is maximally accumulated (Bader, 1990, 33).” But is this line of concentrated electronic charge density literally the “glue”? While the issues here are subtle, the answer is no.

Bader says that the appearance of an atomic interaction line (AIL), associated with an accumulation of charge between a pair of nuclei, is a “necessary condition” if two atoms are to be bonded to one another, but its presence is also a sufficient condition only when “the system possesses a minimum energy equilibrium internuclear separation” (Bader, 1990, 33). It is then that the AIL is designated a bond path. In order to better understand this definition, the role of the accumulation of charge between two nuclei in the process of achieving bonding and the specific meaning of the presence of a bond path in equilibrium need to be carefully distinguished.

If one pictures atoms at a greater-than-equilibrium distance being brought closer together, then bonding is “the situation obtained when the initially attractive Hellmann-Feynman forces acting on the approaching nuclei, and resulting from the accumulation of electron density associated with the formation of the atomic interaction line, vanish…” (Bader, 1998, 7314).” Bader is referring in the quote to the role of electrostatic forces, the reliance on which is justified by reference to the Hellman-Feynman theorem. This theorem, which dates to the 1930’s, implies that, given a wave function (and associated $\rho$), all the forces on a nucleus in a molecule can be calculated based on classical

\[ \text{\footnotesize 230} \] The notion that a region of electron density provides the “glue” or “cement” holding atoms together in a molecule is widespread in chemical texts, presumably for its heuristic value in some contexts (e.g. Loudon, 1995, p. 36; Shusterman & Shusterman, 1997, p. 774). This provides the backdrop for Bader’s comment. \[ \text{\footnotesize 231} \] Popelier (2000) includes a discussion of this distinction (pp. 60-61).
electrostatics. As shown by Berlin (1951), this result can be used to identify so-called binding and anti-binding regions of electron density in molecules. In the case of a diatomic system, this analysis describes the electronic charge in the binding region as something that draws the nuclei together, while electron density in anti-binding regions on the far side of the nuclei works to draw them apart (along with nuclear repulsion).

Outside the equilibrium inter-nuclear separation, we can ascribe to the binding region the responsibility for a (net) attractive force. At equilibrium distances, all forces on the nuclei are balanced. Now, one can still divide the electronic density distribution into binding and anti-binding regions (it should be noted here that the binding region encompasses much more than the line of density marked out by the bond path). But if there is no net force at work, it would be an oversimplification to say the binding region is “holding the atoms together.” To assert this would only tell part of the story, since one could say it is the anti-binding region that also holds the atoms in place by keeping the inter-nuclear distance from compressing beyond the equilibrium separation. This point is perhaps clearer in comparison to Hendry’s formulation of bonds as “material parts of the molecule that are responsible for spatially localized sub-molecular relationships between individual atomic centers.” Binding and anti-binding regions both clearly play a role in defining the equilibrium inter-nuclear distance. And since these regions combined

---

232 The Hellmann-Feynman theorem generally describes how the wave function and energy of a system changes with respect to a change in a parameter that appears in the Hamiltonian (see Atkins & Friedman, 1997, pp. 182-183). When the parameter in question is the nuclear position coordinate, one obtains the result that the force on a nucleus “is the sum of the Coulombic forces exerted by the other nuclei and by the electron density distribution ρ (Gillespie & Popelier, 2001, 134).” See also Bader (1990), pp. 315-322.

233 This is discussed by Gillespie & Popelier (2001), pp.134-36. See also Bader (1964). Note that this work long pre-dates the AIM theory.

234 This electrostatic picture of how bonding is achieved as atoms are brought together has been challenged by a competing theory of the bonding process. As described recently by Needham (2014), primarily referencing the work of Ruedenberg and colleagues, this alternative uses a variational analysis to infer that changes in electronic kinetic energy play an important role. Bader has responded to this line of theorizing (Bader, 2011, pp. 27-29).
include the entire molecule, it would seem there is no basis for concluding that the bond
paths of AIM theory, despite highlighting a concentrated area of ρ, serve to pick out a
sub-molecular region to play the functional role envisioned by the structural conception.
Instead, the presence of the bond path at equilibrium appears to represent the “attractive
restoring forces [which] act on the nuclei for any displacement from their final
equilibrium position (Bader, 1998, 7314-15, emphasis added).” Bader goes on to say the
bond path is a “universal indicator of bonding (Bader, 1998, 7315, emphasis added).” One
must conclude that, according to AIM, the bond path is a sign that the bonding
relationship exists, but it does not directly represent a particular region of density
responsible for holding the atoms in place.

Consistent with this conclusion is one other departure AIM takes from a
traditional view of chemical bonds. Bonds are often pictured as spatially localized
between the bonded atoms in a molecule. However, bond paths (with the exception of the
bond critical point) fall inside the boundary of the atoms as they are defined in AIM. The
atoms lie adjacent to one another along the interatomic surface. To the extent a
conception of chemical bonds envisions a clear separation between those bonds and the
bonded atoms, this is not provided. AIM offers a different picture that “requires the
replacement of the model of structure that imparts an existence to a bond separate from
the atoms it links – the ball and stick model or its orbital equivalents of atomic and
overlap contributions – with the concept of bonding between atoms (Bader, 1998, 7322,
emphasis original).”

---

235 Popelier discusses his and Bader’s (somewhat confusing) use of the word “attractive” to describe the
restorative forces at work in any displacement from equilibrium separation: it is to be distinguished from
“attracting” (net) forces that bring two atoms together (Popelier, 2000, pp. 55-56).
236 Also note that even in the above discussion of binding in terms of electrostatic forces, the forces apply to
nuclei, not atoms.
3. Interpreting the AIM Conception of Atoms and Bonding

Bader’s distinction between “bonds” and “bonding” is in keeping with a difference in perspectives Hendry identified between the structural and energetic conceptions.\(^\text{237}\) Given this, and given its reliance on energy equilibrium in its definition of bonding, one might ask how closely AIM should be identified with the energetic conception. The key difference is that while the energetic conception puts the emphasis on achieving a stabilizing minimum molecular energy, AIM goes further and identifies bond paths as the indicator that particular atoms are indeed bonded. Only some pairs of atoms in a polyatomic molecule at an energy minimum are bonded, and according to the theory, bond paths pick these out. As a first pass, it may seem helpful to view the AIM conception of bonding as a hybrid combining features of the conceptions considered above: while it relies on energetic considerations, its bond paths define directional relationships between atomic centers in keeping with the structural conception.\(^\text{238}\)

But viewing the AIM idea as a hybrid in this way arguably misses what makes it distinctive. The goal of AIM is to provide not only a conception of bonding, but also of

\(^{237}\) Bader is consistent in using this characterization, stressing in his 1990 book that “a bond path is not to be understood as representing a ‘bond,’” rather “the presence of a bond path linking a pair of nuclei implies that the corresponding atoms are bonded to one another (Bader, 1990, 35)”. In a 2009 article, he objects to critics who interpret his intention as that of identifying the bond path with a traditional concept of a bond, even taking them to task for failing to note that “bond” is a noun, while “bonded” is a participial adjective (Bader, 2009, p. 10391). It would seem there is room for some misunderstanding, given that “bond path” features a noun, and the visual representations of bond paths and molecular graphs are suggestive of traditional structural diagrams.

\(^{238}\) On the other hand, it should be emphasized that because bond paths don’t fulfill the envisioned functional role, AIM is consistent with the conclusion that the structural conception as defined above fails to survive the transition to a quantum physical understanding of chemistry.
the bonded \textit{atoms} it seeks to develop the idea that atoms in molecules should be seen as bona fide physical systems in their own right. Bader notes that “quantum mechanics has been shown to account for the properties of isolated atoms and for the total properties of a molecular system” but there is a “lack of a quantum definition of an atom in a molecule (Bader, 1990 131).” The approximated solutions to the molecular wave function feature delocalized electron orbitals around a configuration of stationary nuclei. As discussed, AIM uses a topological examination of $\rho$ to define atoms. In doing so, AIM also provides a way to calculate a number of properties of these atoms. To calculate atomic charge, for example, one integrates $\rho$ over the topologically defined volume of an atom and then subtracts it from the associated nuclear charge: Bader argues that the consistency of these calculated values across molecules that incorporate the same atom demonstrates the success of the approach.\textsuperscript{239} AIM extends this approach to other atomic properties, including atomic energies, although this involves more complex derivations.\textsuperscript{240} Bader’s ultimate claim is that AIM provides a full account of atoms in molecules as quantum physical subsystems: the topologically defined atom is also a quantum atom.\textsuperscript{241} The theory’s success in establishing this claim continues to be the subject of debate in the theoretical chemistry literature, and no definitive judgments on its technical merits can be

\textsuperscript{239} Bader uses Li in LiF, LiO, and LiH as an example (Bader, 1990, p. 135). Gillespie and Popelier (2001) also argue for the usefulness of these calculations (pp. 153-4).

\textsuperscript{240} Popelier (2000, 2016) gives a concise account of AIM’s derivation of atomic energies. In response to some criticism of the approach (e.g. Anderson, Ayers, & Hernandez, 2010), Popelier concedes that it is not ruled out that some molecular fragments, which are not AIM’s topological atoms, may also have a well-defined energy. As a result, he asserts that “all topological atoms are quantum atoms, but not all quantum atoms are topological atoms (Popelier, 2016, 37).” One of the limitations of AIM’s approach to energy calculations is that it only applies at equilibrium. Alternative methods of partitioning energy which relax this assumption are discussed by Popelier (2016, p. 38). Blanco, Pendas & Francisco (2005) develop an approach (labeled “interacting quantum atoms” or IQA) that allows for partitioning into both intra and inter-atomic energies. Other work toward extending the AIM research program is described by Shahbazian (2013a).

\textsuperscript{241} Bader’s arguments that quantum mechanical principles apply to AIM’s atoms is given in Bader (1990), Chs. 5, 6 & 8. An important part of these is a derivation based on Schwinger’s principle of least action (Ch.8).
made here. Rather, with this sense of the goals of the program, we can return to the question of what AIM’s approach implies for the conceptions of atoms and bonding.

Instead of simply describing a molecule as a system featuring interactions between electrons and the various nuclei, AIM posits atoms as interacting systems within the molecule. An atom in a molecule is an “open quantum subsystem, free to exchange charge and momentum with its environment (Bader, 1990, 169).” Of course the relationship of interest is between two bonded atoms along an interatomic surface: “it is through the exchange of electrons and the fluxes in properties across the surface described by the physics of a proper open system that atoms adjust to the presence of their bonded neighbors (Bader, 1998, 7322).” Bonding is a special physical relationship between pairs of atoms in a molecule where displacement (within limits) leads to particular restorative responses within the molecular framework. In equilibrium, the nature of this relationship can be examined by looking at the topological properties of electron density at the point where the atoms meet and where charge or other properties would be exchanged – the bond critical points on the interatomic surfaces. But even though bond paths are defined in terms of equilibrium, I would argue that the distinctive feature of the conception is that it embodies the idea of a particular pairwise interaction between atoms (not just between the electrons and nuclei). Rather than a combination

242 There are challenges to AIM in the chemistry literature that claim that the association of bond paths/BCP’s with stabilizing interactions is flawed based on an examination of cases (e.g. Poater, Solà & Bickelhaupt, 2006). The present discussion is limited to arguing that this notion is conceptually central to the theory.

243 Pendás, Francisco, Blanco & Gatti (2007), extending the IQA analysis mentioned in footnote 240, use their approach to analyze the nature of BP-signal bonding. Taking advantage of the fact that their approach treats interatomic energies explicitly compared to the manner of the original AIM theory, they conclude that bond paths “signal the existence of preferred or privileged exchange channels in molecules (Pendás, Francisco, Blanco & Gatti, 2007, 9370).”
of the energetic and structural conceptions discussed above, the AIM conception can be better labeled an *interactive* conception of bonding.

Before moving on, it should be noted that AIM’s claim of a close match between its definition of bonding and more traditional chemical definitions has been challenged. Bader had put the claim this way: “the network of bond paths…is found to coincide with the network generated by linking together those pairs of atoms that are assumed to be bonded to one another on the basis of chemical considerations (Bader, 1990, 33).” However, Weisberg, Needham and Hendry note that Bader’s approach appears to be “too permissive:” citing Cerpa, Krapp, Vela & Merino (2008), the problem is that bond paths occur in situations that involve atoms which are not considered to be chemically bonded (one example given was that of an Argon atom trapped within a C₆₀ molecule which features a bond path connecting it with all sixty carbon atoms).²⁴⁴ In his more recent papers, Bader does clearly acknowledge that bond paths are present in contexts that are not traditionally associated with chemical bonds. However, he attempts to turn this into a virtue: AIM theory offers a more theoretically precise approach to bonding that is extends beyond traditional notions but also offers analytic tools to more precisely characterize different cases (bonding in molecules vs. e.g., van de Waals interactions).²⁴⁵ On balance, while the criticism has some merit given claims made in Bader’s earlier work, it must be

---

²⁴⁴ See Weisberg, Needham & Hendry (2011, sec. 4.3). Bader’s response to this example is in his (2009). Another criticism was recently offered by Foroutan-Nejad, Shahbazian & Marek (2014), who emphasize the fact that bond paths may disappear/re-appear due solely to nuclear vibrations in some cases. They also note that atoms in molecules that are not linked by bond paths may also be seen as interacting, and AIM’s indirect manner of characterizing the differing nature of these interactions leaves the theory open to “controversies” (p. 10149).

²⁴⁵ See Bader (2009) and (2011): “This definition… transcends all bonding schemes and categories and provides a unified physical understanding of atomic interactions (Bader, 2011, 20).” Note Bader died in 2012.
considered in the larger context of debates about chemical bonds: neither Lewis’s theory nor any successor account of bonding is free of challenges.\textsuperscript{246}

4. The Question of Spatial Localization

The AIM conception has advantages over the others discussed: it offers more detail about how atoms relate to one another inside a molecule compared to the energetic conception, and unlike traditional structural approaches to bonding it relies only on information drawn from quantum mechanically derived calculations. It also invites one to consider that bonding is not best understood in static terms. While the analysis is based on an idealized model of an isolated system in equilibrium, the conception suggests that molecules are better understood as constituted from patterns of repeated characteristic interactions between atoms in an ever changing environment. This picture has a clear affinity with Chapter 4’s proposal that nature’s causal web contains a hierarchy of composite processes that form from repeated interactions of a constituent group of sub-processes.

There is a remaining challenge to AIM’s interactive conception of bonding, however. It is one shared with the structural conception. Recall that in Hendry’s definition bonds “are responsible for spatially localized submolecular relationships (Hendry, 2008, 917, emphasis added).” By analyzing the topology of the electron density distribution, AIM finds a basis for bonding as a special interaction relationship between

\textsuperscript{246} This is true even of a minimalist energetic account. For example, Berson (2008) argues there are cases where covalent bonding leads to energetic destabilization.
atoms in a molecule. But the question now is whether this approach justifies conceiving of this interaction as one taking place in familiar three-dimensional space.

As discussed, much of modern chemistry begins with the calculation of molecular wave functions and associated energies. As also mentioned in section 1, the wave function does not represent something residing in our familiar space, but rather is defined on a multi-dimensional configuration space. The wave function can also be described as a vector, and the most general representation of the state of a quantum system is as a vector in an (infinite dimension, complex-valued) Hilbert space. According to Bader, this abstract representation is an obstacle for understanding molecules. Instead a description is needed “of the system in real space (Bader, 1990, 10).”

Atoms and bonds have meaning in real space and are a reflection of the structure present in real space. This structure is not reflected in the properties of the infinite-dimensional Hilbert space of the wave function (Bader, 1990, 132)

The proposed solution is to rely on the electron density distribution:

The charge density provides a description of the distribution of charge throughout real space and is the bridge between the concept of state functions in Hilbert space and the physical model of matter in real space. (Bader, 1990, 169)

The driving intuition behind AIM’s topological approach is that “atoms exist in real space and they are defined by a partitioning of real space (Bader, 1990, 55).” But how is it that \( \rho \) represents entities in “real space”, when it is straightforwardly derived from the wave
function $\Psi$. Here it is helpful to review the statistical interpretation of the wave function and the derivation of $\rho$.

In quantum mechanics, the probability of finding a particle in a given volume of space in a measurement is calculated from the absolute value (or modulus) squared of the appropriate wave function, or, equivalently, the product of the wave function and its complex conjugate (denoted by $\psi^*$). For an atom with a single electron (in the time-independent case), the probability of finding the electron in given infinitesimal volume of space $r + dr$ is given by:

$$\psi \ast \psi dr$$

(where $r$ represents the three space coordinates with the origin on the stationary nucleus as well as the spin coordinate). For a multi-electron molecular system (given a certain arrangement of nuclei), the similar expression:

$$\psi \ast (1, 2, ..., N)\psi(1, 2, ..., N)dr_i$$

gives the probability of finding the first electron between $r_1 + dr_1$, and the second electron between $r_2 + dr_2$, ..., and the $N$th electron between $r_N$ and $dr_N$. If we are interested only in

---

247 One may also estimate $\rho$ based on results from a scattering experiment. This is discussed immediately below.

248 The notation in this discussion follows that of Veszprémi & Fehér (1999). For Bader’s equivalent presentation, see Bader (1990), p. 6.
finding the probability of finding a particular electron in a particular location, then one must integrate over all possible positions for the rest of the electrons:

\[ p_1 r_1 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} ... \int_{-\infty}^{\infty} \psi^* (1, 2, ..., N) \psi (1, 2, ..., N) \, dr_2 \, dr_3 ... \, dr_N \]

but since electrons are indistinguishable, the probability of finding any electron in the same volume is the same. So by multiplying this value by the number of electrons in the system we can find the probability for finding any of the \( N \) electrons at a particular location.

\[ \rho (r_1) = N \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} ... \int_{-\infty}^{\infty} \psi^* (1, 2, ..., N) \psi (1, 2, ..., N) \, dr_2 \, dr_3 ... \, dr_N \]

In this way, the electron density \( \rho (r) \) can be defined at each point.

The fact that quantum mechanics offers a way to probabilistically predict measurement outcomes (for observables such as position, momentum and spin) is at the core of what makes it a successful physical theory. The calculation of \( \rho \) just outlined is valuable for the same reason. In the context of quantum chemistry, of course, much information and predictive power is derived from approximated wave functions (orbital

---

\(^{249}\) This is not a probability density function, since it sums to \( N \) rather than one. If one integrates the expression over the entire space, one will recover the number of electrons, as expected.
structures) and calculated energies without utilizing this specific link to measurement of position or other properties of basic quantum particles. This aspect of quantum theory tends to be left in the background: in employing visual depictions of atoms and molecules, texts sometimes go back and forth between representations of wave functions/orbitals on the one hand and electron densities on the other without making much of the distinction.\footnote{See Gillespie & Popelier (2001), p. 62.} This practical attitude on the part of chemists is understandable, but given that the AIM approach claims to have a rigorous quantum basis for its view of bonding, the statistical interpretation given to $\rho$ appears to be an obstacle.

As discussed, Bader asserts that $\rho$ can be used to define a model of the molecular system in “real” space, but real space only comes into play in the statistical interpretation when we measure the position of a particle using our experimental apparatus. In the present context, it is electrons that are localized in space via a measurement interaction.

We do not find \textit{density} at particular locations in space, rather we find \textit{electrons}.\footnote{Bader and Zou (1992) discuss how $\rho$ can be linked to an operator, $\hat{\rho}$, that meets Dirac’s definition of a quantum mechanical observable (pp. 43-55). This operator is defined as a function of the position operator: $\hat{\rho}(r) = \delta(\vec{r} - \vec{r})$. Its eigenstates are the same as the position operator, but with eigenvalues that are delta functions at each coordinate. The expectation value for this operator will indeed be the density at each position. This result does not alter the argument of the main text, however, because even if $\hat{\rho}$ is an observable, its expectation value is not something that is observed. As always, the expectation value is a probabilistic assessment of the outcome of a particular measurement. Bader (2009) also defines a “bond path operator” based on this density operator, and the present discussion applies to this notion as well, since the bond path itself is again the expectation value for the operator.} According to the statistical interpretation, when one begins with a calculated wave function for a molecule and then derives the electron density distribution, the context of possible measurement interactions is presumed. The electron density itself is not interpreted as describing something that exists in space.
Bader is proposing, then, an alternative interpretation of $\rho$:

The charge density...is a representation of the time-independent distribution of negative charge throughout three-dimensional space and this is an alternative and equally valid interpretation of the quantity $[\rho]$. (Bader, 1990, 7)

Apart from references to Schrödinger’s endorsement of a similar interpretation in an early paper,$^{252}$ the key support Bader offers for the alternative interpretation is from experimental investigation of molecular structure. Whether this in fact provides support for the view, however, is debatable.

As mentioned, in addition to ab initio calculations, $\rho$ is estimated experimentally. X-ray diffraction, in particular, is used for this purpose. For this reason, sometimes $\rho$ is said to be “measurable” or something that can be “observed”?$^{253}$ However, it may be more accurate to say that the form of $\rho$ is estimated or reconstructed.$^{254}$ In a QM description of X-ray diffraction featuring coherent or elastic interactions (meaning the energy states of the particles are not altered), there is a mathematical relationship between $\rho$ and the amplitudes of scattered beams given certain simplifying assumptions (for additional details see the appendix). Because the expression for this relationship has the form of a Fourier transform, one can turn around and look to estimate $\rho$ from the measured

---

$^{252}$ See Bader (2005), pp. 821-822 and Bader (2010), pp. 2-3. It may be that the use of classical EM physics as licensed by the Hellman-Feynman theorem plays a role in inspiring this interpretation.

$^{253}$ “The charge density $\rho(r)$ is the fundamental property measured in a coherent X-ray scattering experiment (Bader, 1990, 8).” Elsewhere authors have referred to electron density as “a real, measurable property (Veszprémi & Fehér, 1999, 185),” or a “measurable quantity (Bader & Matta, 2013, 255).” Gillespie and Popelier call electron density “a real observable property (Gillespie & Popelier, 2001, 82),” while Matta and Gillespie say it can be “observed by electron diffraction and X-ray crystallography (Matta & Gillespie, 2002, 1141).” In addition to the X-ray diffraction case discussed in the text, I note that Bader has also claimed that his interpretation is consistent with results from atomic force microscopy (Bader & Matta, 2013). Hettema (2013) has responded by asserting that alternative explanations are possible (p. 319).

$^{254}$ The latter term is used by Tsirelson & Ozerov (1996), Ch. 3.
intensities of coherent X-rays reflected in a diffraction experiment. As a practical matter, the process utilizes a crystal formed from molecules of interest (X-ray crystallography). The mathematical form of \( \rho \) is extracted from the information from rays scattered from the various electrons in the crystal. Computer generated images based on these calculations are used to display detailed shapes of molecules.

Bader asserts that this ability to estimate \( \rho \) from coherent scattering supports an alternative to the statistical interpretation: “There is no need here for a probabilistic interpretation of the results. These experiments may be interpreted as providing a measurement of the spatial distribution of charge (Bader, 1990, 7).” In a later paper, he writes: “The probabilistic interpretation is inconsistent with the measurement of the electron density by the elastic scattering of X-rays from crystals, since no excited states are involved in the scattering process, as required by the Born probabilistic postulate (Bader, 2011, 16).” He is drawing a contrast here with a hypothetical inelastic scattering experiment that would locate the position of an electron in a small volume of space at the expense of changing its momentum/energy (with the relationship subject to the uncertainty relation). However, there appears to be no clear basis for concluding that the experimental estimation of \( \rho \) justifies such an alternative interpretation.

In thinking about X-ray diffraction experiments there may be a temptation to picture the rays being scattered by the electron density as if the latter was akin to a classical object or field. But recall that the reason the form of \( \rho \) can be estimated is because it appears in the context of a formal QM description of the interacting system of photons (in the X-ray beam) and electrons (the \( N \) electrons in the molecular system). The building block for this description is the case of a single photon interacting with a motionless one-electron atom: here a wave function for the composite system can be
expressed as a sum of terms whose coefficients, when squared, are proportional to the probability of the photon scattering in various directions. Note that the connection between the wave function and a probabilistic expectation for outcomes is present, as usual. Now, even in this simpler case, the ground state atomic electron probability density appears in the term associated with elastic scattering possibilities. But the physical system being described remains that of an interacting (quantum mechanical) photon and electron. And just as in the many photon/N-electron case, photons are scattered by electrons, not the electron density. The fact that the electrons are not localized in elastic scattering events would seem to have no bearing on the interpretation of ρ.

In summary, while ρ can be estimated via experiment, this does not offer clear support for the conclusion that it is represents an entity that exists in space. The alternative interpretation Bader puts forward appears to need further justification. This is especially apparent given that accounting for the emergence of classical-like chemical properties, such as structure or space in shape, from a quantum mechanical basis is widely viewed as a tremendous challenge. Bader (2011) goes on to claim that AIM indeed provides the path to a successful reduction of chemistry to physics by both having

---

256 I note that Hettema (2013) includes a brief objection to the idea that ρ can be considered an observable, saying “electronic densities are not directly empirically accessible, but rather correspond to observables that are the result of a more convoluted operational procedure (318-319).” In a reply to Hettema, Shahbazian counters that ρ shares with other quantum observables such as spin the characteristic of needing complicated experimental procedures to measure (Shahbazian, 2013b, p. 332). However, for present purposes this exchange does not address the key point. Many objects are difficult to observe, and scientists employ complex indirect methods to do so (this is true for both microscopic and macroscopic cases – the use of gravitational lensing in astronomy is an example). But the transition from a quantum description of a system in Hilbert space to locations and structures in familiar three dimensional space is a distinctive and controversial element in discussions of chemistry’s foundation in physics. So the concern is not the complex indirect estimation techniques involved in estimating ρ, the objection is to any inference that this estimation confers a status as something present in space.
257 For an argument that canonical QM cannot support the concept of (classical) molecular structure, see Primas (1981).
a rigorous foundation in (canonical) QM while also providing new derived elements that successfully ground chemical concepts. While I will not review Bader’s claim at length in the context of the reduction debate,\textsuperscript{258} I do argue that this ambition rests on a problematic foundation: the assumption that $\rho$ is a representation of an entity in real space.

5. An Alternative Interpretation of $\rho$

I will suggest a way to overcome the challenge to the AIM conception summarized in the last section. To set the stage, it is helpful to review the situation in the context of scientific modelling, since AIM uses $\rho$ as a model to explore the characteristics of molecules. Treating $\rho$ as a scalar field and examining its features via the associated gradient vector field, the intent is to associate these features with aspects of the molecular system, including constituent atoms and their bonding relationships. Using Weisberg’s (2013) framework, a model consists of a structure and its interpretation. The structure in this case is the scalar/vector fields. The interpretation or construal describes how the structure relates to the target system in the world. The problem with the AIM construal identified in section 4 is its assumption that the target system is an entity in 3D space. This assumption is problematic given the standard statistical interpretation, and this calls into question the specific assignment of features of the model to those of the target.

There may be a different way to interpret the relation of the model to the target system, however, that puts the AIM picture of bonding on a sounder conceptual footing.

\textsuperscript{258} See Hendry (2010) for a critical examination of the prospects for inter-theoretic and ontological reduction. For a discussion of AIM in the context of Bader’s reduction claim, see Hettema (2013).
The cost of taking this approach lies in the view one must adopt regarding the quantum measurement problem and “wave function collapse.” As discussed, the standard statistical interpretation sees \( \rho \) as giving probabilistic information about the outcomes of measurements: this implies an experimental setting. But it is certainly possible to speculate that such “measurement-type” outcomes, resulting in spatial localization, take place in nature on an ongoing basis with no need of human intervention. Possibly such localization events are triggered by a variety of interactions involving external systems.\(^{259}\)

The upshot is that instead of describing systems using wave functions that evolve strictly according to the Schrödinger equation, one now assumes that in nature this evolution is punctuated by wave function collapse and position localization. With this alternative picture in mind, the idea is then to construe \( \rho \) as a model of a different target system. Instead of being a model representing an entity that exists in space at a given time, one instead sees \( \rho \) as an idealized representation of the positions where electrons are likely to intermittently and periodically localize as the system evolves. It depicts a time-extended pattern of localization events. If these events are indeed associated with particular physical interactions, then it follows that the pattern can be interpreted as a summarized causal history of (at least some of) the interactions involving the molecular system.

\(^{259}\) It should be noted that this suggestion has some similarities to other discussions of how molecular shape can be accounted for given QM. In several articles, Woolley has argued that new ideas are needed to understand the genesis of structures for polyatomic molecules given the possibility of isomerism (see, e.g., Woolley, 1991, 1998 and Sutcliffe & Woolley, 2012). Woolley suggests that perhaps environmental interactions play a role in the emergence of classical-like molecular structures from a quantum mechanical basis (Woolley, 1991, p. 42). Also, Ramsey (1997) emphasizes the link between molecular shape and experimental contexts. But in trying to develop these ideas, one is typically led back to the problem of interpreting or modifying quantum mechanics. Scerri (2011, 2013) has recently proposed that decoherence theory offers the prospect for addressing the problem Woolley has emphasized. However, it appears to be accepted in QM interpretation debates that decoherence theory on its own doesn’t provide true measurement-like outcomes (see Bacciagaluppi, 2016). Also, see Fortin, Lombardi, Camilo & González (2016) for a discussion of the specific problem of optical isomerism and its link to the problem of quantum measurement.
In the last chapter I proposed that causal interactions involve spatial localization, and that in QM explanations, measurements correspond to these interactions. Here, I suggest extending this picture by positing that measurement-style localized interactions take place throughout nature. Among extant approaches to interpreting QM, this idea has most in common with dynamic collapse theories, such as GRW (from Ghirardi, Rimini & Weber, 1986). As discussed in Chapter 5, section 4.1, the GRW approach modifies QM such that an isolated microscopic system may evolve without collapse for a very long time, while macro-sized composite systems will collapse extremely quickly (a “hit” to any one constituent spreads to all of those in the composite). The values of the constants chosen by the theorists (governing localization resolution or “width” and the frequency of hits) allow for predictions in line with experimental findings. As noted in the last chapter, one difference is that collapses in the GRW theory do not have a necessary link to interactions as assumed in the causal theory, although in practice the creation of larger entangled systems through interaction in the GRW picture increases the likelihood of collapse. In the case of an interaction of a micro-system with a macroscopic measuring instrument with multiple possible settings, a collapse into one of the possible positions can expected to ensue with the appropriate probabilities (see Ghirardi, 2016). Returning to the question of the AIM conception, I want to argue that even if the precise process whereby localization occurs dynamically is unknown, positing that collapses do occur in nature allows for an attractive alternative interpretation of $\rho$ and its representation of a molecular system.

---

260 Recall that in this scheme, Schrödinger evolution corresponds to propagation of causal processes between these interactions.

261 The implication is that collapses occur in the position basis. This is another commitment of GRW-type theories: the outcome of any measurement of non-position properties we observe is always via the position of some macro-sized detection device which has established a correlation with the measured system.
Before proceeding to detail this interpretation, one further challenge for this line of thinking should be highlighted. This is the potential that the dynamical collapse process proposed has some disruptive impact on the molecule. This is contrast to the usual implicit (idealized) model assumption that the system retains the same character over time. It is worth noting that the spatial resolution in the GRW framework (approximately $10^{-7}$ m) is chosen to be substantially larger than the size of atoms and molecules in part to avoid this concern.\textsuperscript{262} At the molecular/sub-molecular level, it would seem that any localization would have to be transitory in a nature if it is to avoid undue disruption. Despite the challenge this issue presents to any detailed theory of collapse at the micro-level, I will assume in what follows that localization does occur intermittently and periodically at the level of electrons. With this this assumption in place, it is possible to sketch how this new picture of the target system represented by $\rho$ can support the interactive conception of bonding.

Beginning with an isolated atom, we interpret the mathematical form of the electron density distribution as encoding a pattern of positions where the electrons will likely periodically localize.\textsuperscript{263} When an atom joins others to form a molecule, it forms a new composite system. Generally, the key to identifying the atom as a subsystem in the new composite lies with the degree to which its signature pattern of electron localization is maintained. As long as the interactions between the atom and other subsystems do not unduly alter the pattern that characterizes the isolated atom, the atomic-level properties are conserved. In the formation of molecules, the atomic electrons will physically interact

\textsuperscript{262} "The fact that localization width [in the original GRW model] is large compared to the dimensions of atoms (so that even when a localization occurs it does very little violence to the internal economy of an atom) plays an important role in guaranteeing that no violations of well-tested quantum mechanical predictions is implied by the modified dynamics (Ghirardi, 2016, Sec. 5)."

\textsuperscript{263} Note that any simplifying or idealizing approximations used in the calculation of a wave function for an atom or molecule carry over to the electron density distribution.
A functional group may correspond to an intermediate-level pattern in a similar way.

It is unclear whether this interpretation of \( \rho \) creates a problem for the formal aspects of the AIM theory: one concern is whether it is still appropriate for the interatomic surface to be physically interpreted as a zero-flux boundary (as required for the quantum mechanical derivations in the full AIM theory).

6. Conclusion

The AIM theory offers an attractive, interactive conception of chemical bonding, using a quantum model of the molecule at equilibrium to identify a special interaction relationship between constituent atoms. Given a more realistic picture of continual environmentally-induced fluctuations around equilibrium, the conception suggests that bonding is achieved via repeated characteristic interaction patterns between atoms. The resulting view fits well with the causal theory presented in Chapter 4: molecules and their bonded constituent atoms map onto the composite causal processes and constituent sub-processes of the theory.

---

264 A functional group may correspond to an intermediate-level pattern in a similar way.
265 It is unclear whether this interpretation of \( \rho \) creates a problem for the formal aspects of the AIM theory: one concern is whether it is still appropriate for the interatomic surface to be physically interpreted as a zero-flux boundary (as required for the quantum mechanical derivations in the full AIM theory).
The question of whether bonding relationships can be interpreted as spatial presents a stumbling block for the AIM approach, however. This is due to the lack of spatial localization in quantum mechanical descriptions of molecular systems in absence of measurement. This problem might be addressed by embedding the AIM conception of atomic interactions within a larger interpretative framework which sees wave function collapse and spatial localization as a real, ubiquitous natural phenomenon. While we lack a widely-accepted physical theory incorporating collapse in this way, positing such a process offers a way to support the use of the electron density distribution as a model for analyzing bonding and other chemical properties of molecular systems. It also is in keeping with the argument for the compatibility of quantum theory and dispositional causal process theory presented in Chapter 5.

Appendix: Estimation of \( \rho \) through X-Ray Diffraction

In an X-ray diffraction scenario featuring coherent or elastic interactions (meaning the energy states of the particles are not altered), there is a mathematical relationship between \( \rho \) and the intensities of scattered beams. To briefly summarize,\(^{266}\) in the general quantum mechanical description of the interaction of incident photons (with a certain propagation vector) with a system of electrons, the interaction Hamiltonian includes a term associated with scattering. From this, an operator can be defined, \( \exp(i\mathbf{K} \cdot \mathbf{r}) \), that when integrated over \( \psi^*\psi \), gives the scattering amplitudes associated with photons with a given propagation vector (here \( \mathbf{K} = \mathbf{k} - \mathbf{k}_0 \), with \( \mathbf{k}_0 \) being the incident vector and \( \mathbf{k} \)

\(^{266}\) The notation here follows Coppins (1997).
the scattering vector of the propagating X-rays).\textsuperscript{267} Assuming scattering events are independent, the interactions in a multi-electron scattering system can be expressed in terms of the sum of one-electron interactions. And given that electrons are indistinguishable, this sum can be restated in terms of the electron density $\rho$.\textsuperscript{268} The intensity of a scattered X-ray beam in the elastic case is thus given by:

$$I(K) = \left| \int \rho(r) \exp(iK \cdot r) \, dr \right|^2$$

6.A.1

The intensity is the modulus squared of the amplitude:

$$A(K) = \int \rho(r) \exp(iK \cdot r) \, dr$$

6.A.2

Notably, this expression has the form of a Fourier transform:

$$A(K) = \hat{f}[\rho(r)]$$

6.A.3

\textsuperscript{267} Here $\psi$ is the electronic wave function, which in the elastic case is unchanged in the interaction. Note that while the overall wave function for the system encompasses both the electronic system and the radiation field, the scattering amplitudes only depend on the electronic wave function.

\textsuperscript{268} In addition to assuming scattering events are independent, this assumes there is no re-scattering (first Born approximation). Other idealizing assumptions include ignoring the following factors: interactions within atoms/molecules, nuclear vibrations, magnetic field effects, relativistic effects, and any deviations from strict polarization and monochromacity in the X-rays (see Coppins, 1997, Ch. 1, and Tsilerov & Ozerov, 1996, Ch.3).
This is an important result, since a function can be recovered from its Fourier transform.

Given this mathematical relationship between amplitudes and $\rho$, one can turn around and look to estimate $\rho$ from the measured intensities of coherent X-rays reflected in a diffraction experiment. As a practical matter, the process utilizes a crystal formed from molecules of interest (X-ray crystallography). The crystal is described in terms of a repeating unit cell in a lattice. Each diffracted ray that arrives at a detector has wave characteristics that can be mathematically described as a Fourier sum of contributions from scattering electrons in defined volume elements of the unit cell—this sum is called the structure factor, denoted $F$. In the limit, one can express the sum by integrating over $\rho$, and the resulting equation is a Fourier transform of $\rho$. Taking advantage of the reverse transform, $\rho$ (for the unit cell, which may contain more than one molecule) is then estimated by summing the structure factors associated with the many diffracted rays:

$$\rho(x, y, z) = \frac{1}{V} \sum_h \sum_k \sum_l F_{hkl} \left[ \exp - 2\pi i (hx + ky + lz) \right]$$

Here $V$ is the volume of the unit cell, and $h$, $k$ and $l$ are indices for the reciprocal space of the reflection. In this way, the mathematical form of $\rho$ is extracted from the information from rays scattered from the various electrons in the crystal. Computer generated images based on these calculations are used to display detailed shapes of molecules.

---

269 Notation here follows Rhodes (2006).
270 A caveat is that the inputs required are amplitude, wavelength and phase of the rays: the first two are derived from intensity and position, while phase is not available. Techniques have been developed to infer
Chapter 7: The Causal Explanatory Character of Evolution Models

There is a complex debate regarding whether and how causation figures into the modern understanding of evolutionary change. There is a consensus that the activities of organisms comprising a population can be characterized causally. But the role of causation is controversial when evolution is depicted in models that represent population structure in terms of features such as trait distribution and trait fitness. Some argue that models of evolutionary change employing these notions (modern synthesis, or MS-models) provide non-causal, statistical explanations of the phenomena. I argue below that MS-models, while employing substantial idealization, do provide causal explanations when used to represent particular real-world outcomes. The key to interpreting these models is a proper account of the causal character of their target system. Here, expanding on Millstein’s proposal to apply Salmon’s causal framework to populations, I show how the theory presented in Chapter 4 properly encompasses both the population and its constituent organisms. The explanatory features of the model can then be shown to map onto causal elements of the evolving populations they represent.

1. Introduction
The statisticalist interpretation of MS-models says that they explain (and predict) changes in population structure, but do not do so by representing the causes of selection or drift.\textsuperscript{271} Walsh, Ariew & Matthen (2017) explain that there are various changes a population can undergo. Some of these changes can be described causally. For instance, if one is interested in changes in lineage structure (resulting from the fact that some organisms reproduce more than others), one might cite the survival and reproductive activities of those organisms (including any emigration/immigration). Such interactions among organisms and between organisms and the environment are considered unproblematically causal in nature (although this causal character might be theoretically cashed out in different ways). The effect of these interactions on population structure may be summarized in terms of individual organism-level (or ecological or “vernacular”) fitness, which is often given a causal, or propensity interpretation.\textsuperscript{272} However, MS-models divide populations into classes according to certain heritable traits. They cite trait distribution and trait fitness (a parameter that corresponds to the growth rate of the trait type in the population) when explaining selection and drift:

By the early 20\textsuperscript{th} century, evolutionary biologists understood that one cannot explain and predict the magnitude and direction of evolutionary change in trait distribution solely from the survival and reproduction of individuals. (Walsh, Ariew & Matten, 2017, 4)

\textsuperscript{271} Walsh, Ariew & Matthen (2017) is primarily used to represent the statisticalist position. Earlier articles arguing for the view include Matthen & Ariew (2002, 2009), Walsh, Lewens, & Ariew (2002), and Walsh (2007, 2010).

\textsuperscript{272} This interpretation is introduced by Beatty & Mills (1979) and Brandon (1978). The view is critiqued by Bouchard & Rosenberg (2004). See Millstein (2016) for a review. While the concept of vernacular fitness will not be discussed at length, some brief comments are included below (section 5).
More information is required for MS-models, for instance, details regarding trait inheritance across generations. But the fundamental issue is that trait fitness as understood in these models (expected growth rate in a trait type) is not generally derivable from vernacular fitnesses of individual organisms bearing the various traits. This is because the defining of trait types “cross classifies the population of individuals—each trait class has many individuals as members, and each individual is a member of many trait classes”.

Per the authors “statisticalism encompasses two positive claims and one negative claim (5).” First, MS-models cite statistical properties of trait types. Second, MS-models explain changes in trait distribution. Third and finally, MS-models do not cite the causes of population change. Trait fitness, in particular, “is not a causal property of a concrete entity (14).” In what follows, I will oppose the third claim: MS-models do cite causes of population change. As argued below, they do so when they are employed to explain actual token instances of selection. Here, so-called statistical properties also play a substantive causal explanatory role. In particular, initial trait distribution and trait fitness values together map onto to a productive causal property of the population.

Together, the first two (positive) claims commit the authors to the existence of a type of statistical, non-causal explanation (Walsh, Ariew & Matten, 2017, p.6). I will not argue against the possibility of this kind of explanation in general, or even in the case of evolution. Much of the power of modeling in evolutionary biology lies in the ability to mathematically investigate possible population structures and their trajectories. These

---

274 This amounts to modeling in the absence of a specific target. For a recent discussion, see Weisberg (2013), Ch. 7.
investigations undoubtedly inform our understanding of worldly evolutionary outcomes. Perhaps in these cases, one might argue that the models are providing a non-causal yet scientific explanation of a class of phenomena. This would depend in part on one’s account of explanation and view about its relationship with scientific understanding. These larger issues won’t be debated here. Rather, I will show that in the paradigm case of representing a token instance of actual evolutionary change, the explanation provided by an MS-model is causal.

Putting the focus on the interpretation of models is a helpful way to frame the debate. To argue that MS-models offer causal explanations requires two steps (the discussion here will focus on a very simple MS-model of a population where selection is thought to have occurred). First, an apt causal characterization of the target system must be given. While scientists seek to model worldly phenomena, in practice models are compared to target systems (Weisberg, 2013, 90-91). Target systems abstract from the entirety of a phenomenon to focus on a subset of entities, properties, and causal relationships. In typical cases, most of the abstractions used to pick out the target system will not be controversial: the focus is on a single population of individual organisms with one or a few heritable traits of interest, along with a key feature or features of the environment, with changes tracked over a limited period of time. Many details will be neglected in characterizing the target system and some idealizations will be made (e.g. in subsuming particular entities/properties under types). What appears unusual about evolution is the lack of consensus about how to appropriately describe the causal relationships among the entities and properties in the systems of interest. Some of the background debate on this question is reviewed below in section 2. I then develop and apply an appropriate causal framework for the target system (sections 3 & 4). As detailed below, this work is closely aligned with proposals due to Millstein, who applies Salmon’s
The other type of fidelity criteria concern “dynamical fidelity”: this has to do with how well the model produces an output similar to the real world system (Weisberg, 2013, p.41).

Weisberg describes several varieties of idealization and their purposes. For target-directed models, idealizations are primarily motivated by pragmatic considerations or epistemic ones. In the case of “minimalist” idealization, the modeler seeks to include in the model only “core causal factors which give rise to a phenomenon (Weisberg, 2013, 100).” By featuring the most important causal factors, the model can effectively serve explanatory goals.
2. Causal Controversies: Finding the Right “Level” and the Right Concept

The causal character of evolutionary change and natural selection in particular has been described in several ways. Beginning with Sober (1984), one approach sees selection as a force analogous to Newtonian forces (alongside others such as drift, mutation and migration). Another approach seeks to depict selection using a mechanistic explanatory framework, and several authors have explored this treatment.

In her (2006), Roberta Millstein highlighted a key question for those seeking a causal account: does natural selection occur at the level of the population or the organism? Notably, Walsh, Ariew & Matten allow for causes only at the level of the organism, expanding on statisticalism’s negative claim as follows: “MS-models do not cite the causes of population change, which are all events involving individual organisms (2017, 5, footnote 8).” Millstein explores the idea of whether distinctively population-level causation should be used to characterize selection. While I will argue Millstein’s account falls short, clarifying the notion of population-level causation and its relationship to the activities of organisms is a necessary step toward providing a response to the statisticalist position.

---

277 The dynamical interpretation has also been defended by Stephens (2004, 2010). A recent close examination of the analogy is made by Hitchcock & Velasco (2014).
The importance of this distinction can be seen beginning with Sober’s discussion in *The Nature of Selection.* 279 Sober sees selection and other forces as components that together account for the net force on a population. The final outcome or effect of the net force is described by changes in gene frequencies. 280 But what is responsible for the force? Sober describes the role of source laws and consequence laws: “The former describe the circumstances that produce forces; the latter describe how forces, once they exist, produce changes in the systems they impinge on” (Sober, 1984, 50). Sober says that finding source laws for natural selection is the task of showing how “properties of the organism and environment can render certain kinds of traits selectively advantageous” (58–59). The source laws give rise to differences in the fitness of individual organisms. The consequence laws “are part of the province of population genetics”, and involve the calculations of gene frequencies given the relative fitness values (selection coefficients) and other relevant information (59).

While the causal frameworks discussed below do not employ the force analogy, 281 the key point to notice here is that the productive causation described by Sober’s source laws references only *organism-level* properties and the environment. Consequence laws describe the effects, but these are assessed exclusively at the *population-level* using modern synthesis models. For the statisticalist, the causal story involved in the source

279 Earlier articles defending the statisticalist position, such as Matthen & Ariew (2002) and Walsh, Lewens, & Ariew (2002), were in part responding to Sober in developing their arguments.
280 “The present point is that selection is a force whose effects are calculated in the currency of gene frequencies (Sober, 1984, 46).”
281 Rather than consider natural selection or drift as forces or causes, they are treated as phenomena—types of evolutionary change—for which we are seeking an explanation. There is discussion in the literature about the need to distinguish between conceptions of natural selection and drift as “processes” vs. “products”: see Millstein (2002) and Stephens (2004). Here they will be mainly discussed as products. To be sure, the causal explanatory interpretation discussed below is consistent with claiming that MS-models describe causal “processes” of selection or drift over time. But the focus here is on the debate between two proposals regarding how MS-models can be said to explain evolutionary changes in trait distributions.
laws is no longer in view when it comes to explaining changes in population structure using these models. In addition, while fitness is not itself a causal factor in his account, Sober sees fitness differences among individual organisms as summarizing the impact of source laws.\footnote{See Sober (1984), pp. 49-51. Sober’s view is that an organism’s fitness supervenes on physical properties (which are the true causal factors). He also describes the source laws as “physical circumstances that can generate fitness differences” (p. 51). In a later article, Sober (2013) argues for a causal role for trait fitnesses (as distinct from individual fitnesses).} The statistologist points out that there may be no direct translation possible from vernacular fitness to trait fitness (a population-level parameter), and MS-models utilize the latter.

Millstein (2006) argues that trait distribution is a population-level property that is itself germane to evolutionary change. This is perhaps easiest to see in the example of frequency-dependent selection.\footnote{See Millstein (2006), pp. 629-631.} In these cases, the relative survival and reproductive success of two trait types may be completely reversed depending on their initial prevalence in the population. To a less dramatic degree, the example generalizes. The trajectory of the distribution of heritable traits will depend on the initial distribution.

Given that the distribution of a trait is a property of populations, and since the nature of subsequent evolutionary change depends on this property, Millstein proposes treating the property as a “cause”:

The ‘cause’ is variation in the population – more specifically, heritable differences in physical survival and/or reproductive abilities – a property of the population rather than of the individual. Similarly, I will argue that the ‘effect’ is differences in reproductive success, a property of the population rather than the individual.

(Millstein, 2006, 631-632)
Millstein argues that calling the initial population-level property a cause is appropriate, since the kind of dependency relationship involved fits several theories of causation, including counterfactual and manipulation-based approaches. The significance of this idea for the debate over statisticalism is clear: if trait distribution is a causal property, then the fact that it is cited in MS-models may support interpreting the explanations they provide as causal rather than statistical.

However, identifying this causal role for population-level properties isn’t sufficient to settle the debate. Does the initial distribution really cause the population structure to change? After all, it is the organisms that are doing the living, mating, and dying. They are the tightly organized entities whose localized physical interactions seem paradigmatic of causation. What is the relationship between these organism-level happenings and the population-level causal property?

In a response to Millstein, Glennan (2009) aptly raises the point that there are two concepts of causation, often referred to as difference-making and production. While the concepts can be elaborated in different ways, difference-making is primarily understood as counterfactual dependence. Production, on the other hand, involves bringing about change via a spatio-temporal influence. It is the concept invoked in the notion of a force and it is also the notion defenders of the statisticalist position have in mind:

---

284 Reisman & Forber (2005) argue that natural selection is a population-level cause, using manipulation as a criterion. Shapiro & Sober (2007) also apply the manipulation criterion for causation in their critique of the statisticalist position. For responses, see Walsh (2007, 2010).

285 For Glennan’s discussion of the two concepts, see Glennan (2009) pp. 327-331. See Chapter 2 for a detailed discussion of these concepts and their role in natural science explanations.
Natural selection is ontologically derivative on individual-level events such as births, deaths, and mutations— the latter events are brute physical occurrences—they occur by the transmission of physical influence from one place to another. (Matthen & Ariew, 2009, 216)

The distribution of heritable traits at a time makes a difference to the subsequent evolutionary change (relative to other possible starting points), but what it does not do is produce the change via a spatio-temporal influence. This seems to point the debate back toward an assessment that organisms are the key causal actors.

3. Productive Population-Level Interactions

In a subsequent article, Millstein (2013) offers a revised account intended to show that selection can indeed be explained in terms of productive causation at the level of the population. Her strategy is to apply Salmon’s (1984) mark-transmission causal process theory, a theory of causal production. Recall that “causal processes” are the basic entities in Salmon’s theory. Compared to an event, a process is typically extended in time as well as space: “A baseball colliding with a window would count as an event; the baseball, traveling from the bat to the window, would constitute a process” (Salmon, 1984, 139). There are two aspects to causation for Salmon: propagation and production. A causal process is distinguished from non-causal processes by its ability to propagate (or transmit) its structure through space-time. In doing so, it can transmit a signal, in the form of a modification of that structure (a “mark”). A mark is an alteration of the process,

---

286 See Chapter 3 for a comprehensive review.
In reviewing the possible application of other approaches to mechanistic explanation, Millstein thinks they tend to emphasize a “decompositional” framework, and suggests that this emphasis leads to difficulties in characterizing selection. See Millstein (2013), pp. 152-153. The distinction is that of an emphasis on etiological vs. constitutive causal explanation (notions discussed by Salmon, 1984, pp. 279-80). As discussed further below, the case of selection requires the right treatment of both aspects. On the possibility that mechanistic frameworks such as Glennan’s (1996, 2002) or that of Machamer, Darden & Craver (2000) might be adapted for the case of selection, see Barros (2008). Barros does note the need for a two-level (population and organism) analysis, although a precise relationship between the two levels is not articulated (Barros, 2008, pp. 318-320).

Millstein is attracted by Salmon’s framework in part because she sees potential to causally explain evolutionary change in a population without prioritizing the activities of organisms. The first step in applying Salmon’s account is to establish that a population is the kind of thing that can correspond to a causal process. There are several ways to define populations, but it seems adequate for Salmon’s purposes that it can be seen as an entity that, despite its loose organization, is extended in time and space and can plausibly enter into local causal interactions. A migrating animal population that encounters an obstacle, such as a swamp, might be an example (picture some individuals becoming lost traversing the swamp). The changes wrought by this interaction might make reference to population-level properties such as a reduced size and a change in the distribution of

---

287 In reviewing the possible application of other approaches to mechanistic explanation, Millstein thinks they tend to emphasize a “decompositional” framework, and suggests that this emphasis leads to difficulties in characterizing selection. See Millstein (2013), pp. 152-153. The distinction is that of an emphasis on etiological vs. constitutive causal explanation (notions discussed by Salmon, 1984, pp. 279-80). As discussed further below, the case of selection requires the right treatment of both aspects. On the possibility that mechanistic frameworks such as Glennan’s (1996, 2002) or that of Machamer, Darden & Craver (2000) might be adapted for the case of selection, see Barros (2008). Barros does note the need for a two-level (population and organism) analysis, although a precise relationship between the two levels is not articulated (Barros, 2008, pp. 318-320).

288 See further discussion of populations as entities in Section 4 below, which highlights Millstein’s own causal interactionist population concept.
certain characteristics (the environmental obstacle would be changed as well, although likely in a less interesting way). To see the presence of causal production in these sorts of cases, it doesn’t seem essential to refer to the interactions of organisms within the population. While a detailed examination of these interactions would generally provide a fuller account of what occurs, the claim is that the interaction of the population itself with another entity is the central event producing change. As a result, it seems reasonable to conclude that a population might be conceived as a causal process that can enter into interactions with other processes to produce change in structure (a mark). This change would then be transmitted forward until the next interaction.  

This makes the question more specific: it is about whether natural selection can be produced this way. Using an example, Millstein pictures a population of finches with varied beak lengths (a heritable trait) that repeatedly interacts with a feature of the environment (seeds) “so that some finches are favored over others based on the differences among the finches, producing changes in distribution of types in the population (Millstein, 2013, 159).” The question is whether these kind of productive interactions provide a sufficient account of the changes over time that are characteristic of selection without reference to the activities of organisms. Here, Millstein fails to provide a clear account. Instead, she appeals to a re-interpretation of Salmon’s concept of propagation, suggesting (contra Salmon) that propagation can also produce change. The

\[289\] Millstein considers several examples of mark-transmission (involving a small mark, such as the death of a single organism, and a more dramatic mark, such as that caused by a disease spreading through a population) and concludes that “mark transmission is not only possible for a population, but commonplace (Millstein, 2013, 159).” Millstein does not address Salmon’s stipulation that a mark is transmitted in the absence of any further interactions. As a pragmatic matter a population will interact with a variety of aspects of the environment on an ongoing basis. The necessary assumption is that these other interactions can be ignored since they don’t impact the structural change of interest. See Salmon (1997), p. 464.

\[290\] This example was featured in Skipper & Millstein (2005).
motivation appears to be to deliver the desired outcome by supplementing the external population-level productive interactions:

The genotype and phenotype frequencies of a population at one point in time probabilistically propagate the genotype and phenotype frequencies to future points in time. This propagation is reflected in transition probability models, equations that describe the probability of various possible future states, given the current state of a population...As I suggested above, this propagation is a type of causal production, albeit different than the type of causal production that occurs as the result of an interaction. (Millstein, 2013, 159)

Instead of maintaining a structure between interactions, “propagation” now helps deliver the changes which are of central concern. Note also that Millstein implies that a property of the population (the initial frequencies) produces the change without reference to anything external. Asserting that a property of the population is caused by a property possessed at an earlier time essentially returns to the account offered in the 2006 paper. This is a relationship of dependence, not production.

The account thus falls short in offering a fully productive causal picture of natural selection. However, it is helpful to consider carefully how far population-level interactions alone can take us in an explanation. Picture again the population of finches with varied beak lengths. This population interacts with a key feature in the environment, plants whose variably-sized seeds are the sole, scarce, food source (assume these replenish during an annual growing season). Some seeds are small enough for all finches to eat, while larger seeds can only be eaten by long-beaked members of the group. Assume also the finches mate and reproduce annually. After several years/generations the distribution
has changed, and long-beaked members have increased in proportion. We have a candidate instance of selection.

Now picture the interaction of the population of finches with the population of seed-producing plants as a Salmon-style productive one. This is a time-extended interaction, lasting during the annual feeding season, and then repeated over a number of years. Now, do these productive interactions account for the changes in structure we associate with selection? We can conclude that each interaction has a differential impact on the prospects for survival of organisms in the population. And it remains the case that the initial distribution of beak-types, a population-level characteristic, will be a key factor in the outcome. But the actual survival rates will be driven by the detailed activities of the organisms. In a condition of scarcity, the impact on the population depends on the outcome of Darwinian struggles for survival among the finches. Can these organism-level survival activities be ignored? At this point, one might suggest that perhaps the outcomes are regular enough so that not much information is lost if they are ascribed solely to the population-level interaction. Here, we can note that MS-models do exclude organism-level details. However, as discussed in section one, modelling may be viewed as a two-step process. While abstraction plays a role in picking out the target system, important causal explanatory details should not be excluded at this stage.

The need to include reproductive activities seems even clearer. In order to get to the trait distributions in each new generation the offspring must be produced. Perhaps it is the case that we can sensibly assume a particular, highly regular mating and heritability scheme in building a model. But again, for the purposes of a proper causal characterization of the target system, the issue is that mating and reproduction-related
activities are so obviously organism-level interactions crucial to the detailed outcomes we are interested in.

Stepping back, we can see how suppressing organism-level survival and reproductive activities works against the goal of offering a full productive causal characterization of the target system. The population-level causal account is incomplete. The etiological account at the population-level must be supplemented with a constitutive account of how organism-level interactions account for the changes in the population.

4. Deploying the Dispositional Causal Process Framework

Salmon outlines a distinction between etiological explanations and constitutive explanations. The former trace the relevant preceding processes and interactions leading up to a phenomenon. The latter, on the other hand, cite the interactions and processes that compose a phenomenon. He did not flesh out this distinction, however. This is accomplished by the theory presented in Chapter 4.

Recall that in place of Salmon’s characterization of causal processes as entities that transmit “structure” or “characteristics” through space-time, the revised approach defines a causal process as something that transmits a cluster of dispositions toward particular interactions—in short, a dispositional profile. A causal interaction is a mutual manifestation of dispositions at a location in space-time involving two or more processes. The interaction produces change, which takes the form of altered profiles. This framework underpins explanations in the natural sciences by providing a target for their causal elements: the scientific entities cited (particles, molecules, organisms) will
correspond to causal processes while properties (velocity, charge, color pattern) correspond to regularities in their dispositional profiles. In the simplest cases of etiological explanation, the phenomenon to be explained will be either an entity and certain of its causal properties (corresponding to a causal process and aspects of its dispositional profile) or a change in an entity’s properties (corresponding to a causal interaction), and the explanation will reference causal antecedents. Of course, an explanation will also feature idealizations that may distort the correspondence to the causal notions to some degree. These reasons are varied and may include the pragmatic paring of elements for simplicity and a striving toward wide scope of application.

For instance, in the simple collision example from Chapter 4 (section 2.2) the scientific explanation of the final conditions cites the initial physical properties of each particle and the conservation of momentum/kinetic energy. What makes the explanation a causal one is the fact that these properties and their changes in the collision can be mapped onto the dispositions of causal processes and their manifestations in a causal interaction. The explanation idealizes by assuming perfect elasticity and omitting other forces, thereby capturing an important regularity in this type of scenario at the expense of a more precise description of the particular case.

With this framework for etiological explanations in place, Chapter 4 further introduces composite causal processes as the target for the constitutive explanations envisioned by Salmon. The basic idea is that one can pick out causal processes and their interactions at a higher level by attending to patterns at the lower level. In particular, differential interaction rates give rise to a hierarchy of processes. Following this line of thinking, a composite causal process is defined as two or more sub-processes (the constituting group) that interact with a greater frequency than each does with other
processes, and that together transmit a “higher-level” cluster of dispositions between space-time regions. A constitutive causal explanation, in turn, seeks to explain the entity (or properties thereof) by referencing properties and interactions of the constituting group (at both levels, the entities/properties map onto causal processes/dispositional profiles). As discussed in Chapter 4, constitutive explanations in science will rarely attempt to explain the entire entity in full. It is more common for a scientific explanation to target a single characteristic regularity in the dispositional profile of a composite process (a causal property or a capacity).

As also discussed, the complexity of biological systems may make it hard to identify the boundaries of a composite process. Here, the theoretical interests of a research program will come into play when defining entities. Still, the causal account of composition is an appropriate starting point for evaluating biological individuality in an explanatory context. In the case of biological populations, a number of accounts have been offered. Millstein (2009, 2010) has notably proposed a definition that has an affinity with the present framework. Inspired by Ghiselin and Hull’s work on biological individuality, as well as by Simon, Millstein proposes a “causal interactionist” population concept:

*Populations* (in ecological and evolutionary contexts) consist of at least two conspecific organisms that, over the course of a generation, are *actually* engaged in survival or reproductive activities, or both. The boundaries of the population are the largest groupings for which the rates of interaction are much higher within the grouping than outside. (Millstein, 2010, 67, emphasis original)

The reproductive activities go beyond successful mating to include unsuccessful mating interactions and rearing interactions involving parents and offspring. Survival activities
include classic Darwinian struggles but also broader interactions including cooperating to survive (Millstein, 2010, 67-68). 291 This proposal offers a theoretical entity type whose instances can map onto the concept of a composite causal process. 292

Returning to the present debate, the advantage of using the Chapter 4 framework lies in the way it clarifies the relationship between composite processes and those of their constituents. In particular, recall that the dispositions characterizing a composite process themselves are not constituted by a static collection of the dispositions characterizing its sub-processes. The dispositional profiles of the sub-processes in the constituting group are changing as they engage in their interaction pattern. Rather, the dispositions of the composite are constituted by the spatio-temporally extended pattern of interactions and changing disposition profiles among its sub-processes. The same is true when we consider interactions at the composite level. An interaction involving two or more composite processes takes place in a space-time region that is large relative to their respective sub-processes and their interaction rates (note that sub-processes continue to interact with their own fellow sub-processes as well as also interacting with sub-processes of the other process involved in the higher level interaction). As a result, any causal properties that map onto to the dispositions at the different levels cannot be said to have an inter-level (“vertical”) relationship of either composition or supervenience at an arbitrary point in time.

291 Millstein (2015) discusses exceptions to this definition.
292 To clarify, it is not necessary for present purposes to endorse all of the details of Millstein’s account. Also, I don’t mean to suggest any one population concept serves all theoretical purposes. I do propose that defining a population (at least in part) in reference to survival and reproductive-related causal interactions among constituent organisms is appropriate for describing the target system represented by an explanatory MS-evolutionary model. Different population concepts may be appropriate in other contexts. For an example, see Spencer (2016). For a defense of population pluralism, see Stegenga (2016).
Applying this framework to the finch example, the population corresponds to a composite causal process. The population is composed of organisms engaged in a pattern of survival and reproductive interactions: these correspond to its constituting sub-processes. The population enters into multiple yearly interactions with seed-bearing plants: these correspond to composite-level interactions. Note that the population, its dispositions, and the interaction itself can only be understood as constituted over an extended period. The constitutive survival and reproductive activities take time to manifest (a snapshot of a collection of organisms at a point in time is not a population). For instance, reproductive activities (in our example) unfold in an annual cycle and propagate the population as a continuing entity through the years.

In the case of the finches, the distribution of beak types corresponds to an aspect of the dispositional profile of the population that is of particular interest. If we picture multiple, repeated interactions of the population with the seed-producing plants, we can expect that the distribution of beak-types will be associated with a certain pattern of mutual outcomes (for a given size population, more long-beaked types mean more large seeds consumed). Calling this regularity in the dispositional profile a causal property of the population involves some idealization, since any particular outcome may deviate from the expected pattern. The analogy here would be to the role of momentum in causal explanations of physical collision scenarios: we can picture repeated trials where the precise actual outcomes of a collision are impacted by various unmeasured influences in the environment (friction, electric charges). But applying the causal property of momentum to the objects gives us the best expectation for the outcomes.

Unlike in Millstein (2006), the causal property in view here is productive: it designates a power of the population to produce certain (mutual) changes in an
interaction with the environment. While we may still refer to the property as “beak
distribution”, it is important that its causal character goes beyond the role the breakdown
of beak-types might play in a difference-making framework, say, as an initial condition.
Also, in keeping with the discussion above, “beak distribution” is also not equivalent to
the collection of individual beak-types at one point in time even when the latter are
considered as productive organism-level causal properties in their own right. To see this,
picture an early part of a year when seeds are still plentiful: at this stage the impact of the
interactions of finches and plants is broadly to reduce the number of seeds and fill finch
stomachs. Considered as organism-level causal properties, all of the beak types manifest
in similar outcomes. Later on, we can picture the supply of small seeds as having been
depleted: the long-beaked finches continue to eat large seeds while the short types go
hungry. The beak-types now manifest in very different organism-level outcomes. The
collective result of these interactions is different at these different stages. However, both
stages are part of the constitution of the extended population-level interaction, which in
our example may last a season or a year. As discussed below, to do justice to this kind of
causal property in an explanation it is not enough to represent it with a tabulation of trait
types at a point in time. Its role in producing outcomes must be acknowledged. In MS-
models, this role is played by the trait fitness values that are linked to the types.

5. Interpreting an MS-model

The description of the target system involves some abstraction and idealization. In
the finch example, the target system of interest abstracts from most of the details
regarding the population, its constituents, and its interactions with its environment. It
focuses on one productive interaction, repeated over a period of years, and on one population-level property relevant to determining the outcome of these interactions. A model may omit yet more information, and will introduce mathematical representations of various aspects of the system. In the case of target-directed modelling, as discussed, the model will maintain a degree of fit with the target, and it is typically the case that there will be a mapping between model elements and key features of the target’s causal structure. The question for MS-models of evolution is whether in their case this kind of mapping is abandoned, particularly in light of the fact that organism-level details are excluded.

For an example like the finches, researchers might analyze selection pressure for longer beaks using the breeder’s equation, using observations and measurement of parent and offspring beaks to infer heritability. To more easily engage the arguments of the statisticalists, we will consider a simpler analog of the finch population so that a model based on relative fitness of discrete types can be applied. So, picture a population of haploid organisms that come in two trait types corresponding to two alleles at a single locus. These have an annual life cycle, and we will assume that a discrete model with non-overlapping generations provides an adequate approximation. To simplify heritability, assume the organisms are asexual and reliably produce their own type. These types interact with a food source in their environment in the manner of the finches, with the two having different abilities to access a varied food source over the year (a la short and long beaks). This interaction with the food source is the main driver of differential survival, but all surviving organisms reproduce at the end of the year and we can thus

293 See Crow & Kimura (1970), Ch.1.
plausibly separate fitness into two components: viability and fertility. In fact, we will initially assume fixed and equal fecundity for both types so that fitness differences are entirely due to viability.

Table 7.1: Beginning of year census by type (annual fitness values derived)

<table>
<thead>
<tr>
<th></th>
<th>Year 1</th>
<th>Year 2</th>
<th>Year 3</th>
<th>Year 4</th>
<th>Year 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Long</td>
<td>50</td>
<td>60</td>
<td>78</td>
<td>140</td>
<td>154</td>
</tr>
<tr>
<td>Short</td>
<td>50</td>
<td>45</td>
<td>41</td>
<td>81</td>
<td>65</td>
</tr>
<tr>
<td>Total</td>
<td>100</td>
<td>105</td>
<td>119</td>
<td>221</td>
<td>219</td>
</tr>
<tr>
<td>$W_vL$</td>
<td>0.60</td>
<td>0.65</td>
<td>0.90</td>
<td>0.55</td>
<td></td>
</tr>
<tr>
<td>$W_vS$</td>
<td>0.45</td>
<td>0.45</td>
<td>1.00</td>
<td>0.40</td>
<td></td>
</tr>
<tr>
<td>$W_fL$</td>
<td>2.00</td>
<td>2.00</td>
<td>2.00</td>
<td>2.00</td>
<td></td>
</tr>
<tr>
<td>$W_fS$</td>
<td>2.00</td>
<td>2.00</td>
<td>2.00</td>
<td>2.00</td>
<td></td>
</tr>
<tr>
<td>$W_tL$</td>
<td>1.20</td>
<td>1.30</td>
<td>1.80</td>
<td>1.10</td>
<td></td>
</tr>
<tr>
<td>$W_tS$</td>
<td>0.90</td>
<td>0.90</td>
<td>2.00</td>
<td>0.80</td>
<td></td>
</tr>
</tbody>
</table>

Table 7.1 gives census information with fitness values inferred from observed changes from the beginning of one year to the next ($W_v$ refers to viability, $W_f$ to fertility, and $W_t$ to total fitness). Next, in the usual fashion we ignore the absolute magnitudes and focus on the proportions of types and on relative fitness (Table 7.2). Here $p$ and $q$ are the

---

294 See Orr (2009), pp. 531-2.
proportions of L and S types in the population and \( w_L \) and \( w_S \) are total trait fitness values (with \( w_L \) normalized to 1).

Table 7.2: Type frequencies and relative fitness (annual fitness values derived)

<table>
<thead>
<tr>
<th></th>
<th>Year 1</th>
<th>Year 2</th>
<th>Year 3</th>
<th>Year 4</th>
<th>Year 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>( p )</td>
<td>0.50</td>
<td>0.57</td>
<td>0.66</td>
<td>0.63</td>
<td>0.70</td>
</tr>
<tr>
<td>( q )</td>
<td>0.50</td>
<td>0.43</td>
<td>0.34</td>
<td>0.37</td>
<td>0.30</td>
</tr>
<tr>
<td>( w_L )</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>( w_S )</td>
<td>0.75</td>
<td>0.69</td>
<td>1.11</td>
<td>0.73</td>
<td></td>
</tr>
<tr>
<td>( s (w_L-w_S) )</td>
<td>0.25</td>
<td>0.31</td>
<td>-0.11</td>
<td>0.27</td>
<td></td>
</tr>
<tr>
<td>( \Delta p \text{ actual} )</td>
<td>0.071</td>
<td>0.087</td>
<td>-0.024</td>
<td>0.070</td>
<td></td>
</tr>
<tr>
<td>( \Delta p \text{ calc.} )</td>
<td>0.071</td>
<td>0.087</td>
<td>-0.024</td>
<td>0.070</td>
<td></td>
</tr>
</tbody>
</table>

Here, the change in the proportion of types (and implied fitness) has been observed (“\( \Delta p \text{ actual} \)”), but a calculated change based on fitness can be derived from the following formula (“\( \Delta p \text{ calculated} \)”).\(^{295}\)

\[
\Delta p = \frac{pq}{1 - q}
\]

The relative fitness of the types is captured by the selection coefficient, \( s \), which is the key factor in the equation computing the change in population structure.

\(^{295}\) The denominator as presented is equivalent to the mean relative fitness of the population (which often appears in presentations of this formula). See Orr (2009), p. 533.
We can imagine that researchers investigating this outcome as one of selection will seek to estimate the “correct” value for the difference in trait fitnesses. For the period observed, one can calculate (geometric) mean and variance of fitness values (here the means imply $s$ equal to about .20). But since this is a short period of time and the actual outcomes in each year are influenced by a variety of factors, many other considerations may come into play. In the simplified example concocted here, the annual outcomes for years 1, 2 and 4 imply similar relative fitness, but year 3 looks very different. Perhaps the weather was wet in year 3 (leading to plentiful food) and relatively dry in the others (leading to scarcity). But why was relative fitness actually greater for the short type? Perhaps researchers see no reason for this and attribute it to random drift: a wet year should imply no expected fitness difference. Further, weather observations over longer periods suggest wet and dry years should be equally likely. Given all of these considerations, researchers settle on estimated fitness values that imply outcomes like the mean of years 1, 2 & 4 half the time and no fitness difference the other half. Table 7.3 compares actual evolution to expected annual changes based on these estimates ($s = .15$).

**Table 7.3: Type frequencies and relative fitness (mean fitness values estimated)**

<table>
<thead>
<tr>
<th></th>
<th>Year 1</th>
<th>Year 2</th>
<th>Year 3</th>
<th>Year 4</th>
<th>Year 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>$L$</td>
<td>0.50</td>
<td>0.57</td>
<td>0.66</td>
<td>0.63</td>
<td>0.70</td>
</tr>
<tr>
<td>$S$</td>
<td>0.50</td>
<td>0.43</td>
<td>0.34</td>
<td>0.37</td>
<td>0.30</td>
</tr>
<tr>
<td>$wL$</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>$wS$</td>
<td>0.85</td>
<td>0.85</td>
<td>0.85</td>
<td>0.85</td>
<td></td>
</tr>
<tr>
<td>$s (wL-wS)$</td>
<td>0.15</td>
<td>0.15</td>
<td>0.15</td>
<td>0.15</td>
<td></td>
</tr>
<tr>
<td>$\Delta p_{actual}$</td>
<td>0.071</td>
<td>0.087</td>
<td>-0.024</td>
<td>0.070</td>
<td></td>
</tr>
</tbody>
</table>
Now let’s turn to an exploration of the fit between the model and the causal features of the target system. According to the statisticalist, trait frequency and trait fitnesses explain the changes in the population structure associated with selection and drift without citing causes. However, I would argue that the key population-level causal property in the target system is cited in the model. Beginning-of-year type frequencies together with the estimated relative trait fitnesses correspond to this causal property. Recall, using the framework developed in section 4, that the population is understood as a composite causal process with dispositions toward certain interactions with the food source. These dispositions (mutually) manifest in particular annual outcomes, and “beak distribution” describes a causal property that picks out a regularity in these dispositions/manifestations. As such, it describes the expected outcome of this kind of interaction. This is just what researchers investigating the real world population are keen to estimate. The model breaks this causal feature down into two components: for each possible frequency combination, there is an expected outcome that is quantified by applying the selection coefficient.\textsuperscript{296} Since these are the key explanatory variables in the model, and they map onto a causal property of the population, the model causally explains the trajectory of frequencies associated with selection.\textsuperscript{297}

\begin{center}
\begin{tabular}{|c|c|c|c|c|}
\hline
 Δp \textit{calc.} & 0.040 & 0.039 & 0.035 & 0.037 \\
\hline
\end{tabular}
\end{center}

\textsuperscript{296} Here we assume relative fitness is unrelated to frequency, but of course it may not be (frequency-dependent selection).

\textsuperscript{297} Sober (2013) concludes that initial trait frequencies and trait fitnesses are population-level causes. However, he applies a difference-making model in conceptualizing the relationship between these causes and the effect (change in frequency).
To emphasize again, because the causal feature in focus is a population-level property, it is not related in a straightforward mathematical or statistical way to properties of the organisms in the population at a given time. Rather, it is constituted by the extended pattern of survival activities (and resulting changing properties) of the organisms in the manner discussed above. There is a complication, however. It is only the viability component of relative fitness that corresponds to this property. Reproductive activities sustain the population, but are not constitutive of this population-level feature. If we change the example above so that the expected fertility of the two types differ, then the initial frequency and (total) fitness values in the model would not map only onto the single population-level property as before. In this case, trait fitnesses would correspond to an expected outcome associated with both the population-level causal interaction with the food source as well as organism-level reproductive activities. But given the theory of composite causal processes, we can see that these model elements would now correspond to a combination of expected changes due to the interaction as well as further changes that occur during the propagation of the population.\textsuperscript{298} Since propagation and production are the two aspects of causation in our Salmon-style theory, the mapping of these model elements is still to causal features of the population.

Of course, other complexities are overlooked in the example. Different heritability or mutation rates would require adding yet more components to the model’s trait fitness values.\textsuperscript{299} The more the single fitness parameter has to encompass, the less straightforward is the mapping relationship to particular population-level features. Nevertheless, because

\textsuperscript{298} Now we can see how propagation can be associated with change. This was the idea that was not well-founded in Millstein’s (2013) application of Salmon’s original causal framework (discussed in section 3).

\textsuperscript{299} For a sexual species, fitness would incorporate different prospects for mating success.
trait distribution and relative trait fitness still correspond (if imperfectly) to causal features of the target system, the explanatory character of the model remains causal. The fact that trait fitness may compress a lot of information into a single number likely helps to account for the difficult philosophical debate over its nature. I would note in passing that the concept of vernacular fitness also involves this kind of compression: it summarizes in a single value the expected outcome of a large number of varied productive causal interactions, including those between organisms and environment and organisms with each other. While it is a causal concept broadly speaking, it is not a single causal property as defined in the framework of section 4.

6. A Brief Look at Four Pillars

With this account of the causal explanatory character of MS-models in hand, we can examine what Walsh, Ariew & Matthen (2017) describe as four core commitments or “pillars” that are central to statisticalism. The first pillar states that “natural selection is a higher-order effect (7).” The authors begin by observing how a change in trait distribution from one point in time to another is a simple consequence of organisms with that trait entering or leaving the population. From this starting point, they conclude that “evolutionary changes in population trait distribution are analytic consequences of the activities of the organisms (7, emphasis original).” This is what it means to be a higher order effect: a change that is a non-causally entailed by lower-level changes (which themselves may have a causal explanation). But keeping in mind that the authors claim that MS-models explain the change without referencing the organisms, more must be said.
To flesh out the idea they have in mind, the authors compare the case of changing trait distribution to the case of the changing location of the center of mass (c) of particles in a container (which may be in motion and colliding with the walls of the container). The location of c is entailed by the locations of the particles at any given time: it is a higher-order effect. There is no causal story at the level of c explaining its motion through time. The authors also stress that detailed individual level causal information is often not needed. Using the example of diffusion of one liquid in another, while there is again a lower-level story involving moving particles, information in the form of a summary statistic (e.g. a density gradient) can be sufficient. This is important for the argument since detailed organism-level information is not used in MS-models.

Unfortunately these analogies are not apt. A distinction must be drawn between the relationship of higher order and lower-level facts at points in time on the one hand and the relationship between the explanation of changes at the higher and lower levels on the other. In the center of mass case, causal explanations of the individual particle locations at a time would cite the properties of mass and velocity at some earlier time, utilizing conservation laws if a collision has occurred. The explanation of the location of c adds to the particle-level explanations a simple mathematical function connecting it to the time-indexed values of the properties of the particles. In the diffusion case, the authors do not discuss a particular mathematical model, but generally diffusion is a function of the random motions of ensembles of particles, where if lower-level causal explanations were sought, they would again be framed using physical properties. MS-models differ from these cases. While the beginning and ending trait distributions can indeed be straightforwardly tabulated from information about the organisms, the explanatory information in the MS-models does not have a mathematical or statistical relationship to that of the lower-level.
In the MS-models, a key explanatory role for the change in trait frequency is played by trait fitness differences. Where do trait fitness values come from? If they are inferred directly from an actual change in population structure (as in Table 2 above), then they can indeed be calculated from the beginning and ending trait distributions (along with other information about migration, etc.) But if derived in this way, then the trait fitness differences have not truly explained the outcome. Perhaps the “true” means and variances differ from what has been observed. Perhaps drift should be ascribed a role, etc. But setting this aside, we can also ask what the relationship is between any explanation of an observed frequency change given in terms of trait fitness differences on the one hand and the collection of causal explanations for the organism-level outcomes on the other. The causes of an organism surviving and reproducing would be given in terms of its own properties, summarized perhaps in terms of vernacular fitness. But the authors themselves agree that there is no straightforward relationship between trait fitness and the causal properties of organisms: in particular they deny that trait fitness is derivable from vernacular fitness (see the second pillar below). There is thus a difference between how the explanations at the higher and lower-levels are related in the evolution case compared to the other examples of “higher order effects” offered. In the latter cases, we accept the idea that the explanation of the higher-level outcome is parasitic on the lower-level causal explanations because there is a mathematical/statistical relationship between the higher-order change and the changes in the properties invoked in the lower-level explanations. In the evolution case, while there is a straightforward relationship between trait distribution and organism-level facts at particular times, there is no such relationship between the key explanatory factor cited to explain the change in trait distribution and the properties invoked in organism-level causal explanations.
What is the underlying reason for this disanalogy? Returning to Simon’s terminology, some complex systems are “decomposable” while others are not. For instance, consider an ideal gas. Here the system is decomposable (Simon, 1996, 197). This is due to an assumption that interactions between molecules are negligible: the molecules do not have a characteristic pattern of interacting with each other at any greater frequency than they do with the external system (the container). With this assumption in place, the system can be decomposed into the subsystems (the individual particles) and the familiar properties of the gas are mathematically derived. An evolving population, in contrast, cannot be decomposed, as it is comprised of organisms engaged in an internal pattern of survival and reproductive activities that serve to distinguish it from a mere collection of organisms. As a result, as discussed above, a population has certain causal properties (manifested in its population-level interactions) that cannot be derived from the causal properties of the organisms at any particular point in time. There are exceptions of course: conserved physical quantities such as mass and net charge are examples of what Wimsatt calls “aggregative properties”: here the relationship between a property of the whole and that of the parts is insensitive to their internal causal organization (see Wimsatt, 2007, 274-287). But in the case of the population-level causal properties of interest in cases of selection, these conditions are not met.

The second pillar is that trait fitness is “primitive”, meaning that “in particular, (unlike other aggregate-level parameters, e.g. pressure and temperature) it has no definition in terms of the causal properties of individuals that make up an ensemble (Walsh, Ariew & Matthen, 2017, 9).” The present analysis agrees with this conclusion, although for different reasons. The authors first discuss why trait fitness does not follow simply from vernacular fitness, given the number of other factors that can impact the reproductive output of organisms possessing a trait. But since these factors may also
derive ultimately from (or perhaps supervene on) the organisms in some way, “this does not establish that the distribution of trait fitnesses cannot be defined in terms of the properties of individuals in the population (10).” Instead, the authors appeal to the idealizations that go into creating an MS-model to reach two conclusions. First, given the many causal influences on any actual population that are ignored in the model, the model’s trait fitness values cannot “be defined in terms of the properties of actual individuals in the population. (11)”. Second, they say trait fitness, as a statistical property of an idealized population, is itself an “abstraction” that “cannot be defined in terms of the properties of actual individuals in a concrete population (11).” For these reasons, trait fitness should be “considered to be a primitive theoretical concept (11).”

Pinning this conclusion on the abstraction/idealization process inherent in modeling is misguided, however. Certainly much detailed causal information is omitted and/or distorted, but as we have seen this does not eliminate the interpretive correspondence with key causal aspects of the target. It is true that, strictly speaking, an idealized population’s features cannot be “defined in terms of” the properties of an actual system. But this, along with the authors’ concern about defining an “abstraction” in terms of something concrete, is not germane to the debate over whether or not the explanation provided by MS-models is causal. The reason why trait fitness cannot be straightforwardly derived from organism-level causal properties (even at a similar degree of abstraction) is that it corresponds to a distinctively population-level causal property.

---

300 This is not to say that question of how to define similarity conditions between abstract entities and concrete phenomena is an easy one. In his discussion of comparing mathematical models to target systems, Weisberg suggests that the target can also be represented abstractly, resolving the concern (Weisberg, 2013, p. 95). Regardless of how this question is viewed, it would be untenable to claim that causal features of concrete worldly systems can never be abstractly represented.
The third pillar can be mentioned very briefly. Since MS-models involve a high
degree of abstraction and idealization, the authors note that they can also be applied to
non-biological phenomena: they are “substrate neutral (11)”. This is a common fact about
mathematical and computational models. One of the virtues of idealization in modeling is
the potential for expanded scope of application. It is still the case that in target-directed
modeling, the features of a highly idealized model will map onto the target system. If the
model is used to explain, many of these features will typically map onto causal elements
of the target.

The fourth pillar states that “we can only say that a population is undergoing MS-
selection, or drift, or both, relative to an MS-model (12).” The authors present a simple
example where, given the interpretation of a limited stretch of data, a model for a given
evolutionary change might be created that implies either selection or drift. Generally,
given that trait fitness differences play a key role in these models and their long-run mean
and variance values are not directly observable, the presence of selection and drift will
depend on the judgments made by the modeler in representing the system. This seems
uncontroversial, but the authors believe it is very significant:

The model relativity of MS-selection and drift marks a watershed between the
statistical and causal interpretations. The model-relativity of MS-selection and
drift sits uneasily with the causal interpretation, while it is perfectly consonant
with statisticalism. (Walsh, Ariew & Matthen, 2017, 13)

Why do the authors think there is a particular problem here? They say that “the tension
between model-relativity and the causal interpretation arises from the commonly-held
intuition that causal relations are objective, description-independent features of the world
(13).” But the modeler can accept this as well as the point about model-relativity: the
objective causal character of the world is obviously not always transparent to our
inspection. Assessing the presence of selection in a population is a challenge. The fact that
an MS-model can describe different possible causal structures depending on the
estimation of its parameters is a feature everyone can embrace. As more data is collected
and analyzed, the model can be modified to better fit the target.


doi:10.1093/bjps/axy058


Sober, E. (2013). Trait Fitness is not a Propensity, but Fitness Variation is. *Studies in the History and Philosophy of Biological and Biomedical Sciences, 44*, 336-341.


