



1996

## Explanation and Ontological Reasoning

Dwayne Mulder  
*Loyola University Chicago*

Follow this and additional works at: [https://ecommons.luc.edu/luc\\_diss](https://ecommons.luc.edu/luc_diss)



Part of the [Philosophy Commons](#)

---

### Recommended Citation

Mulder, Dwayne, "Explanation and Ontological Reasoning" (1996). *Dissertations*. 3675.  
[https://ecommons.luc.edu/luc\\_diss/3675](https://ecommons.luc.edu/luc_diss/3675)

This Dissertation is brought to you for free and open access by the Theses and Dissertations at Loyola eCommons. It has been accepted for inclusion in Dissertations by an authorized administrator of Loyola eCommons. For more information, please contact [ecommons@luc.edu](mailto:ecommons@luc.edu).



This work is licensed under a [Creative Commons Attribution-Noncommercial-No Derivative Works 3.0 License](#).  
Copyright © 1996 Dwayne Mulder

LOYOLA UNIVERSITY OF CHICAGO

EXPLANATION AND ONTOLOGICAL REASONING

A DISSERTATION SUBMITTED TO  
THE FACULTY OF THE GRADUATE SCHOOL  
IN CANDIDACY FOR THE DEGREE OF  
DOCTOR OF PHILOSOPHY

DEPARTMENT OF PHILOSOPHY

BY

DWAYNE HUDSON MULDER

CHICAGO, ILLINOIS

JANUARY, 1997

Copyright by Dwayne Hudson Mulder, © 1996

All rights reserved.

## ACKNOWLEDGMENTS

I wish to thank Paul Moser for directing this dissertation and for years of valuable philosophical discussion, training, guidance, and support. He has been an encouraging and generous colleague and friend. I thank Suzanne Cunningham and J. D. Trout, not only for their valuable assistance in the completion of this dissertation but also for their very helpful and enjoyable philosophical discussions and friendship. I am grateful for the financial support I have received over the years from Loyola University of Chicago and the Schmitt Foundation, through graduate assistantships, a teaching fellowship, and the Schmitt dissertation fellowship, which allowed me to complete the lion's share of this work. Finally and most importantly I thank my wife, Pamela Van Halsema, for always encouraging me and tolerating me through the good times and the tough times of graduate school. To her I dedicate this dissertation.

D.H.M.

*Chicago, Illinois*  
*May, 1996*



# CONTENTS

ACKNOWLEDGMENTS . . . . .	iii
Chapter	
1. INTRODUCTION . . . . .	1
The Importance of Explanation	
Explanation and Existence	
Intuitions about Explanation	
Summary of Chapters	
2. THE PHILOSOPHY OF EXPLANATION . . . . .	22
Why Study Explanation?	
What Is Explanation?—The Importance of Examples	
What We Don't Need from Theories of Explanation	
3. METAPHYSICAL EXPLANATION . . . . .	55
What We Should Expect from the Philosophy of Explanation	
Plato on Adequate Explanation	
Aristotle on Adequate Explanation	
The Threat of Circularity	
Aristotle's Arguments for Substance based on Explanatory Considerations	
Rationalists on Adequate Explanation	
4. EXPLANATION IN CONTEMPORARY ONTOLOGY . . . . .	109
Examples of Explanation in Contemporary Ontological Reasoning	

Inference to the Best Explanation as the Basic Form of  
Nondeductive Inference

Inference to the Best Explanation in the Philosophy of Science

Inference to the Best Explanation in Ethics

Inference to the Best Explanation in Epistemology

Inference to the Best Explanation in the Philosophy of Mind

The Explanation of Experience *via* Existence Claims: A Broad Sketch

Explanation in Arguments Regarding Existence

5. HEMPEL'S ANALYSES OF EXPLANATION . . . . . 149

Deductive Explanation

Statistical Explanation

Modelling Explanation

Hempel's Method: Conceptual Analysis

6. CONCEPTS OF EXPLANATION, AND REALITY . . . . . 195

Two Types of Justification

Causation in Scientific Explanation

Objective Homogeneity

The Statistical and the Propensity Interpretation of Probability Statements

Intuitions Contrary to Salmon's

Contrastive Why-Questions

7. EXPLANATION, UNDERSTANDING, AND ONTOLOGY . . . . . 230

Intuitions About Explanation, and Understanding

The Subjective Sense of Understanding

The Connection Between Explanation and Understanding

Understanding and an Ontological Background

The Threat of Circularity Again

Conclusion

REFERENCES . . . . . 274

VITA . . . . . 285

## CHAPTER 1

### INTRODUCTION

#### The Importance of Explanation

Explanation is one of the most important, fundamental, and ubiquitous intellectual tasks we perform. Our need to know *what* is the case often subserves our desire to understand *why* things happen as they do. Perhaps this sounds peculiar, for we think of descriptive knowledge as being preliminary to explanatory knowledge in the sense that we must first know what exists and what happens in order to have something to explain, and we are often in the situation of knowing what is the case without knowing why it is the case. It turns out, however, that we often draw conclusions about what exists and what happens on the basis of how well those conclusions explain some set of observations. As David L. Miller puts it, “description without explanation is blind and sterile,” although “explanation apart from description is incomplete and insignificant” (1946, 241).

It is usually only when we are focussed on some specific task we must perform that explanation and understanding have merely secondary importance. A worker might not care at all *why* chemicals have to be combined in a certain ratio to accomplish a desired effect, but only that they actually do combine in that ratio and do produce that effect. An undergraduate might not care at all *why* existential instantiation in a proof in quantificational logic must introduce a new constant, but only that one must follow that rule on a test to get correct answers. In many other matters, however, it is highly

unsatisfactory to know what is the case without knowing why. It is not enough to know that a close friend has died, we need to know why. It is not enough to know that the car isn't working, we need to know why. It is not enough to know that the light from other galaxies is red-shifted, we need to know why. Knowledge without explanation is generally deeply unsatisfying and provocative.

Explanation appears to be an important intellectual desideratum for humanity from the earliest periods of recorded history. One purpose of explanation is to engender understanding, and understanding is the most characteristic function of human reason, or the human intellect. Asking explanation-seeking questions and trying to answer them require intellectual creativity in an attempt to model the world, an activity that is most fully developed in us humans.

Explanation remains today one of our foremost intellectual goals. Technology (practical know-how) is another highly prized commodity as well, however. We remain primarily tool-users. We continue to be occupied to a great extent with providing for ourselves and controlling our environment. Indeed, far too many still suffer deprivations with respect to the basic needs of life. While some people continue to struggle with irrigation systems, construction, and manufacture, others explore exotic technologies for curing diseases, playing recorded music, shopping by computer, and videotaping their eventful lives. Although we clearly desire more than mere know-how, it is unclear whether many people desire understanding through genuine explanations for its own sake or just as a means of advancing technology, a means of developing the next generation of video amusement. It is unclear whether most people see explanation of the world

around us as a primary goal and advanced technology as simply a fortunate by-product, or see technology as the primary goal and explanation as a means.

Certainly, many people *have* put supreme value on explanation for its own sake, on pure theoretical science simply for the sake of greater understanding. Just as surely, others see value only in practical advances, perhaps because they tend to be more easily marketable. Even these latter people, however, would assign a high value to explanation insofar as it can be put into service for the advancement of technology and practical know-how.

In book I, chapter 1 of the *Metaphysics*, Aristotle comments on the value of explanatory knowledge *for its own sake*, calling it the essence of wisdom.

With a view to action experience seems in no respect inferior to art, and men of experience succeed even better than those who have theory without experience.... But yet we think that *knowledge* and *understanding* belong to art rather than to experience, and we suppose artists to be wiser than men of experience ... and this because the former know the cause, but the latter do not. For men of experience know that the thing is so, but do not know why, while the others know the 'why' and the cause (981a13-31, 1941, 689-690).

Explanation is at the heart of philosophy. A primary task of philosophy is to help us understand the world and ourselves. Western philosophy originates in the attempts of ancient Greek cosmologists to understand the world by finding a rational order in the basic elements of what exists. Socrates and Plato sought primarily explanations of who we are and why we think and behave as we do in terms of a human nature. Most philosophers have tried to explain how we come to know about the world, why the world has the most general features that it does, and why we make the value judgments we do.

Explanation is indispensable to science, even though a large part of science consists also in *discovery*. The discovery of the outer planets, the discovery of DNA, the discovery of galaxies and quasars, the discovery of tectonic plates—these and many other discoveries serve as markers of the great achievements of science. Science discovers not only things in the physical world but also *laws* describing the regular behavior of things in the world. It was necessary to discover the laws of light and optics and to master the techniques of lens-grinding before scientists could discover the outer planets. In turn, the discovery of the outer planets helped to confirm the laws of planetary motion. The discovery of tectonic plates spurs on the search for laws describing the behavior of these plates.

A basic task of science is also to explain, however. It explains, with a theory of gravity, why the laws of planetary motion are as they are, why the planets orbit as they do. One of the more famous episodes in the history of science involves the failure of one theory and the success of a rival theory theory in explaining unique features of Mercury's orbit. Scientists invoke tectonic plates and fault lines to explain earthquakes. The structure of DNA and RNA explains the inheritance of characteristics and patterns of inheritance.

We all discover and explain in the humdrum routine of daily life. Discovering patterns helps us deal successfully with our world, and providing explanations helps us to understand our world and ourselves. When I discover that the B-trains stop only at B-stations, I can get home more successfully, more efficiently, by waiting for an A-train to take me to my A-station. When I wait for thirty minutes without getting any trains,

there is a part of my world I fail to understand, some anomalous event that calls for a special explanation. Provided with the explanation that a man has thrown himself under a train, thereby disrupting service, I come to understand this atypical aspect of my world.

### Explanation and Existence

Two aspects of explanation are of special importance in what follows. First, people use explanatory considerations in figuring out what the world contains and how it behaves. They draw inferences on the basis of what would be the best explanation of something that demands an explanation. This use of explanation is not simple; it involves operations of subtly varying sorts in different areas of experience. For example, scientists (John Couch Adams in England and Urbain-Jean-Joseph Leverrier in France, 1845) inferred that there was another planet beyond Uranus (although none had ever been observed) because this seemed the best way (but not the only way) to explain certain anomalous characteristics of the orbit of Uranus. Nearly a year later astronomers observed the planet (September 23, 1846, Berlin), calling it Neptune. Scientists drew a similar conclusion from some anomalous characteristics of the orbit of Mercury, inferring that there was a planet (they named it Vulcan) between Mercury and the sun. (Vulcan has never been observed, and probably never will be, because Einstein gave us an explanation of the unexpected characteristics of Mercury's orbit without postulating an inner planet.) We might make the same sort of inference when we conclude that a person has been this way on the basis of footprints in the snow, or conclude that S



committed the murders on the basis that his glove was at the murder scene with his and the victims' blood on it.

A somewhat different but closely related sort of inference can lead one to conclusions about the existence of things one seems to observe. The best explanation of my visual tree-ish experience might be that I am seeing an existing tree. The best explanation of my experience of a rosy smell might be that I am smelling a rose. The best explanation of my auditory horn-ish experience might be that I am hearing a car horn. Alternatively, the best explanation of my visual experience of big squiggly things floating around the room might *not* be that there are big squiggly things floating around the room but rather that the experience is caused by very tiny squiggly things floating around my vitreous humor. (It sounds a bit strained to say that my experience is a visual experience *of* tiny things floating around my vitreous humor.) The best explanation of my experience of hearing my dead maternal grandmother speak to me in foreboding tones (as she tended to do) is not that my dead grandmother was talking to me just then but rather that I was having a nightmare. In general, we have experiences, and we can draw inferences on the basis of those experiences about what exists, whether it's Vulcan or the neurological activity characteristic of dreaming.

Most of the time, of course, we don't consider the proposition that we are being appeared to in a certain way and infer from it the proposition that there exists a certain object of which we are having an experience. When appeared to in a certain characteristic way, I just find myself believing that there is, for example, a computer on

my desk.<sup>1</sup> When the automatic beliefs are disrupted, however, we reevaluate the belief that the usual explanation is the best explanation of the experience. If the computer just vanishes into thin air, for example, I start fishing around for an explanation of my experiences *better* than the twofold claim that there *was* a computer on my desk and now there *isn't* a computer on my desk. The inferential moves we normally make smoothly and automatically sometimes rise to prominence for special consideration. Many of those inferences (perhaps all) depend on assumptions about how best to explain something.

The second central topic in what follows is the widespread disagreement about what explanation *is*, what distinguishes an explanation from a non-explanation, what distinguishes a better from a worse explanation, what distinguishes a full explanation from a partial explanation, and so forth. People disagree about what explanation is. This is most evident when they disagree about particular attempts at explanation. The explanation that giraffes have long necks because they help them reach food high in trees might be an acceptable explanation to one person, a completely unacceptable explanation (no explanation at all) to another person, and merely an oblique explanation<sup>2</sup> to a third person. These differences of opinion about explanation tend to be most obvious across different historical periods or across different contemporaneous cultures. Aristotle's explanation (see Ross 1949, 75) of why many objects fall down and some things (e.g.,

---

<sup>1</sup> Alvin Plantinga seems fond of reminding us in Reidian fashion of how we “find ourselves” believing all sorts of things. (See his 1993, p. 66, for example.) I don't restrict inference to *conscious* acts of forming a belief on the basis of other beliefs. I use ‘inference’ to refer primarily to any act of coming to believe something that one would, if pressed, defend or justify with an inference from other beliefs.

<sup>2</sup> Richard Dawkins calls such explanations “metaphorical” (1995, 98).

fire) go up, relying on the notion of “proper place,” appears to have few adherents today. Astrological explanations might be popular in some cultures (or, sub-cultures), but they are roundly denounced in scientific circles. Finally, one of the most hotly contested points about explanation in science and the philosophy of science today is the explanatory status of descriptions of low-probability quantum events. My central question concerns the justification, or rationale, one might have for advancing one doctrine of explanation and rejecting all the rest. What is our most basic source of evidence about what explanation is?

A doctrine of explanation not only determines acceptable methods of explanation but also determines what *calls* for explanation. Hence, shifts in the predominant doctrine of explanation can lead one generation to believe that a certain type of phenomenon that previous generations felt the need to explain calls for no explanation at all; the current generation considers it rather to be a *brute fact*. It is, therefore, difficult to understand the felt need on the part of scientists and philosophers of the past to explain certain phenomena unless we understand the doctrine of explanation that led to that *felt explanatory need*. It is also sometimes hard to understand their conclusions about what exists without first understanding the way considerations of explanation functioned in their *reasoning* about what exists.

The felt need for an explanation can sometimes come not from a doctrine of explanation but from some quite unexpected areas. Leibniz, for example, felt the need to explain the very existence of the physical world around us, including ourselves, in a way that seems alien to many people today (as we will see in Chapter 3). It appears that

his interpretation of modal statements was responsible, at least in part, for his feeling such an explanatory need. His interpretation of modal statements also seems to have determined to a significant extent how he went about satisfying that explanatory need. Aristotle felt the need to explain how knowledge is possible in a world of apparently universal flux. It appears that it was his reliance on a particular sort of linguistic formulation as a paradigm of the expression of empirical knowledge that was responsible, at least in part, for his feeling such an explanatory need. After Newton published his *Principia*, there was great debate among leading intellectuals (including Leibniz) about whether his postulated universal force of gravity demanded an explanation or could be left as an unexplained brute fact. In this case, the view that it must be explained simply faded away as the empirical success of Newton's theory proved overwhelming. Oddly enough, an explanation did eventually come, long after the felt need for the explanation disappeared, in Einstein's General Theory of Relativity. Now again, the felt need to explain the fundamental forces of nature (by reducing them, eventually in a Grand Unified Theory,) is widespread.<sup>3</sup>

### Intuitions about Explanation

The distinction between adequate and inadequate (partial, oblique) explanation, or between an explanation and no explanation, seems to be second-nature to most people. In general, I think people would regard the distinction between an explanation and a non-explanatory description as self-evident. This is not to say that a *theory of explanation*

---

<sup>3</sup> See, for example, Weinberg 1992.

is self-evident. On the contrary, it seems exceedingly difficult to come up with a satisfactory general characterization of what explanation is. I believe people have, however, very firm convictions about particular cases of explanation.

For example, a certain tower stands fifty meters high and the sun is at an elevation of  $45^\circ$ . Consequently, the tower's shadow extends for a length of fifty meters in the opposite direction from the sun. The height of the tower and the elevation of the sun, together with some simple trigonometry and the law that light travels in straight lines, *explain* why the shadow is fifty meters long. The length of the shadow, together with the elevation of the sun and the trigonometry, clearly does *not* explain why the tower is fifty meters high. People generally know an explanation when they see one, and if they see an explanation but do not know it is an explanation, they can tell precisely when, through the instruction of others, they come to understand the explanation. It is accompanied by that distinctive experience expressed by "A-ha, now I see it."

On the other hand, people are quite firm in their convictions about what is *not* an explanation, given that they have fully understood and considered the alleged explanatory account. For example, an astrological explanation of a political assassination is not even a candidate for an explanation to some people. Even if there were a strong correlation between some astronomical signs and political assassinations, most people would not even entertain the possibility that there is something in the astronomical signs that can play a significant part in any genuine explanation of a political assassination.

The *depth* of intuitions about explanation indicates that we have hit upon a topic of considerable philosophical significance. It is a sign that the notion of explanation is

probably pulling very heavy duty in our cognitive lives. Topics such as causation, probability, knowledge, meaning, time, reality, value, and duty, among others, exhibit similar characteristics. The intuitions about particular cases are strong, but stating the general principles proves exceedingly difficult. Most importantly, attempts at communication between parties with different sets of intuitions generate immense difficulties. The notion of explanation has the marks of being a member of what Kuhn (1970) calls a paradigm. Disagreement about what explanation is does not allow for the usual channels of resolution, such as would resolve disagreement about how many people live in the city, how best to get to Symphony Hall, what the boiling point of water is at five thousand feet, or even what it is to be a mammal. Explanation is at or near the foundation of our conceptual schemes; much is built upon it and depends on it.

### Summary of Chapters

In chapter 2, I specify what philosophical value there is in studying explanation, and of how best to study it. I claim that explanation is an important epistemological topic and that it is crucial for empirical knowledge in particular. Hence, the philosophy of explanation fits squarely within epistemology traditionally conceived. I am, however, quick to distance myself to some extent from the type of philosophical study of explanation found in a number of competing theories of explanation (primarily in the philosophy of science) over the past five decades.

Theorizing about explanation has proven to be significant, if only because it makes apparent the complexity of even the restricted concept of *scientific* explanation.

It does not, however, serve the function of clarifying the extensional meaning of the term 'explanation' as, say, a theory of justice can serve to clarify the extension of the term 'justice'. We do not benefit greatly from a theory of explanation as a tool for distinguishing explanatory from non-explanatory accounts of something. Mainstream theorizing about explanation in this century has not been useful primarily for giving us, by way of its successes, a deeper understanding of what explanation is and a more accurate capacity for identifying explanations. It has been useful, by way of its numerous failures and ongoing debates, for making clear the multiplicity and great intricacy of concepts of explanation and, consequently, drawing our attention to the complexity of our attitudes toward explanation and the multiplicity of ways in which we make use of appeals to explanation.

The primary pay-off of philosophizing about explanation is a better understanding of what we do intellectually with explanations when reaching conclusions about what exists. I claim that the earliest empiricist philosophers of explanation in this century, such as Carl G. Hempel, began with an acute sense of the importance of notions of explanation in our reasoning, and especially in our reasoning about what exists. That point was quickly lost, however, in the subsequent technical battles over conceptual analyses of 'scientific explanation' and of more general concepts of explanation.

In Chapters 3 and 4, I will contend that the only plausible arguments for existence claims rely essentially on considerations of best explanation, thereby showing the fundamental importance of the concept of explanation in our cognitive lives. Arguments for existence claims find their home in metaphysics and ontology. Hence, I shall look

at some classic arguments from metaphysics and general ontology, trying to show that the deepest and most plausible basis for argumentation in this area involves potentially controversial explanatory criteria. I do not claim that the great metaphysicians *took themselves* to be offering arguments that depend on substantive explanatory criteria. I merely try to show some plausibility for the claim that their background thinking on proper explanation and on what calls for explanation actually guided their metaphysical reasoning to a significant extent.

Before there was metaphysics in the West, there was the complex mythology of the ancient Greeks, and explanation flourished within these myths. Myths order experience, unify experience, provide coherence and meaning to experience. “It is likely that the ancients recognized certain intellectual problems and asked for the ‘why’ and ‘how,’ the ‘wherefrom’ and ‘whereto’.... [For them,] an account of [individual] events and also their explanation can be conceived only as action and necessarily take the form of a story. In other words, the ancients told myths instead of presenting an analysis or conclusions” (Frankfort et al. 1946, 6). Myths answer the pressing questions regarding *why* things happen as they do, and they sometimes even support general sorts of predictions. Myths provide explanations by describing a *general picture of the world*, a worldview. Such general pictures allow one to place an event in a *context*, and this allows one to draw out connections between events. A very important aspect of explanation of this sort is its capacity to render unpredictable and threatening events less frightening by subsuming them under familiar categories. Myths serve many other



purposes as well, of course, but they do provide reasons for much of what happens in life and they give what happens a meaning by putting it in a *structured context*.

Myths eventually gave way to more empirical philosophy, cosmology, and early science. From myth to philosophy there is both continuity resulting from the inheritance of some explanatory responsibilities and discontinuity resulting from a dramatic change in some fundamental ideas and methods of explanation. According to Edward Hussey,

What Hesiod does in the *Theogony* is like Presocratic thought in many ways. He attempts to create a concrete, unified and reasonable picture of the workings and history of the universe. He employs a single basic mechanism (the begetting of gods by gods) to achieve this picture. He is by no means constrained by tradition, and he is open to non-Greek ideas. Yet between Hesiod and even the earliest Presocratics there is a great gulf, created by a revolution in thought (Hussey 1972, 11-13).

The gulf is the result of a shift in ideas about the justification of explanatory accounts of the universe. Xenophanes introduces the new way of thinking.

In attacking head-on the traditional beliefs and putting forward utterly different ones, Xenophanes makes no appeal to the authority of a prophet or teacher, still less to any personal revelation. He relies entirely on certain general principles—certain conceptions of what it is reasonable or fitting that a god should be.... What seems especially important for the revolution in thought is the emergence of the concept of *law* as something determinate, impartial, and unchanging, and the spread of political equality. A debate between equals, in the popular assembly or the law-courts, must be conducted by appeals to general, impartial principles of law or reason—otherwise the parties will not be equal. The notion of ‘reasoned argument’ will begin to develop. There will grow up a habit of seeing particular situations as applications of a superior, abstract law. And law of this kind will be seen as the necessary arbiter of any complex whole in which order is apparent (*ibid.*, 14-15).

Even though the shift from the Greek mythology to philosophy can be explained in terms of many non-intellectual forces, its primary consequences involve intellectual life. According to Kirk, Raven, and Schofield,

[T]he transition from myths to philosophy, from *muthos* to *logos* as it is sometimes put, is far more radical than that involved in a simple process of de-personifying or de-mythologizing.... Rather it entails, and is the product of, a change that is political, social and religious rather than sheerly intellectual, away from the closed traditional society ... and toward an open society in which the values of the past become relatively unimportant and radically fresh opinions can be formed both of the community itself and of its expanding environment.... [I]n the speculative and cosmopolitan societies of Ionia, not least in Miletus itself, [the rational undertone of the Homeric tradition and the classificatory craft of Hesiod] took on a sharper form and were applied, without too much distraction from myths and religion, to a broader and more objective model of the world (Kirk, Raven and Schofield 1983, 73-74).

The pursuit of this broader model of the world takes place in philosophy and in science.

The earliest philosophers in the West, the Presocratics, were scientists of a sort; they are best described as cosmologists. We might call them scientists because they dealt with questions appearing to us today as scientific in content. They sought rational explanations of the movements in the skies and of the natural world around them, explanations identifying regularities and supporting predictions. They sought very general principles governing events in the world and the basic building blocks, the elements, of the things in the world. Yet their methods of inquiry were not what we recognize today as scientific methods. As already noted, they made a radical break from traditional myths and legend, but they did not have a methodology involving controlled experimentation. They relied quite heavily on *a priori* methods of reasoning about the world, and considerations of explanatory power played a very large part in this method. Early cosmological theories were tested against experience in a way, but only against common experience, not against special experience in a controlled experimental setting. The ultimate test of theory against experience was the test of explanatory power. Successful theories made better sense of the world of experience than their competitors.

Myths had been constructed in large part for the same purpose, but the shift from myth to philosophy was itself a shift from one set of criteria for explanatory goodness to another.

In chapter 3, I examine what Plato, reacting to the Presocratic philosophers, has to say about explanation. His entire doctrine of the Forms can be seen as a discussion of what is required of philosophical explanation of the world. In the *Phaedo* in particular, Plato has many explicit comments on explanation. We will see, for example, that Plato attacks a notion of explanation in terms of physical causes that he claims to find in the Presocratic philosophers. He recommends explanation in terms of the mind as a cause and what is *best* as a goal. “Somehow it seemed right that mind should be the cause of everything, and I reflected that if this is so, mind in producing order sets everything in order and arranges each individual thing in the way that is best for it” (*Phaedo*, 97c). His metaphysics of the Forms is then clearly recommended for its capacity to provide explanations of the requisite type. We have in Plato, I argue, a clear example of a metaphysician’s thinking about the nature of reality being strongly influenced by his prior commitment to specific criteria for explanation. I do not, however, argue that Plato would have interpreted his own thinking in this way.

Aristotle’s doctrine of explanation dominated philosophy and science for centuries. We shall see that his doctrine of the four causes functions as a doctrine of adequate explanation. Aristotle emphasized not the virtues we associate with experimental science but rather full explanation in terms of all of his four causes: final cause, formal cause, efficient cause, and material cause. This notion of explanation, it turns out, is firmly

rooted in Aristotle's metaphysical views on the nature of substance. A full explanation requires accounts of all four kinds of causes because they reflect the four principal aspects of the fundamental reality called primary substance. Aristotle's views on substance, however, also turn out to rely in important ways on his notion of explanation, but not explanation as a full account in terms of the four causes. His doctrine of substance is based in some way on his beliefs about what reality must be like in order to explain the possibility of the kind of knowledge, *epistemé*, that he presents in his works on nature.

This dual function for notions of explanation is of central importance, and is typical in theorizing about the most fundamental aspects of reality. This dual nature is also dangerous. In Chapter 3, I present the possibility of a vicious justificatory circle resulting from the mutual influence that exists between a doctrine of explanation and a metaphysical view of the world.

While it appears that Aristotle avoided a justificatory circle by using a notion of explanation to motivate his metaphysics *different from, and independent of*, the notion of explanation that is justified *by* his metaphysics, I present a case at the end of Chapter 3 that illustrates how a vicious justificatory circle can emerge. A vicious circularity shows up in one version of the argument from contingency, or the cosmological argument for the existence of God, based on Leibniz's Principle of Sufficient Reason. Without a prior commitment to the distinctive metaphysical claims involved, the distinctive views about proper explanation seem to be unsupported; without a prior

commitment to the relevant view of explanation, the metaphysical claims seem to be unsupported. The entire line of metaphysical reasoning is in this way undermined.

In Chapter 4, I examine the central place of concepts of explanation in some areas of contemporary ontological theorizing. Inference to the best explanation is widely regarded as the strongest type of argument one can give for existence claims. Appeals to explanation appear in arguments for theoretical entities, moral facts, mental entities, and ordinary middle-sized physical objects. Gilbert Harman is primarily responsible for focussing the attention of the philosophical community on inference to the best explanation as an important type of non-deductive inference.<sup>4</sup> I consider his claim that it is the basic form of non-deductive inference, without fully endorsing it. I do not find it necessary for my purposes to decide whether induction can be “reduced” to abduction or *vice versa*. I rather provide an argument at the end of this chapter that every non-deductive inference works on the basis of substantive (non-analytic) *presuppositions* about explanation.

In Chapter 5, I examine Carl Hempel’s classic analyses of the concept of scientific explanation: the deductive-nomological, inductive-statistical, and deductive-statistical models. Along the way, I indicate briefly how ontological or metaphysical assumptions can affect how one analyzes explanation. I argue that certain deterministic assumptions are responsible for some characteristic and highly controversial aspects of Hempel’s theories. I also raise questions about his method of conceptual analysis used

---

<sup>4</sup> There are Quinean antecedents to Harman’s ideas. See Harman 1967, especially pp. 350-1. See also Quine’s articles “On Mental Entities” and “Posits and Reality” in Quine 1966, and Chapter One (especially §5 and §6) of *Word and Object* (1960).

to generate theories of explanation. That very method of conceptual analysis provides some important clues as to what is really the most fundamental source of our intuitions about what explanation is.

We begin to see a line of influence from metaphysical assumptions to doctrines of explanation. One's general metaphysical view of the world and its ways will not uniquely determine some specific theory of explanation, but it will put substantial limitations on theorizing about explanation. This is so primarily because the ingredients of explanation have to come from the stock of entities and mechanisms existing in the world. In chapters 5 and 6, I study prominent examples of theorizing about explanation that are influenced by metaphysical assumptions without, however, arguing that there *must* be such influence or arguing about the specific nature of such influence.

In Chapter 6, I note that Wesley Salmon, one of the most influential theorists of explanation since Hempel, bases his theory openly on metaphysical theses about causation and "underlying mechanisms" that produce and propagate structure in the world. He also employs a methodology that gives us important clues about the most fundamental source of our intuitions about explanation. He claims that his theory of scientific explanation is superior to its competitors because his can account for the explanations of (apparently) irreducibly stochastic events in quantum mechanics and its competitors cannot. Some of those who hold the competing theories, however, do not admit that there are genuine explanations of irreducibly stochastic (and low-probability) events. So Salmon has to do something to defend the claim that such scientific accounts are genuinely *explanatory*, and his comments strongly suggest that he relies on a personal,

subjective sense of gaining increased *understanding* of the phenomena through those accounts.

In Chapter 7, I present my main conclusions. I argue that the familiar subjective and relative sense of *understanding* is indispensable as a guide to what explanation is, although some theorists of explanation have tried to downplay the significance of the subjective sense of understanding, searching rather for objective criteria to define 'explanation'. I argue that there is no clean conceptual analysis of 'explanation' either as the sole producer of the sense of understanding or as something that always produces a sense of understanding. It is impossible, however, to give an adequate account of what explanation is without any reference to the subjective sense of understanding. The phenomenon of understanding, or comprehension, also provides the key to showing just how one's metaphysical view of the world shapes one's account of explanation, for one's metaphysics, or general ontology, deeply influences what one finds understandable and what one finds incomprehensible. In general, one can understand a fact if and only if one can successfully integrate it into one's overall metaphysical view of the world. Thus, a presupposed metaphysics constrains how one thinks of explanation *by way of* its effect on one's subjective sense of understanding.

At the end of Chapter 7, I confront the danger of justificatory circularity in the interplay between explanation and metaphysics. I argue that the proper way to avoid such circularity is to ground ontological reasoning in a fundamental notion of explanation that is not justified only relative to some metaphysical system but is justified by its usefulness as a source of order and cohesion in the most basic levels of our experience.

I also stress the importance of keeping distinct (a) arguments based on a concept of explanation that derives from metaphysical assumptions and (b) arguments for metaphysical or ontological claims that depend on a notion of explanation logically prior to those claims. Arguments appealing to explanatory considerations are philosophically dangerous because of the relativity of doctrines of explanation to metaphysics. One should not assume that appeals to explanatory considerations are philosophically neutral and hence not in need of some philosophical defense or justification. Philosophers should be slower than they typically are to draw conclusions about what the world is like based only on considerations of what would or would not count, *for them*, as an adequate explanation of some aspect of their experience.

Theorists sometimes have to revise their thinking about how proper explanation proceeds on the basis of new developments in their fundamental ontology, and sometimes theorists have to revise their fundamental ontology as the result of attempts to explain new areas of experience. The most basic guide for all such revisions, however, is the primary concern all theorists have with fitting all aspects of their experience of the world into a coherent picture that leads smoothly from one area of experience into another and uncovers a rational order in the world that supports predictions and explanations. It is thus concerns about explanation, in the sense of providing order and structure, that support all speculations about objective existence. Fundamental explanatory needs provide the foundation and original impetus for all ontological reasoning.



## CHAPTER 2

### THE PHILOSOPHY OF EXPLANATION

#### Why Study Explanation?

The field of the philosophy of explanation, if there in fact is one, is very young. As recently as 1981, Alan Garfinkel was able to write that “a general philosophy of explanation is really needed.... It is something of a scandal how little attention has been paid to this need by traditional philosophy, which, with just a few exceptions, has had virtually nothing to say about the forms of explanation” (1981, 19). The philosophical investigations of explanation that do exist, along with the general conception of a philosophy of explanation, might seem to be vulnerable to charges of obscurity and irrelevance. Philosophical work on explanation has been the product mostly of philosophers of science, working in the past four to five decades, dealing with the topic of *scientific* explanation. Although there was once a dominant view in this area that enjoyed wide acceptance, philosophical work on explanation today exhibits considerable diversity. The discussions of explanation in the philosophy of science have for some time been quite technical, and one can easily begin to wonder what the point of studying explanation, in the form of obscure logical and epistemological puzzles, is. In a review of Wesley Salmon’s *Four Decades of Scientific Explanation*, Andrew Lugg expresses this view. “... [I]t is difficult to shake the impression that the debate has become amorphous,

some would even say aimless.... [I]t is not so easy to see the need for a general philosophical account of explanation.... In particular if the aim is—as it now seems to be—one of understanding scientific practice, why can't we get by with a perspicuous account of the multitude of ways in which scientists proceed?" (1991, 69) It seems appropriate, therefore, to begin with some brief comments on the philosophical significance of such a work and how it relates to more traditional areas of philosophical investigation.

The study of the forms of explanation fits squarely within epistemology and has a very important place in the study of logic, both inductive and deductive. As David-Hillel Ruben puts it, "The concept of explanation, like the concept of knowledge to which it is closely related, is an epistemic concept, and therefore has a philosophical location with the theory of knowledge, widely conceived" (1993, 1, see also Ruben 1990, 6). *Consequently*, it is of major importance in the philosophy of science. I believe that the value of studying scientific explanation, contrary to Lugg, does indeed lie in the larger goal of understanding scientific *practice*, and the value of a general philosophical account of explanation lies in the larger goal of understanding theoretical practice and theoretical knowledge in general. We cannot gain an understanding of scientific practice, as Lugg claims, with just a perspicuous account of the ways scientists proceed. It is entirely conceivable that one have a perspicuous view of how scientists proceed, a complete inventory of their *activities*, without being able to interpret, or give a meaning to, what they do.

I have no idea whether Lugg meant to bring to mind Wittgenstein's comments on the importance of a perspicuous view of philosophical grammar, but his mention of a perspicuous account suggests some of Wittgenstein's remarks. Nowhere, however, does Wittgenstein say that a perspicuous account of how certain people proceed is adequate for philosophy. He always says that perspicuity in *grammar* is the main objective. The following remark is typical.

Savages have games (that's what we call them, anyway) for which they have no written rules, no inventory of rules. Let's now imagine the activity of an explorer, who travels through the countries of these peoples and takes an inventory *of their rules*. This is completely analogous to what the philosopher does (1991, 16, emphasis added).

The concept of perspicuous representation is of fundamental significance for us.... Above all, our grammar is lacking in *perspicuity* (1991, 11; cf. 1953 #122).

To understand scientific practice we need not only an account of how scientists proceed but also an account of the rules that guide them. We need to *explain why* scientists proceed as they do, and this requires an understanding of the way scientists *reason*. More generally, if one wants to understand how theorists (i.e., all of us) proceed, one must understand the rules theorists follow, i.e., how they infer, and one cannot get very far in understanding inference without coming up against the topic of explanation, for, as I shall try to show, the need to meet explanatory demands guides many of our most important inferences.

Wesley Salmon defends the importance of studying scientific explanation in a somewhat different way. Salmon responds to Lugg, "My answer is that the aim is not one of understanding scientific *practice*—it is the aim of understanding scientific *understanding*. Although it is obviously important to take note of the ways in which

scientists proceed, we need a great deal more. Scientific explanation of what transpires in our world is a complex matter, partly because we live in a complicated world. The philosophical problem is to clarify what is meant by *scientific understanding of the world*" (1993, 16). Ten years earlier, in his book *Scientific Explanation and the Causal Structure of the World*, Salmon said the same about the value of philosophical work on explanation: "Our aim is to understand scientific understanding. We secure scientific understanding by providing scientific explanations; thus our main concern will be with the nature of scientific explanation" (1984, ix). In this way Salmon argues in defense of "the value of scientific understanding."

Why do we need a general philosophical account of explanation? Why do we need to explain explanation? One cannot fully understand inference without an account of explanation. Why is an understanding of inference important? Because inference is an essential part of most, if not all, of our knowledge of the world. When it comes to matters epistemic, explanation is of central importance. As Karl Popper has put it, "In seeking pure knowledge, our aim is, quite simply, to understand, to answer how questions and why questions. These are questions which are answered by giving an explanation. Thus all problems of pure knowledge are problems of explanation" (1972, 263). So also with *practical* knowledge; inference goes on constantly, regardless of whether one is aware of one's inferring. Explanation is at the heart of virtually all of our decision-making and should thus be of central importance to anyone interested in making good decisions.

More specifically, explanation is at the heart of philosophy. Philosophy is, in all its manifestations, primarily an activity of giving and comparing explanations. One cannot understand what philosophy is or how it works without understanding explanation. Competing philosophical theories are evaluated by their power to explain something. Even though philosophy often has somewhat different ostensible forms, it is always reducible to problems of explanation. For example, Paul Moser (1993, 19) has said that philosophical questions typically take the form, “What is *X*?” Philosophers, somewhat like scientists, investigate the nature of *reality*—the nature of knowledge, justice, goodness, meaning, validity, etc. Yet answers to such questions (that is, philosophical theories) are always evaluated as explanations, and the evaluation is *via* our experience of instances of *X*. They are tested against actual experience of *X*'s and against our intuitions regarding hypothetical instances (intuitions based in part on actual experience and in part on background theories, even though such intuitions are often called “pre-theoretical”).

Philosophical theories are tested primarily for explanatory power, however, not for (empirical) accuracy, because philosophical theories, by their very nature, cannot be evaluated by comparing what they imply about how the world is with how the world is observed to be. (Of course, if a philosophical theory has some definite empirical consequence that is contrary to actual experience, it is considered to be flawed—assuming, however, that it does not offer an *explanation* of its empirical inadequacy.) If, for example, one philosophical theory predicted the observable bending of light rays near the sun and another predicted no bending of light, those theories would

be, insofar as they have precise empirical consequences, scientific. Philosophical theories do not make predictions that allow for crucial tests, otherwise we would see the same signs of progress in philosophy as we do, say, in physics.<sup>1</sup> We do not find, for example, ethical theories implying the impossibility of Hitler's behavior in his Final Solution, although we do find some theories explaining more satisfactorily than others why it was wrong. Philosophical theories are meant to explain experience and the world that we experience, and they are evaluated *as* explanations. Since they do not make predictions allowing for empirical tests, they do not explain experience by providing for a deduction of an experienced event from some general law. Philosophical theories do not explain experience in the sense that Einstein's theory explains our experience of the bending of light rays and Newton's theory fails to explain the same experience. So we must be using some other sense of explanation when we say that explanation is at the heart of philosophy. Philosophy often does not explain *why this* (rather than that) but explains *how-possibly this*. It often does not explain interpreted experience but explains *by* interpreting experience.

I am arguing that there must be a sense of explanation other than the one Carl Hempel developed in his covering-law models. There is, of course, a sense in which we can explain why this wire conducts electricity by stating that this wire is copper and all copper conducts electricity. There are cases, however, where we face a choice of two or more competing (i.e., contrary) theories that all explain all of the relevant data in the

---

<sup>1</sup> There are, of course, philosophical problems regarding the possibility of crucial tests in the natural sciences as well, but for now I am assuming a naive falsificationism for scientific theories, a falsificationism I do not endorse.

covering-law sense of explanation, but we still distinguish these theories as being better or worse explanations of the data. There must be some criteria for explanatory goodness, criteria *internal* to theorizing in general, other than those embodied in the covering-law models. Understanding those criteria for explanatory goodness, and the accompanying sense of 'explanation,' is crucial for understanding how we select what theories to believe. Hence, it is crucial for understanding how we select what *philosophical* theories to believe, i.e., how we evaluate the products of philosophizing, i.e., how we *do* philosophy.

Unlike ordinary science, philosophy can seem perplexing and even pointless. One can see the point of considering theories about the behavior of light. If they say different things about the behavior of light, then at most one of them can be true. We can eliminate the false theories by comparing their claims about the world with observations of the world. If a theory says that light rays will not bend as they pass near the sun, and we look near the sun and observe the bending of light, then the theory is wrong at least on this point. In this way we get closer to the truth about how the world is. One of the perennial complaints about philosophy (metaphysics in particular) is that it goes beyond the bounds of any possible experience. More precisely, this means that philosophy deals with empirically equivalent yet distinct theories. Empirically (or, predictively) equivalent theories are consistent with exactly the same sets of possible experience. Metaphysical realism and metaphysical idealism, for example, are empirically equivalent. There is no crucial empirical test to distinguish them. No matter what one ever could experience, both realists and idealists will be able to say, "Yep, just as it should be, we

can give a metaphysical explanation of why your experience has the character it does.” This is so whatever your experience happens to be like! Realism and idealism are evaluated, however—as all philosophical theories are—by their explanatory power, and philosophers expend much energy defending one or the other as the better explanation. In this way philosophy is supposed to make sense of our world, to explain why the world is the way it is.

### What is Explanation? — The Importance of Examples

Most books and articles on explanation are full of *examples*. This is obvious at least as far back as the classic Hempel-Oppenheim paper, “Studies in the Logic of Explanation” (1948, 246-247), and it goes back even to Aristotle. Examples are very important to any study of explanation, because what we are studying is intimately related to the experience of *understanding*, for which the final authority is always the first-person. There is a *feeling* associated with explanation, insofar as explanations are intended to produce understanding, and understanding produces an *awareness* of understanding. Such examples are typically meant to provide paradigm cases for the reader to keep in mind as a source of clear intuitions of what is and what is not an explanation, and what makes a better or worse explanation. Hempel, for example, follows his list of examples with considerations of “the basic pattern of scientific explanation,” gleaned from the examples. Examples can also constitute an arsenal for an author to use, not only to support a favored theory of explanation, but also to point out weaknesses in a competing theory of explanation. It appears that most, if not all, authors in this area take as their



ultimate criterion of success for theories of explanation adequacy to a host of considered judgments, or intuitions, about particular examples of explanations. Michael Friedman expresses a typical view while dealing specifically with scientific explanation, “Most, if not all, scientific theories that we all consider to be explanatory should come out as such according to our theory.... Although it is unreasonable to demand that a philosophical account of explanation should show that every theory that has ever been thought to be explanatory really is explanatory, it must at least square with most of the important, central cases” (1974, 13).

Some examples have become staples of many works on explanation, classic counterexamples to prominent theories of explanation. For example, the case of the flagpole and its shadow has had such exposure that it is now synonymous, for those in the field, with the problem of the asymmetry of explanation for Hempel’s D-N model of explanation.<sup>2</sup> The example states that one (clearly) can explain the length of a flagpole’s shadow at a particular moment by giving the flagpole’s height, the position of the sun, and the law that light travels (for the most part) in straight lines. The example further states that one (clearly) cannot explain the height of the flagpole by giving the length of

---

<sup>2</sup> The example of the flagpole is often attributed to Sylvain Bromberger, although he never published it. He explains, “I am often credited with [the flagpole] example, and ‘Why-questions’ is usually cited as its original source. Ironically, though I did think of the example, it appears nowhere in ‘Why-questions.’ In fact it appears in none of my published papers! It is described in papers by Hempel in which he generously recalls that I once presented him with the difficulty. ‘Why-questions’ has obviously not been read by many people who refer to it” (Bromberger 1992, 8). “Why-questions” does, however, present examples (one involving the Empire State Building and another involving a telephone pole with a taut wire running from its top to the ground) having the same import (see Bromberger 1966, 83 and 98).

the shadow, the position of the sun, and the law of the rectilinear propagation of light. (For an interesting twist on this example, meant to exploit yet others of our intuitions, see van Fraassen 1980, 132-134.) Philosophers of science take this as a counterexample to the D-N model of explanation because the D-N model would count each case as an explanation, since each is a case of a valid deduction of the fact-to-be-explained from a general law and specific initial conditions. The counterexample shows that the D-N model violates a well-established considered judgment about explanation and is therefore a defective theory of explanation.

I shall use examples of explanations not primarily as standards of evaluation for competing theories of explanation (although they do, *for us*, possess this kind of authority), but rather as subjects of philosophical investigation. I want to analyze them rather than build theories on them. After all, why would one *want* a theory of explanation that fits with all our intuitions when the intuitions themselves are secure enough to refute any theory that does not count them as explanations? Further, if one takes a privileged set of intuitions regarding explanation as basic one cannot address, for example, the question of what brings about fundamental changes in a group's intuitions about explanation. What guides one group of scientists to regard occult entities such as gravitational force as anathema to explanation and another group of scientists to embrace them as explanatory?<sup>3</sup> I want to discover the *sources* of such considered judgments—sources undoubtedly consisting of further intuitions about explanation, but intuitions of a much more general nature and serving perhaps unrecognized purposes.

---

<sup>3</sup> See Kuhn 1970, 104-106.

I also intend to use examples not for the sake of unification in a single theory of explanation but rather for exhibiting diversity and for making a particular distinction between two fundamentally different explanatory structures. Ernest Nagel did something similar in *The Structure of Science* where he begins with a list of ten detailed examples of various explanations and goes on to distinguish within those examples four distinct types of explanation (1961, 15-26; Nagel's four types are deductive, probabilistic, functional or teleological, and genetic explanations). One of my main objectives is to identify two explanatory structures, or strategies, and argue that they mix in varying proportions in both philosophical and scientific explanations. The distinction between the two different explanatory strategies is conceptually more fundamental than any proposed distinction between scientific and philosophical explanations.

For this reason, my list of sample explanations comes from science and from philosophy. Some might already find this objectionable. Philosophers of science typically make some effort to begin with a demarcation (perhaps somewhat vague) of *scientific* explanation from other types of explanation. Carl Hempel, I believe, set the precedent for such taxonomic concerns.

[T]o put forward the covering-law models of scientific explanation is not to deny that there are other contexts in which we speak of explanation, nor is it to assert that the corresponding uses of the word 'explain' conform to one or another of our models. Obviously, those models are not intended to reflect the various senses of 'explain' that are involved when we speak of explaining the rules of a contest, explaining the meaning of a cuneiform inscription or of a complex legal clause or of a passage in a symbolist poem, explaining how to bake a Sacher torte or how to repair a radio. Explicating the concept of scientific explanation is not the same thing as writing an entry on the word 'explain' for the *Oxford English Dictionary*. Hence to deplore, as one critic does, the "hopelessness" of the deductive-nomological model on the ground that it does not fit the case of

explaining or understanding the rules of Hanoverian succession is simply to miss the intent of the model (1965, 412-413).

Some might wonder about the prospects for trying to draw philosophically significant conclusions about explanation from a set of examples that obviously straddles the boundary between science and philosophy. The usefulness of such a comparative study of scientific and philosophical explanations is a large part of what I want to establish. For now, let it suffice to say that the boundary between science and philosophy is not as clear as many scientists and philosophers would claim, and examination of the common elements in scientific and philosophical explanations should help to clarify the similarities and differences between philosophy and science. In fact, I believe that an adequate study of philosophical explanation must deal with scientific explanations, both past and present, for as Paul Humphreys claims, "Philosophical explanations seem most frequently to occur at the interface between science and philosophy" (1984, 174). Examples of explanation will illustrate the complex connections between the search for explanations and claims about objective existence. They will also present multiple styles of explanation.

1. *Why is chalk white?* Steven Weinberg explores this question at length in his *Dreams of a Final Theory* (1992, 21-31). His objective is to defend a particular thesis about the convergence of all scientific explanations in the direction of a yet unformulated final theory. I merely lift it as an example of a scientific explanation. The simplest explanation is that the chalk is not any other color (Weinberg, 21). More specifically, this means that the chalk does not absorb very much electromagnetic radiation from

frequencies within the visible portion of the spectrum. An explanation of chalk's reflection and absorption pattern gets us into the atomic structure of calcium carbonate, the explanation of behavior at the atomic level gets us into quantum mechanics, and explanation of the quantum-mechanical equations gets us finally to the standard model of elementary particles and forces. In this case, we explain an observed phenomenon (the chalk's appearing white) by describing the reality observed (the chalk) at a level beyond observation, and the laws that operate at that level of reality. In other words, we tell a story about an unobserved reality<sup>4</sup> and cite laws connecting it to what is observed. This is *explanation by reduction*, and an essential ingredient of the explanation is the ontology of the atomic and sub-atomic world. The explanation does not work if the ontology is rejected, but the power of the explanation *supports* the ontology.

2. *Why does a helium-filled balloon move forward (in the direction of acceleration) when suspended in an accelerating cabin?* Wesley Salmon provides this example, one he claims comes from an actual incident involving a physicist friend of his talking with a child holding a balloon in an airplane just prior to take-off (1989, 183). Why does this event make us *feel the need* for an explanation in a way, for example, that the movement of a toy car toward the back of the cabin would not? It seems counter-intuitive, because the pull of physical bodies backward is something we are intimately familiar with for the case of our own bodies, something we can compare directly to the

---

<sup>4</sup> I do not say unobservable. I want to sidestep controversies over the observable/unobservable distinction for now. It is sufficient to point out that the entities mentioned in the explanation are not observationally discriminable in the original experience-to-be-explained.

pull of a toy car backwards. The behavior of the balloon is *disanalogous* in a significant way from our common experience of a backward pull.

Salmon offers two explanations of this event. First, one can explain that the movement of the back wall of the airplane cabin causes a pressure gradient in the air in the cabin from front to back. The pressure gradient causes a greater force on the back of the balloon than on the front, thereby causing it to move forward. Another explanation involves Einstein's principle of equivalence. This principle says that an acceleration is physically equivalent to a gravitational field. The acceleration of the plane is then physically equivalent to placing a gravitational force behind the plane. Since the balloon tends to rise in the earth's gravitational field, it tends to move forward in the plane's new field (i.e., away from a gravitational source behind the plane). I would like to offer yet another explanation, one similar to the explanation based on Einstein's principle of equivalence but one that could easily have been given long before Einstein. Helium-filled balloons have a tendency (as does fire, for example) to move in a direction opposite to massive bodies (such as our own). When we sit still, we can feel the tendency of our bodies to move downward, and we observe the balloon's tendency to move upward. When the plane (or car of a train, if you like) moves forward, we feel the tendency of our bodies to move backward. So, by analogy, we should expect the balloon to exhibit a tendency to move in the opposite direction, that is, forward.

The first explanation is a *causal explanation*. It explains in terms of the relevant causal mechanisms responsible for bringing about the event-to-be-explained. In this way

it relies on prior theses about the causal structure of the world.<sup>5</sup> The second explanation is an *explanation by unification*,<sup>6</sup> since it appeals to a general physical principle to show how this particular event “fits into the universal scheme of things” (Salmon 1989, 183). The second explanation relies, however, on a sophisticated physical principle, a principle that depends on the backing of a lot of physical theories, specifically relativistic theories of space-time. The third explanation is also an explanation by unification, but it relies only on more common experiences and proto-theories. We could usefully label it “*explanation by analogical unification*,” since it depends on the *analogy* between (rather than the physical equivalence of) the pull of gravity and the pull of acceleration. It answers the initial felt need for an explanation by replacing the disanalogy between our own bodies (and most other objects) being pulled backward and the balloon (apparently) being pulled forward with the analogy of our bodies and the balloon being pulled in opposite directions both at rest and during acceleration. How do these three types of explanation differ in the way they explain, and in virtue of what are they all explanations?

3. *How can something remain numerically identical through change?* This question is as old as Western philosophy itself. In the fifth century B.C.E., Heraclitus apparently raised the problem of change by claiming, “Upon those that step into the same rivers different and different waters flow... They scatter and ... gather ... come together

---

<sup>5</sup> Obviously, this is the type of explanation that Salmon is most interested in throughout his book.

<sup>6</sup> On explanatory unification see Philip Kitcher 1976, 1981, and 1989. See also Michael Friedman 1981.

and flow away ... approach and depart” (fragment 12 in Kirk, Raven and Schofield 1983, 194-5). Plato paraphrased Heraclitus: “Heraclitus is supposed to say that all things are in motion and nothing at rest; he compares them to the stream of a river, and says that you cannot go into the same water twice” (*Cratylus* 402a; Plato 1961, 439). Heraclitus seems to challenge the claim that anything can remain the same through change, since everything is in flux like the river and is *not* the same from one instant to the next. Thus, Heraclitus was at least taken by some (including Plato) to have believed that change is the only reality and there is no identity through change.

Parmenides, on the other hand, had denied the existence of change, claiming that the only reality is eternally self-identical and changeless: “But changeless within the limits of great bonds it exists without beginning or ceasing, since coming to be and perishing have wandered very far away, and true conviction has thrust them off. Remaining the same and in the same place it lies on its own and thus fixed it will remain” (from fragment 8 in Kirk, Raven and Schofield 1983, 251). Zeno of Elea had challenged the very concept of change with his paradoxes of motion and of plurality.

Plato and Aristotle certainly took it as one of their main tasks to explain the possibility of change (see, for example, Aristotle’s *Metaphysics* XII, 2 and *Physics* II, 3-7). Plato offered his theory of the Forms as an answer—change is explained in terms of the changeless. Aristotle used his distinctions between potentiality and actuality, between matter and form, between primary and secondary substance, and between essential and accidental features to construct an answer. Primary entities are mixtures, and identity through change is just the replacement of one part with another while



something else (the essence and the material substratum) is held constant. Plato and Aristotle offer *metaphysical explanations*, and we shall study them in more detail in the next chapter.

Anyone who takes this problem seriously says something about a substratum remaining the same while some aspect of it changes. From the seventeenth century on, however, not many people took this problem seriously. The central problem for science became the problem of giving accurate quantitative descriptions of various forms of change in terms of universal (changeless) physical laws. This is a shift in intuitions about what *calls for* explanation. While the problem of the physical possibility of change was either explained in terms of a substratum or else ignored, the conceptual difficulties raised by Zeno lingered on.

4. *How is motion possible?* Zeno's paradoxes raised conceptual difficulties that were not adequately resolved until this century. "It is perhaps amusing to be confronted by a simple argument which purports to demonstrate the unreality of something as obviously real as motion; it is deeply intriguing to find that the resolution of the paradox requires the subtlety of modern physics, mathematics, and philosophy" (Salmon 1970, 5). The solution required clarification of the mathematical concepts of function, limit, continuity, derivative, anti-derivative, differential, and definite integral. The solution also required the removal of infinitesimals (Leibniz's "ghosts of departed quantities," [Coffa 1991, 25]) from the calculus in order to free it from a foundation in intuitions of space and motion (Salmon 1970, 20-21). "Until these foundations had been laid, certain

logical and mathematical questions concerning Zeno's paradoxes could not be given satisfactory answers" (Salmon 1970, 21).

These developments led to the "at-at theory of motion," applied directly to Zeno's paradoxes by Bertrand Russell (Salmon 1970, 23). This theory says:

Motion consists in being *at* a particular point *at* a particular time, and regarding each individual position at each particular moment, there is no distinction between being at rest at the point and being in motion at the point. The distinction between rest and motion arises only when we consider the positions of the body at a number of different moments. This means that, aside from *being at* the appropriate places at the appropriate times, there is no *additional* process of *moving* from one to another (Salmon 1980, 41).

A. J. Ayer expresses the solution as follows: "The answer is that [the arrow's] being in flight simply consists in the fact that over a continuous period of time it occupies a continuous series of places. If we are asked how it gets from one place to another, our answer must again be that its getting from one place to another simply consists in the fact that it occupies some intermediate place at any given intermediate time" (Ayer 1973, 19).

Russell's discussion of Zeno's paradoxes of motion in "The Problem of Infinity Considered Historically" is a modern classic. Written in 1914, it brought together a good deal of the then-recent historical scholarship on Zeno, and discussed the paradoxes in the light of the most up-to-date mathematical knowledge available. It seemed to many scientifically minded philosophers that Russell had, in this and other writings, cut the ground completely from under the metaphysical interpretations of Zeno's paradoxes by such philosophers as Bergson and Bradley, and had provided something approaching a definitive resolution in the light of modern logic and mathematics (Salmon 1970, 26-27).

Assuming for the moment that Russell's treatment of the problem is adequate, the resolution of the problem is an explanation of how to apply concepts associated with non-denumerable infinities to experience without running into contradictions. One does not offer a physical, causal explanation here, as one would explain how the motion of an

automobile is possible with a physical explanation of the workings of the internal combustion engine. The problem here is a problem with a fundamental concept of measurement—the application of the system of real numbers (or even just the rational numbers) to physical reality in the act of measurement. The problem is at bottom a conceptual problem calling for a *conceptual explanation*. A conceptual problem consists either in an internal inconsistency in a set of concepts or an inconsistency in concepts together with empirical data, and conceptual explanation removes the inconsistency by redefining one or more concepts to remove the inconsistency. Russell's at-at theory provides a concept of motion that is consistent with the concept of an infinite number of identifiable tasks involved in motion.

In the late nineteenth century, another major conceptual problem arose, one that involved a conflict between concepts and experience. The concepts were those of time and space, and the experience was provided in the null-result of the Michelson-Morley experiment.

5. *How can the velocity of light be finite but constant in any direction and for every observer?* Einstein's special theory of relativity provides an explanation to answer this question. Like Zeno's paradoxes, this problem is a conceptual problem. Einstein writes, "Experience has led to the conviction that, on the one hand, the principle of relativity [of measured velocity] holds true and that on the other hand the velocity of transmission of light *in vacuo* has to be considered equal to a constant  $c$ . By uniting these two postulates we obtained the law of transformation for the rectangular coordinates  $x$ ,  $y$ ,  $z$  and the time  $t$  of the events which constitute the processes of nature.

In this connection we did not obtain the Galilei transformation, but, differing from classical mechanics, the *Lorentz transformation*” (Einstein 1961, 42). Actually, Einstein did not unite the principle of relativity and the constant velocity of light; he replaced the principle of relativity with his theory of relativity, and in doing so he replaced the old concepts of absolute space and absolute time with their relativistic counterparts. Experience had led us into contradiction with our concepts, and in this way it made us feel the need for an explanation. The Michelson-Morley experiment had convinced us (given certain background assumptions) that the velocity of light is a constant  $c$ . The principle of relativity, however, tells us that the velocity of light should be different relative to different frames of reference if those frames are in motion relative to each other. On the one hand, we are led to believe that the velocity of light is relative to a frame of reference; on the other hand we are led to believe that the velocity of light is not relative to a frame of reference. Einstein removes this contradiction by changing the definition of ‘measurement of velocity,’ i.e., changing the definition of velocity. We could call this a *conceptual-empirical explanation*, emphasizing the way in which specific experiences led us into apparent contradictions.

6. *How are the results of the two-slit experiment possible* (e.g. the Stern-Gerlach experiment described in Hughes 1989, 1-8)? This is really the question of how to explain quantum theory, the second great theoretical development of the twentieth century. Quantum mechanics has probably been more resistant to attempts at explanation than any other set of scientific laws in human history. It is certainly the most empirically successful set of laws that have resisted explanation. Even the resistance of quantum

theory to explanation seems to have a special character to it. Richard Feynman expresses the exceptional inexplicability of it: “Do not keep saying to yourself, if you can possibly avoid it, ‘But how can it be like that?’ because you will get ‘down the drain’, into a blind alley from which nobody has yet escaped. Nobody knows how it can be like that” (1965, 129).

Quantum theory presents the most problematic quest for explanation we face today, because it challenges many deeply-rooted convictions about what explanation is. Quantum theory challenges *determinism*, and this is a fundamental challenge to traditional assumptions about explanation that are based on deterministic assumptions. Philosophers and scientists have begun discussing probabilistic causation and corresponding styles of explanation. The attempt to give an account of explanation that will count quantum theory as explanatory of irreducibly stochastic events at the sub-atomic level will probably produce revolutionary developments in our most fundamental ideas about what explanation is. We will return to controversies over quantum explanations in later chapters.

7. *Why is the symmetry between the weak nuclear force and the electromagnetic force broken?* This is an explanation-seeking why-question from science that has not yet been answered satisfactorily. “The most important thing that we still do not know about the standard model is just what causes the breaking of the electroweak symmetry” (Weinberg 1992, 196). Past attempts to provide an explanation are illustrative, however. Physicists have introduced a new field into the standard model just for the sake of explaining this particular broken symmetry, but the scientific community at large remains

unsure whether the account of symmetry breaking in terms of this field is correct. As Weinberg notes, “This is not the first time that physicists have proposed the *existence of a new field or particle* in order to satisfy some theoretical requirement” (*ibid.*, emphasis added). So this question and the proposed answer tell us something about a very common explanatory strategy in science, a strategy that has an impressive track record. It is a clear example of an explanatory need being met by an *ontological postulation*.

8. *Why is there something rather than nothing?* Although many people consider this to be a pseudo-question, many today still consider it to be an explanation-seeking why-question. Derek Parfit claims, “It might have been true that nothing ever existed: no minds, no atoms, no space, no time. When we imagine this possibility, it can seem astonishing that anything exists. Why is there a universe? And things might have been, in countless ways, different. So why is the Universe as it is?” He adds, “These facts cannot be causally explained” (1992, 3). In *Philosophical Explanations*, Robert Nozick examines several possible answers to the question, and although he does not single out one answer as correct, he says that “it is late enough in the question’s history to stop merely asking it insistently, and to begin proposing possible answers” (1981, 116). There have, however, been proposals enough in the history of philosophy on how to answer this question. Answers have postulated a necessary being beyond the universe of *contingent beings* that explains the universe. Spinoza is one representative of the philosophic drive to answer this question. Stuart Hampshire claims, “There is no doubt that Spinoza regarded Nature or the world as something the existence of which has

somehow to be explained; the mere existence of things of any kind seemed to him to constitute a problem, and a problem which it is the work of the philosopher to solve. The scientist explains the existence and properties of particular things, and kinds of thing, within the universe; but the existence of the Universe itself, or the fact that *anything* exists, seems to the metaphysician to require an explanation of another kind” (Hampshire 1951, 216-7).

F. C. Copleston, in a 1949 debate with A. J. Ayer, gives contemporary expression to the belief that this is a typical philosophical question, and he mentions an element common to most philosophical answers to it.

[T]he scientist, precisely as scientist, does not raise, for example, the question why anything is there at all. To raise this question is, in my opinion, one of the functions of the philosopher. You may say that the question cannot be answered. I think that it can; but, even if it could not be answered, I consider that it is one of the functions of the philosopher to show that there is such a problem.... [I]f I ask whether anything lies behind phenomena, whether anything is responsible for the series, finite or infinite, of phenomena, the answer—supposing that there is an answer—must, in my opinion refer to a reality beyond or behind phenomena. But in any case to ask why any finite phenomena exist, why there is something rather than nothing, is to ask a different sort of question from the question why water tends to flow downhill rather than uphill (Ayer and Copleston 1949, 144-5).

I introduce this question not to study proposed answers to it as much as to study the question itself and issues it raises about explanation. It seems to assume certain things about *explanatory regress*, for example. Apparently, the question is motivated by the belief that an endless regress of why-questions, explanations, further why-questions and further explanations is unacceptable. The nature of explanatory regresses, it turns out, is very important. The question also touches on the relationship between necessity and explanation, because, historically, answers to this question have appealed to

necessary being. The question also raises the issue of whether it is things (beings) that explain and need to be explained, or states of affairs, or linguistic entities, or perhaps something else. The very question of whether *this* question is really an explanation-seeking why-question elicits some very general and illuminating intuitions about how one explains. It also yields good examples of how *variable* felt explanatory needs and presupposed standards or criteria of explanation lead to metaphysical conclusions.

### What We Don't Need from Theories of Explanation

Most theorists of explanation have taken their task to be one of formulating a list of individually necessary and jointly sufficient conditions for something's being an explanation of a certain sort. Peter Achinstein asserts that the main theorists of explanation "are concerned primarily with providing [necessary and sufficient] conditions for E's being a good explanation of q. In general, the accounts have this form:

E is a good explanation of q if and only if E is a proposition (or argument, which is a certain kind of compound proposition) satisfying conditions C" (1983, 13).

For example, most philosophers object to Hempel's three models of scientific explanation (D-N, D-S, and I-S) on the grounds that they fail to give a list of sufficient conditions for scientific explanation. This is the point of the many counterexamples in circulation, such as the flagpole and its shadow. There is controversy over whether his models provide necessary conditions for scientific explanation,<sup>7</sup> although most philosophers think

---

<sup>7</sup> Michael Redhead argues that the conditions given in the D-N model are necessary for complete scientific explanations (see his 1990, 136). Opposition to this claim has been widespread ever since Salmon 1965 argued that some statistical scientific



they do not. Given such concerns, considered judgments (or so-called pre-theoretical intuitions) about explanation are taken as *data* for which any adequate theory of explanation must account. Theorists take their task to be the construction of a theory adequate to the data presented in considered judgments of cases of explanation.

Yet no one has explained why we need such a theoretical account of explanation. It is indeed a task that requires the utmost intellectual efforts. The literature from the last four decades on this topic presents elaborate theoretical constructions, followed by clever counter-examples, followed by ingenious tinkering with the theories, accompanied by competing theoretical accounts, and so on. Clearly, this is a good way for one to exercise one's mind, if one has a taste for pondering complex puzzles. But what does all this theorizing get us, other than some valuable intellectual exercise?

Let's ask, more generally, why one theorizes. Naturally, theories are supposed to fill an explanatory role; so one theorizes in order to understand. Setting that function aside for the moment, we can identify some more concrete functions of theorizing.

In empirical matters, theorizing seems to perform two functions. First, theorizing should generate *predictions* of future experiences. As a consequence, theories indicate what experiences to seek in order to *broaden* the scope of our (collective) interaction with the world. Finding a theory that generates correct predictions has obvious instrumental value. It was theory that led us to begin smashing atoms, for example; theory told us that there might be something interesting inside. Broadening the scope of our interaction with the world also has value, an apparently intrinsic value. The value,

---

explanations fail to satisfy Hempel's conditions for either D-N, D-S, or I-S explanations.

other than instrumental, of experiencing such things is hard to explain. It is, however, akin to the value of experience in general, when experience is taken as experience of a real (rather than illusory) world. Hallucinatory experiences, if they are pleasant, can have some value, but experience taken to be *reliable* has a special importance. Reliable experience is important to us insofar as the world we suppose ourselves to experience is important to us. Experiencing the effects of X-rays is important just because the X-rays are really out there, and we want in general to be aware of what's out there independently of what use we might be able to make of it (e.g., in medicine, or airport security). Theory often points the way to such discoveries (the *accidental* discovery of X-rays being one important exception, of course, among others). So theorizing aids us in making predictions and in discovering new kinds of things in the world.

Second, theorizing usually clarifies some of our concepts. Generally, a theory will suggest a more precise intensional definition of a concept, thereby clarifying the limits of its extension. The theoretical claim, based on experiments with some samples of water, that all and only water is made up of H<sub>2</sub>O suggests that the characteristic of being H<sub>2</sub>O be incorporated into the definition of water; theory sediments into definition. This refined definition (embodying practical techniques of detection and measurement) helps us delimit the extension of the term 'water'; it allows us to distinguish cases of non-water that we might have been unable to distinguish before.

Theorizing in non-empirical matters seems to perform only the second of the two functions above. Theories of justice, to take just one example, offer refinements of a common-sense understanding of justice, refinements that clarify the boundary between

the just and the unjust. Theories of justice need not, however, generate predictions about when and where one will find just or unjust acts. They do not lead to discoveries either. A theory of justice will not tell us, for example, about some chemical released in the brain of anyone performing unjust acts. They propose criteria for *identifying* acts or institutions as being either just or unjust.

Non-empirical theories are useful insofar as they clarify concepts. The function of pre-theoretical intuitions in this context is to serve as a set of data over which to theorize. When constructing a theory of justice, for example, one begins with one's most secure intuitions regarding cases of justice and cases of injustice. Philosophers sometimes rely on artificial and simplified hypothetical examples in order to draw on the most secure intuitions possible, intuitions more secure than we might ever find in realistic situations.<sup>8</sup> The philosopher then begins hypothesizing, looking for general criteria that all the sure cases of justice fill and none of the sure cases of injustice (or anything else) fills. This is the process of hypothesis-formation similar to that in the natural sciences. We need not pause here to consider the details of how hypotheses are formed. For my present purposes, I want only to notice what one does after one has a hypothesis, or theory, in hand.

Once one has a theory of justice that one has found to be satisfactory with respect to one's secure judgments regarding cases of justice and injustice, one will typically use it to decide cases that were not easily decidable before. In fact, it seems that this is the

---

<sup>8</sup> See Rawls 1971, 11-13 for his account of an idealized "original position." He explains that we test our ethical intuitions on such examples (p. 19). Hare 1963, 90-93 is another example of the use of artificial examples to develop secure intuitions.

whole point of theorizing about justice. Plato would say, of course, that we want to find out what *justice itself* is, the essence of justice. So Plato would not say that the *whole* point of theorizing about justice is to have a tool for ruling on unclear cases. Even Plato, though, would have to admit that *a* benefit of knowing justice itself is greater ease in recognizing instances of it in the realm of Becoming. We can throw off the Platonic metaphysics and any sort of essentialism (embracing, say, some form of conventionalism) and *still* claim that the point of a theory of justice is to aid us in deciding unclear cases in a way (*one* way among many, of course) that will be *consistent* with our most secure intuitions. Theorizing about justice is a way of extending the principles that seem to be at work in our firmest intuitions to unclear cases.

The situation is more complicated than I have just described, however. There is not always a clear boundary between particular cases we feel sure about and those we feel unsure about. The cases we feel sure about are supposed to serve as data, but we are rarely in a situation where we can be confident we have *all* the data. We can theorize over a certain body of data and head out to decide all other particular cases with that theory. Often, however, we will come across some particular cases that somehow seem as though they should be part of the data, although they were not originally included. In such situations, we face the question of whether the particular case is to be ruled by the theory or whether the particular case falsifies the theory—*modus ponens* or *modus tollens*. Through deciding in a number of such situations which way to go, one approaches what Goodman and Rawls have called reflective equilibrium between theory and particular judgments (see Goodman 1955, 65-68 and Rawls 1971, §§ 4 and 87).

We can safely sidestep such complications for now. It is enough to say that there are clear cases of justice and injustice (even if we have to resort to artificially simplified situations) and there are cases that we find very difficult to decide. One function of a theory of justice is to give us a way of deciding the very difficult cases in a way that is consistent with our other judgments. We value consistency; consequently we value having a set of *principles* on which to base our decisions.

Let's now apply this digression on the functions of theories to *theorizing about explanation* in particular. Can we expect a theory of explanation to yield predictions or facilitate discoveries? It seems not. A theory claiming that every explanation has characteristic *C* will support the prediction that the next explanation we encounter will have the characteristic *C*, but this is not the kind of prediction I am thinking of. When I say that we cannot expect a theory of explanation to support predictions, I mean that it will not support non-analytic predictions such as a prediction of when we will be able to explain phenomenon *X*. Nor will it lead the way to discoveries such as how the brain processes information that is explained versus information that is not explained.

Will a theory of explanation clarify our concept of explanation in a way that fixes the extension of the corresponding term? One would expect this, since a theory of explanation is, like a theory of justice, a non-empirical theory. One might think that a theorist of explanation will go out into the world, theory of explanation in hand, and enlighten the world about what are the *real* explanations and what are not *really* explanations. Logical Positivists such as Carnap and Hempel apparently would use a theory of scientific explanation to legislate against, for example, explanation by

entelechies. Michael Friedman also *allows* for such activity: “*Most*, if not all, scientific theories that we all consider to be explanatory should come out as such according to our theory.... Although it is unreasonable to demand that a philosophical account of explanation should show that *every* theory that has ever been thought to be explanatory really is explanatory, it must at least square with most of the important, central cases” (1974, 13, emphases added).

I believe, however, that this is not the primary benefit of theorizing about explanation. Particular examples of explanation are all as clear as one could want. Although intuitions can differ from person to person, an individual’s intuitions about explanation are not at all likely to change due to consideration of a theory of explanation. It is not difficult to imagine changing one’s mind regarding the justice of a particular act in light of a favored theory of justice. What this amounts to is bringing a difficult case in line with clearer cases, under the governance of general principles. I believe that a similar situation does not arise (at least to any considerable degree) with respect to explanation, however.

This last claim should sound surprising. After all, aren’t there many conflicting intuitions about explanation, and don’t people sometimes change their intuitions about explanation? If so, isn’t it natural to suppose that people change their minds about the explanatoriness or explanatory satisfactoriness of a statement or law on the basis of theoretical considerations about the nature of explanation? Consider, however, the character of a change in view regarding explanation. Teleological explanations, for example, have been wholeheartedly accepted by many and thoroughly rejected by many.

Some people have undergone the conversion from the former to the latter position. This transition resembles a *shift in paradigm* much more than a revision in a single judgment, however. The difference is that changes in attitude toward explanation occur in shifts of large bodies of judgments rather than piecemeal. So acceptance of a new theory of explanation can lead one to change one's particular judgments regarding individual proposed explanations, but only, it seems, *en masse*. A theory of justice, by contrast, can bring a few isolated, errant judgments into line.

The methodology of all the leading philosophers of explanation since Hempel provides confirmation of my contention. The way they use individual examples of explanations (scientific, common-sense, psychological, etc.) presupposes that the reader will respond with a whole-hearted "Yes, that is an explanation!" They do not explore realms of contended explanations. They try to bring everyone on board the bandwagon with a number of secure examples of explanation and proceed to look for defining criteria. After a theory of explanation is produced, they rest content as long as it has not failed to account for any of the canonized examples of (proper) explanation. That is, they strive only for immunity to counterexamples. I know of no case where a philosopher of explanation was confronted with a counterexample and refused to acknowledge the force of the counterexample.<sup>9</sup> It is, indeed, admirable for a philosopher

---

<sup>9</sup> There are many *apparent* exceptions to this claim. They follow the model of Hempel's rejection of Michael Scriven's objections to his model (Hempel 1965, 413). Hempel's rejection of the alleged counterexamples was, however, based on the claim that the counterexamples are not examples of *scientific* explanation, and Hempel intended to theorize about scientific explanations only. Hempel would say these are explanations, but not of the sort he was trying to characterize. What I am claiming here is that Hempel and others do not focus their discussion of alleged counterexamples of the

to acknowledge forthrightly the force, if any, that counter-arguments actually have, and this aspect of the discussion testifies to the high philosophical caliber of the participants. But it is just these “rules of the game” that tell us that the end product, the theory of explanation, will not serve primarily to set us straight about what the *real* explanations are and what the merely apparent explanations are. The discussion depends on our knowing all along an explanation when we see one.

Theorists of explanation have taken their task to be that of articulating either some necessary conditions for an explanation, a set of sufficient conditions for an explanation, or, ideally, a set of individually necessary and jointly sufficient conditions. I have questioned whether this is really what we should look for in a *philosophy* of explanation. I raise this question because I am impressed with how *little* one learns about the nature of explanation itself when one reads the literature on theories of explanation. One can get through the literature and look at it critically only if one is already secure in one’s sense of when one has hold of an explanation and when one hasn’t. The exceptional recalcitrance of individual judgments about what sort of account can be explanatory will be obvious in later chapters when we examine competing views on the most recent unresolved questions about explanation. For now, I want only to defend my strategy of *not* entering the ring with those theorists who engage in the combat of theorizing and counter-examplng. There are benefits in what they do, but I am hunting different prey. I am after a deeper *understanding* of the judgments about explanatoriness that theorists

---

appropriate type on disputes over the very *explanatoriness* of the alleged counterexamples. If the counterexample comes from the appropriate field, it is generally given substantial weight.



of explanation take as their starting points, the data over which they theorize. Hence, I now turn to a positive account of what I think a philosophy of explanation should deliver, and later I develop my own philosophy of explanation.

## CHAPTER 3

### METAPHYSICAL EXPLANATION

#### What We Should Expect from the Philosophy of Explanation

We can now explore what the payoff from philosophizing about explanation might be. The most interesting and significant feature of concepts of explanation is what people *do* with them in the way of inference. It is in inferential moves from explanation to *truth* that our discriminations between explanation and non-explanation, and between better and worse explanations, become philosophically significant. Discriminations in the area of explanation become important for philosophy when they are responsible for discriminations in the area of truth.

In this chapter, I will show how various quests for explanation can lead to metaphysical theorizing. This is the first of the two central themes in this dissertation. On the one hand, I have said that someone who has criteria for what counts as a satisfactory explanation and for discriminating better from worse explanations (even if the criteria are far from precise and explicitly acknowledged) will use these criteria in making decisions about what exists. My thesis is that metaphysical theories, like all theories about what exists, are formulated primarily as responses to felt explanatory needs. They are accepted in large part, at least, for the sake of their explanatory power. This chapter presents historically important examples of this aspect of the connection

between explanation and existence claims. On the other hand, one's beliefs about what exists (especially at the most fundamental and general levels of reality) will strongly influence what criteria one accepts for identifying satisfactory explanations and distinguishing better from worse explanations. This aspect of the connection between explanation and existence claims will be the central topic of later chapters.

The Logical Positivists, most notably Hempel, who started up the contemporary discussions of scientific explanation were well aware that it was how explanation was used to justify inferences to metaphysical conclusions that made it philosophically controversial. The early Positivists had been the ones, after all, to place a strict ban on explanation in science. In *The Grammar of Science* in 1892, Karl Pearson expressed a view that came to dominate the thinking of many early Logical Positivists: "Nobody believes that science explains anything; we all look upon it as a shorthand description, as an economy of thought" (1957, xi; see also 136 and 140). Comte, Mach, Poincaré and others had expressed similar sentiments. The Logical Positivists had been led to this view in particular, as they were in general, through strong objections to the speculative metaphysics dominating philosophy in the late nineteenth century. David Miller claims, "For the most part, the rebellion against explanation has been of this general sort; it is a rebellion against the belief that there is a metaphysical (beyond sensation) condition for change, whether that condition be a purpose in the mind of God or a more natural one such as the 'inner nature' of matter" (Miller 1946, 242). Rudolf Carnap gives a characteristically clear and direct description of the situation:

In the nineteenth century, certain German physicists, such as Gustav Kirchoff and Ernst Mach, said that science should not ask "Why?" but "How?" They meant

that science should not look for unknown metaphysical agents that are responsible for certain events, but should only describe such events in terms of laws. This prohibition against asking “Why?” must be understood in its historical setting. The background was the German philosophical atmosphere of the time, which was dominated by idealism in the tradition of Fichte, Schelling, and Hegel. These men felt that a description of how the world behaved was not enough. They wanted a fuller understanding, which they believed could be obtained only by finding metaphysical causes that were behind the phenomena and not accessible to scientific method. Physicists ... objected to why-questions because they were usually metaphysical questions.... When I was young and part of the Vienna Circle, some of my early publications were written as a reaction to the philosophical climate of German idealism. As a consequence, these publications and those by others in the Vienna Circle were filled with prohibitory statements similar to the one I have just discussed.... Today, especially in the United States, we seldom make such prohibitions. The kind of opponents we have here are of a different nature, and the nature of one’s opponents often determines the way in which one’s views are expressed (1966, 12).

Carl Hempel worked against this background of hostility to explanation, and his goal was to explicate a scientific and non-metaphysical meaning for ‘explanation’, a meaning that fit with the idea that the primary activity of science is describing events “in terms of laws.” He intended to give an empiricist account of scientific explanation.<sup>1</sup> The flood of literature to follow has shown that he did not meet with unqualified success, regardless of what degree of success he did have. Interestingly, the attempts to eliminate difficulties in the early works of Hempel, Popper, Nagel, Braithwaite, and others, have led right back to some of the key areas of contention between empiricists and more ontologically liberal metaphysicians. The idea that we can have scientific explanations

---

<sup>1</sup> For an examination of the question whether science ought to go beyond mere description to provide explanations, see David L. Miller’s “The Meaning of Explanation” (1946) and “Explanation Versus Description” (1947). These articles take into account the logical positivists’ hostility to explanation as a part of metaphysics. Miller’s strategy is not so much to isolate explanation from metaphysics (as Hempel’s was) but to give an account inoffensive to the contemporary empiricist of the “metaphysics” that is essentially involved in scientific explanation.

without reference to anything beyond experience (and descriptive laws) has fallen on hard times. We will examine the reasons for this in chapters 5 and 6.

The Logical Positivists objected to metaphysical extravagances of a Platonic, Spinozistic, or Hegelian nature. They objected to inferences to Platonic Forms, a Spinozistic *Deus sive Natura* (God, or Nature), or the Hegelian Mind. One might charge that these inferences have no empirical ground. A metaphysically-minded philosopher *can* reply, however, that there is an empirical ground for such metaphysical realities, because such metaphysical realities are inferred in order to explain the empirical content of experience. F. C. Copleston, in a debate with A. J. Ayer, expresses this view of the relationship between experience and metaphysics: “I do not mean to imply that no sense-experience is in any way relevant to establishing or discovering the existence of a metaphysical reality. I certainly do believe that metaphysics must be based on experiences of some sort” (Ayer and Copleston 1949, 162). Plato seems to have meant something similar about the role of experience in philosophical reasoning when he says, starting at 523b of the *Republic*, that

Some reports of our perceptions do not provoke thought to reconsideration because the judgment of them by sensation seems adequate, while others always invite the intellect to reflection because the sensation yields nothing that can be trusted.... And for the clarification of this [sensation that calls for reflection], the intelligence is compelled to contemplate [for example] the great and small, not thus confounded but as distinct entities, in the opposite way from sensation.... [A]nd it would compel the soul to be at a loss and to inquire, by arousing thought in itself, and to ask, whatever then is the one as such, and thus the study of unity will be one of the studies that guide and convert the soul to the contemplation of true being ... facilitating the conversion of the soul itself from the world of generation to essence and truth (523b-525c, Plato 1961, 755-58, trans. Paul Shorey).

Plato appears to be saying that philosophy begins with experience, or perception, and proceeds through intellectual reflection on that experience to turn the soul from contemplation of the objects of sense to the realm of Being, the Forms, to “essence and truth.” Beginning with experience, we arrive at metaphysical truth. Aristotle also seems to adopt a similar method, according to Terence Irwin: “Aristotle’s inquiries, described in very general terms, should be ways to reach what is known by nature [i.e., first principles] from what is known to us [i.e., experience]. He claims, equally generally, that an inquiry should begin from ‘appearances’, *phainomena*” (Irwin 1988, 29). One can find similar comments about the method of intellectual reflection on experience as a method for metaphysics in other philosophers. Metaphysicians might rely on considerations of adequate explanation to bridge the gap between (illusory?) sense experience and metaphysical reality, to make the inference from appearance to reality.

One might reason that a certain metaphysical story is just the *best* (but not the only) explanation of various features we experience in the world. Aristotle apparently took Plato’s metaphysics and the natural philosophy of the Presocratics as serious contenders for truth, but argued that his own metaphysics of substance was a superior explanation of a vast array of features of the world and of many aspects of our knowledge of the world (see especially his criticisms of his predecessors in *Metaphysics* I, viii-ix). On the other hand, one can argue that a certain metaphysical doctrine is the *only* possible explanation of (some aspect of) experience, and this method is perhaps more typical of traditional metaphysics. This is one way of interpreting the method of the traditional *deductive* metaphysics found in the Rationalists, especially Spinoza and

Leibniz. In general, a metaphysician of this type argues that there is only one logically consistent way of accounting for the world of Appearance. Transcendental argumentation also falls under the heading of metaphysical argumentation that reduces the explanatory options to one. One argues to a metaphysical reality (*Ding an sich*) that lies beyond the realm of any possible experience, as a *necessary* condition for the possibility of any experience.<sup>2</sup> This is to infer to the truth of an explanation of how-possibly X. In either case, whether inferring to the best explanation or to the only (logically consistent) explanation, a metaphysician using such arguments makes essential use of concepts and criteria of explanation to justify the desired claim about objective reality.

My attempt to characterize metaphysical argumentation as involving an inference to an explanation goes against standard interpretations of the great metaphysicians of the past. I do not intend to claim that they took themselves to be constructing such arguments. In fact, they all probably took themselves to be constructing arguments of different sorts (typically involving a special role for definitions), because the form of

---

<sup>2</sup> More precisely, transcendental arguments “prove that certain *concepts* are necessary for thought or experience,” according to Barry Stroud (1968, 118, emphasis added). Further, the transcendental argument *justifies* those concepts as being “objectively valid”: “For Kant a transcendental argument is supposed to answer the question of ‘justification’, and in so doing it demonstrates the ‘objective validity’ of certain concepts. I have taken this to mean that the concept ‘X’ has objective validity only if there are X’s, and so demonstrating the objective validity of the concept is tantamount to demonstrating that X’s actually exist. Kant thought that he could argue from the necessary conditions of thought and experience...to the actual existence of the external world of material objects...” (Stroud 1968, 129). Notice the importance of arguing to necessary conditions, for one could conceivably argue to sufficient but unnecessary conditions, but then one would have at best an inductive argument to the best explanation of how thought and experience are possible, not a deductive argument to the only explanation.

argument we call “inference to an explanation” was probably not explicitly articulated before the work of C. S. Peirce. I believe, however, that all metaphysical reasoning depends on premises, either explicit or implicit, about criteria for adequate explanation<sup>3</sup> and that the early Logical Positivists rejected explanation as being somehow distinctly metaphysical because they saw, perhaps dimly, this dependence. It is clearly fallacious to infer that all explanation is metaphysical because all metaphysics depends on a demand for explanation of a certain sort, but it is fairly easy to see how one might choose to adopt the strategy of rejecting all attempts at explanation in order to prevent excursions into metaphysics. Carnap explains clearly how his early prohibitions were reactions to the Idealist philosophical climate of the time.

I want now to examine how assumptions about adequate explanation could serve to buttress some traditional metaphysical arguments. I will do so by trying to reconstruct some exemplary traditional metaphysical arguments as embodying substantive premises about the criteria for adequate explanation. Apologizing in advance to historical scholars sensitive about historical accuracy, I will be plundering the history of Western philosophy looking for clues as to what kinds of implicit assumptions about explanation might lead one to draw various sorts of metaphysical conclusions.

---

<sup>3</sup> I should say, the only *plausible* arguments for metaphysical or ontological conclusions are inferences to an explanation. I am assuming that other types of arguments (*a priori* arguments, say) have been effectively ruled out by recent critics who have argued that existential premises cannot be known *a priori* and that non-existential premises cannot lead deductively to existential conclusions.



Plato on Adequate Explanation

It is unclear what form of inference Plato thought would lead one from experience of particulars in the realm of Becoming to contemplation of a Form “in itself” in the realm of Being. He typically addresses questions regarding the proper form of education or training that will bring about such contemplation in an individual, and he typically emphasizes a direct apprehension, or intellectual intuition, of the Forms rather than an *inference* to the conclusion that the Forms exist (see, e.g., *Republic*, 504d-518b). There is room, however, for the interpretation that considerations of the demands of adequate explanation played some part in leading him to postulate the Forms.

We find some support for this interpretation in Aristotle. Aristotle’s comments on Plato must be treated with care, for Aristotle mentions Plato’s views usually to criticize them. On the other hand, Aristotle knew Plato personally, conversed with him in a shared native language, had a great many philosophical discussions with him, and undoubtedly had access to many more expressions of Plato’s thoughts than we do. It is reasonable to expect that Aristotle knew Plato’s teachings as thoroughly as anyone else could. At *Metaphysics* 990b1-5, Aristotle writes, “But as for those who *posit* the Ideas as causes, firstly, in *seeking to grasp the causes of the things around us*, they introduced others equal in number to these, as if a man who wanted to count things thought he would not be able to do it while they were few, but tried to count them when he had added to their number. For the Forms are practically equal to—or not fewer than—the things, in *trying to explain which these thinkers proceeded from them to the Forms*” (translated by W. D. Ross in Aristotle 1941, 706, emphases added). Richard Hope’s

translation of the *Metaphysics* is more conducive than any other to an interpretation placing explanatory considerations at the root of metaphysical reasoning, for Hope translates the key term *αἰτία* as “explanation” or “basic explanatory factor.” *Αἰτία* is usually translated as “cause,” as in Aristotle’s doctrine of the four causes. In Hope’s translation, the passage above reads: “To begin with, *in their attempt to explain things visible*, [those who believe that ideas explain all] *invented* an equal number of other things, as if they thought it was easier to count many than few. For their ideas are precisely equal in number to or not fewer than *the objects which they are supposed to explain by referring each object to its idea*” (Aristotle 1952, 27, emphases added).<sup>4</sup> Either translation makes it clear, however, given the phrases I have emphasized, that Aristotle, at least, took Plato’s main task to be the explanation of the objects of experience in metaphysical terms. Plato and his followers meant to “explain things visible,” to “grasp the causes of the things around us,” and in order to do this they “posited,” “introduced,” or “invented” the Ideas, or Forms.

It might be, however, that Aristotle presents Plato’s philosophy in this way simply because it was *Aristotle’s* main concern to explain visible things. What Plato has to say himself about the demands of adequate explanation is found primarily in the *Phaedo* 95e-102b<sup>5</sup> (followed by some examples of Formal explanations, most notably of the

---

<sup>4</sup> See also 987b19-20, Aristotle 1952, 19: “Since ideas serve as principles of explanation for other beings, he [Plato] believed their elements to be the elements of all things.”

<sup>5</sup> For an illuminating account of the *Phaedo* on explanation, see Ruben 1990, ch. 2; for accounts of other aspects of Plato’s views on explanation, see Ashbaugh 1988 and Brown 1988.

possibility of change, and finally, at about line 107, application of Formal explanation to the central problem of the dialogue, the question of the immortality of the soul). There Plato sets out his doctrine of causation, stating that “whatever else is beautiful apart from absolute beauty is beautiful because it partakes of that absolute beauty, and for no other reason” (100c, Plato 1961, 81). He describes a kind of causation such that the Forms are the causes of things’ being instances of the Forms. This, of course, leads to the view that an adequate explanation must mention the Forms in which a thing participates. Again taking commentary from Aristotle, we read at *Metaphysics* 991b3, “In the *Phaedo* this doctrine is stated by saying that the ideas [Forms] are explanations for being and becoming” (Aristotle 1952, 29. Cf. *On Generation and Corruption* 335b8-25, Aristotle 1941, 524). Although it is unclear how the doctrine of the Forms, the principle of causality, and the doctrine of explanation are all related (specifically, which is most basic in Plato’s thinking) there seems to be some support in Plato’s middle writings for the thesis that he was led to his doctrine of the Forms as the result of accepting a particular doctrine of explanation (and, just as importantly, rejecting others, such as the materialistic-deterministic explanations of some Presocratics). It would then follow that Plato was motivated at least in part in his own thinking by an argument to the best explanation, given his specific criteria for adequate explanation.

Plato provides arguments in favor of his doctrine of explanation, arguments based primarily on counterexamples to various presocratic methods of explanation and illustrative examples of his preferred Formal explanation. In this presentation, he is assuming without argument the reality of the Forms and trying to show that explanation

in terms of the Forms is superior to other explanations, but we can fruitfully ask whether Plato actually came to believe in the existence of the Forms *because* they fulfill perfectly some prior criteria he held regarding adequate explanation.

Plato begins at *Phaedo* line 96 by having Socrates survey some mechanistic-deterministic explanatory strategies of the presocratic natural philosophers. Socrates tells us that he was, when he was young, interested in questions such as whether it is fermentation produced from heat and cold that generates living creatures; whether it is with the blood, or with the air or fire in us, or perhaps with the brain that we think, hear, see, smell, remember, opine, and know (96b); or whether it is due to eating and drinking that a human being grows (96d). He then tells us that he came to believe that such mechanistic accounts as those he knew of were not explanations: “I was so befogged by these speculations that I unlearned even what I had thought I knew...” (96c). “I am [now] very far from supposing that I know the explanation of any of these things” (96e). “Nor can I now persuade myself that I understand how it is that things become one, nor, in short, why anything else comes or ceases or continues to be, according to this method of inquiry” (97b). This latter quote shows specifically how concerned Plato was with the central philosophical task of accounting for how things come to be and cease to be.

Plato has Socrates go on to describe how he first heard of the teachings of Anaxagoras, that mind is the cause and explanation of everything, and was very pleased by this method of explanation (97c).

Somehow it seemed right that mind should be the cause of everything, and I reflected that if this is so, mind in producing order sets everything in order and

arranges each individual thing in the way that is best for it. Therefore if anyone wished to discover the reason why any given thing came or ceased or continued to be, he must find out how it was best for that thing to be, or to act or be acted upon in any other way.... These reflections made me suppose, to my delight, that in Anaxagoras I had found an authority on causation who was after my own heart. I assumed that he would begin by informing us whether the earth is flat or round, and would then proceed to explain in detail the reason and logical necessity for this by stating how and why it was better that it should be so.... [I]f he made this clear, I was prepared to give up hankering after any other kind of cause (97c-98a).

This is what Socrates at first expected from Anaxagoras, and it reveals much about Socrates' (and, I assume, Plato's) demands (we could even say, *a priori* demands) of explanation. Socrates favors explanations in terms of what is *best*, or the highest *good*; that is, he demands of an adequate explanation some mention of a goal, purpose or end (a good presumably aimed at) for what happens. This comes out again in a subsequent passage where he produces further criticism of materialistic explanation.

Socrates was so excited about this explanation of all things in terms of mind that he says, "I lost no time in procuring the books [by Anaxagoras], and began to read them as quickly as I possibly could, so that I might know as soon as possible about the best and the less good" (98b). Well, he tells us, he was quickly disappointed, for he found Anaxagoras using the same old material explanations he had already rejected. "I discovered that the fellow made no use of mind and assigned to it no causality for the order of the world, but adduced causes like air and aether and water and *many other absurdities*" (98b-c, emphasis added).

He offers an example of something that calls for an explanation, in order to show that the only acceptable explanation is in terms of what is best, and that explanations in terms of material causes are "too absurd." Why is Socrates sitting in prison conversing

with his friends? The only proper explanation is that Socrates thinks it best to submit to the will of the people of Athens who have condemned him, and therefore to stay and accept his punishment. This explanation fits very well Richard Swinburne's description of *personal explanation* (1979, 32-35). Roughly, a personal explanation cites the intentional action of a rational agent. Socrates is sitting in prison conversing with his friends because he, a rational agent, intends to do so (thinking it best to do so), and acts on that intention. Plato seems to think that the only candidate mechanistic explanation for this situation would cite merely Socrates' physical state at the time, the state of his "bones and sinews ... his flesh and skin." Such an account explains nothing, he says.

At 99d Plato begins to present his account of Formal explanation in full. He has already said some important things about what is required in an explanation, but he has yet to describe how the Forms fill the bill perfectly. He has told us (1) that explanation must be in terms of what is best, or the highest good, and (2) that material conditions, although they might provide a *sine qua non* for an event, cannot by themselves explain it (99b); for material conditions answer only "how?", we could say, not "why?" Now Plato goes on to tell us, through Socrates, the way in which the Forms provide genuine explanation, giving us the real reasons, satisfying the conditions he set out when he considered the theory that mind is the genuine cause of all things. His account of Formal explanation is actually brief and sketchy. He writes, "It seems to me that whatever else is beautiful apart from absolute beauty is beautiful because it partakes of that absolute beauty, and for no other reason" (100c). He also writes, "I cling simply and straightforwardly and no doubt foolishly to the explanations that the one thing that makes

that object beautiful is the presence in it or association with it, in whatever way the relation comes about, of absolute beauty” (100d). This is really the full extent of what Plato has to say in this dialogue directly about Formal explanation. Socrates leaves it to his interlocutors, and Plato leaves it to his readers, to make the connection between the conditions he set out for explanation and the nature of Formal explanation. It is only Formal explanation that explains things in terms of what is best, for the Form in which a thing participates is the highest good for that thing in that respect.

Is Plato’s belief in the reality of the Forms *based on* antecedently accepted conditions of adequate explanation, and if so, how? One passage in particular threatens the interpretation that Plato somehow inferred the existence of the Forms based on a theory of explanation and causality. At 100b of the *Phaedo*, he writes, “As I am going to try to explain to you the theory of causation which I have worked out myself, I propose to make a fresh start from those principles of mine which you know so well—that is, I am assuming the existence of absolute beauty and goodness and magnitude and all the rest of them.” This makes it sound like Plato is basing his theory of explanation on his doctrine of the Forms; hence, he would be inferring criteria for adequate explanation from a prior commitment to the metaphysical reality of the Forms. I believe this is an incorrect reading, however. The passage quoted begins, “As I am going to try to explain to you the theory of causation which I have worked out myself....” Socrates’ purpose here is to present his theory of causation to his interlocutors (and Plato’s purpose is to present it to his readers). It is only for pedagogical reasons, I believe, that Socrates invokes the doctrine of the Forms, relying

on general agreement in, or at least familiarity with, this doctrine to make his theory of explanation understood. Some support for this reading comes from a later comment by Phaedo. Phaedo breaks his narrative of the dialogue to exchange some comments with Echebrates, to whom he is relating the whole dialogue. He tells Echebrates: "I think that when Socrates had got this accepted, and it was agreed that the various forms exist, and that the reason why other things are called after the forms is that they participate in the forms, he next went on to ask..." (102b). This statement seems to suggest that Socrates intended to give at least some *support* to the doctrine of the Forms, for Phaedo says that Socrates got it accepted, and got all to agree "that the various forms exist."

Stronger evidence for my claim comes from the order of presentation. Plato first presents Socrates' admiration for the theory that mind is the cause of all things, or the ultimate explanatory factor. In doing so, he sets out his fundamental criterion for an acceptable explanation. It is only after this that he presents his doctrine of the Forms, demonstrating how it affords explanations fitting his criterion and how other explanations fail. Plato's presenting first the criterion and then the metaphysical doctrine allowing for explanations that fit the criterion makes me think that Plato actually inferred the existence of the Forms on the basis of prior commitments regarding adequate explanation, i.e., that the pattern of presentation in the dialogue is supposed to reflect the train of thought that led Plato to believe in the Forms. Going back to Aristotle's comments again, I believe we find only confirmation of this hypothesis: "In their attempt to explain things visible, they invented an equal number of other things [the Forms]."



Finally, Plato's comments, cited earlier, on the place of sense experience, or perception, in philosophical method support my claim that Plato *arrived at* the doctrine of the Forms as a result of his attempt to explain various problematic aspects of experience. This time I will use Cornford's translation (because it makes certain points clearer), in which we read, "Take our perceptions, then. I can point to some of these which do not provoke thought to reflect upon them.... But in other cases perception seems to yield no trustworthy result, and reflection is instantly demanded" (*Republic* 523b, Plato 1941, 238-9).

Now sight, too, as we said, perceives both big and little; only not as separate, but in a confused impression. In order to clear up this confusion, intelligence was driven to look at bigness and smallness in the opposite way, as distinct things. *It is some such experience as this that first prompts us to ask what is meant by bigness or smallness. And that is how we came to distinguish what we call the object of intelligence from the thing seen* (524c, Plato 1941, 240, emphasis added).

This passage provides powerful evidence, I believe, that it was the attempt to explain in a satisfactory way certain problems of interpretation presented in experience that led Plato to the theory of the Forms. It appears, further, that Plato thinks such patterns of reflection should lead others to accept his theory; that is, he recommends it explicitly on the basis of its explanatory power. Socrates' comments in the *Phaedo* about the superiority of explanation in terms of what is best account for Plato's choosing the theory of the Forms as the best (or only) explanation of these problems given in experience.

### Aristotle on Adequate Explanation

It seems fairly clear that Aristotle would endorse what we would call an inference to an explanation. It also seems clear what Aristotle regards as the conditions for adequate explanation, for he spells them out explicitly in his doctrine of the four causes.<sup>6</sup> An adequate explanation must be a causal explanation in one of his four senses (for Aristotle it is the causes themselves, non-linguistic and non-mental elements of reality, that explain), and purported explanations in terms of something other than those four types of cause (in terms of *chance*, for example) are not genuinely explanatory. This much is clearly true for Aristotle's account of theoretical scientific knowledge. Scientific knowledge, he says, is knowledge of causes and of essences. Such a methodology supports a scientific ontology, for one can justifiably infer the existence of the causal agents invoked to explain observed objects. It supports an inference, for example, from a sculpture to a sculptor (as the efficient cause) and to an end (as a final cause). Most importantly, it supports inferences to essences as ultimate explanatory principles (formal *αἰτίαι*), and this is how one gains theoretical scientific knowledge of kinds of things. Does Aristotle apply a similar methodology to the science of first principles, however? Is metaphysics, in this sense, an extension of physics? Or is there some significantly different method, or a significantly different explanatory demand, appropriate for metaphysics?

---

<sup>6</sup> See Moravcsik 1974 for a defense of the thesis that what is commonly called Aristotle's doctrine of four causes is actually Aristotle's account of what constitutes an adequate explanation.

In this section, I will argue that the science of first principles, or metaphysics, is for Aristotle primarily an explanatory activity. The main evidence for this comes from *Metaphysics A*, I-II, where Aristotle describes the development of theoretical knowledge and the nature and goal of wisdom (i.e., philosophy). Just as we can infer to explanatory causes that are natural objects in natural philosophy, we can likewise infer to explanatory causes in first philosophy that are supernatural (divine), or metaphysical.

To begin, we get some initial support for the claim that Aristotle took metaphysics to be based primarily on demands for an explanation of problems presented in experience by noting that he takes the method appropriate for metaphysics to be inductive (i.e., non-syllogistic). This rules out, for example, a conception of metaphysics (and, consequently, of philosophy as a whole) as being based on self-evident, certain, or necessary starting-points and proceeding through pure deduction. For Aristotle, the starting-points are clearly given in experience and the method is primarily inductive, as Terence Irwin explains:

Induction (*epagôgê*) is the method that Aristotle describes for the approach to first principles.... Induction presupposes that experience has sorted out the true appearances from the false.... Inductive generalization must therefore presuppose some explanatory hypothesis about kinds. Aristotle's suggestion that induction is primarily generalization would be more reasonable if he could justifiably assume that inquiry and experience have already produced perceptual judgments involving the right kinds, and so incorporate the right explanatory hypothesis. But his description of inquiry and experience justifies no such assumption.... Aristotle's remarks about induction, and about discovery in general, leave out many of the most important and most difficult questions (Irwin 1988, 32-3).

Aristotle's explicit recognition that argument to objective first principles is inductive (non-syllogistic) facilitates the interpretation of his metaphysical arguments as being based on prior commitments regarding the demands of adequate explanation.

Aristotle acknowledges that *demonstration* is not the appropriate method for arriving at first principles. He apparently relies on something like the dialectical method, but Terence Irwin (1988) argues extensively that this method is not adequate for the task of establishing *objective* first principles. Irwin claims to find a modified method, “strong dialectic”, in Aristotle’s writings (especially *Metaphysics*, Book IV) adequate to the task. Setting that issue aside, I want to assert that Aristotle’s inductive reasoning will rely in some way on premises about adequate explanation. The induction he uses is not merely generalization, as Irwin points out in the quote above. That Aristotle’s inferences are guided by explanatory considerations is also evident in the fact that, for Aristotle, “a correct theory must not only yield, explain, and support the true appearances; it must also explain why the false appearances might have seemed true even though they are false” (Irwin 1988, 47). Thus, Aristotle’s goal is to explain the appearances, or experience, *in general*.

In *Metaphysics* A, I, Aristotle distinguishes experience from art (*τεχνέ, techné*), reasoning, or scientific knowledge (*επιστεμέ, epistemé*). “Experience, though it seems quite like scientific knowledge and art, is really what produces them.... Art is born when out of the many bits of information derived from experience there emerges a grasp of those similarities in view of which they are a unified whole” (981a2-7, 1952, 3-4). Ross’ translation reads, “[S]cience and art come to men *through* experience....” (Aristotle 1941, 689). Experience (“connected experience” in Ross) is not bare sensation or perception for Aristotle, but requires some higher cognitive functioning. So we have more experience than animals, although the more developed animals have some

experience. People can, however, have experience without having what Aristotle calls art or knowledge or understanding. Aristotle's comments make it clear that men having experience can have a kind of practical knowledge allowing them to excel "in a situation in which something is to be done" (981a12, 1952, 4). Without art or scientific knowledge, however, they lack *understanding* of why things are better done one way rather than another.

Nevertheless, we believe that knowing and understanding characterize art rather than experience. And so we take experts in an art to be wiser than men of mere experience; because wisdom presumably comes only with knowledge, and we believe that the experts can analyze and explain, whereas others cannot. Men of experience discern the fact "that", but not the reason "why"; whereas experts know the reason why and explanation [αἰτία]. (981a25-30, 1952, 4)

Aristotle says here that it is characteristic of those having wisdom (σοφία) to possess an explanation of what others might know to be the case but cannot explain. The wise know the reason (λόγος) for what is the case. Wisdom requires possession of explanations, and what is to be explained is presented in experience.

What we intend to say in the present account is that all men take what is called "wisdom" to be concerned with the first principles of explanation [αἰτία]: so that, to sum up what we have said, the man of experience is deemed wiser than those who merely sense things, the artist than men of experience, and the master worker than manual laborers. Thus, wisdom is thought to consist in theoretical rather than in productive kinds of knowledge. Clearly, then, wisdom [σοφία] is rational knowledge concerning basic factors [αἰτίαι] and principles. (981b28-982a4, Aristotle 1952, 5-6)

This is Aristotle's account of wisdom. Philosophy (love of wisdom, φιλοσοφία), then, is the pursuit of the sort of theoretical knowledge Aristotle has just described, knowledge of explanatory factors.

Knowledge of explanatory factors is the primary objective of *all* theoretical knowledge for Aristotle. In every case, one acquires knowledge of essences through the search for explanations of the appropriate subject matter. In natural philosophy, for example, Aristotle infers that the elements of bodies are primary matter and certain contraries (*On Generation and Corruption* II, 1) on the basis that this offers a better explanation of coming-to-be and passing-away than alternative theories (II, 3-7). He claims in *De Anima* that it is soul that explains the difference between living things and non-living things (413a20) and that it is essential to the soul to be in a body (414a20). And in metaphysics, “God is thought to be one of the reasons for all things and to be in some sense a beginning....” (*Metaphysics* 983a8, Aristotle 1952, 8).

Theoretical knowledge in metaphysics is of the same kind as theoretical knowledge in other areas of the pursuit of wisdom. “We must acquire knowledge of factors [*αἰτίαι*] that are primary.... These primary factors are fourfold” (*Metaphysics* 983a25, Aristotle 1952, 9). Aristotle goes on to describe the four *αἰτίαι*. In physics, metaphysics, the history of animals, and so on, one strives for explanations in terms of these *αἰτίαι*. Such explanations lead one to infer the existence of various things: the elements of things, the infinite (not as actual but as potential), place, the soul, God.

There is another kind of knowledge, however. There is another task for first philosophy: justifying the four *αἰτίαι*. Scientific knowledge, of things natural or eternal, is based on explanations in terms of the four *αἰτίαι*. But why should we think that there are just these four *αἰτίαι*? Aristotle’s most explicit response is to provide a number of examples: he shows (a) that a number of familiar phenomena can be explained in these

four ways (they do have these four causes), (b) that familiar attempts to give different explanations (causes), such as chance and spontaneity, are not genuinely explanatory (genuine causes), and (c) that no other candidates for explanatory causes are forthcoming (see *Metaphysics* I, *On Generation and Corruption* II. 9, and *Physics* II. 4-6).

There is another justification for the four αἰτίαι. We are justified in thinking that there are just these four αἰτίαι because they are part of the best explanation of change, or coming-to-be and passing-away. Aristotle believed in his four causes at least partly for this reason, as we see in his criticisms of others for failing on this count. “It is clear that those who reduce all to one ... err in many ways.... In their eagerness to explain generation and destruction and in explaining all things as processes, they ignore the very cause [αἰτία] of change” (*Metaphysics* 988b25-27, Aristotle 1952, 23). “But the third ‘originative source’ must be present as well [for coming-to-be and passing-away]—the cause vaguely dreamed of by all our predecessors, definitely stated by none of them” (*On Generation and Corruption* 335b8-9, Aristotle 1941, 524). Aristotle’s desire to explain motion, generation and destruction, and all other forms of change is well-known. How does his doctrine of the four αἰτίαι help explain change?

First, we must recognize that his four αἰτίαι are grounded in his conception of substance. Later we shall see that his metaphysical doctrine of substance is, for Aristotle, ideally suited to be the best explanation of change. Julius Moravcsik (1974) has argued that Aristotle’s doctrine of the four causes is actually a doctrine regarding the criteria for adequate explanations. After presenting his argument for this interpretation of the αἰτίαι, Moravcsik says that “we still have to find the reasons for [Aristotle’s]

selecting the well known four kinds. Why these four, why not others? And what ties the four together?" (1974, 5) He then claims that the answer lies in the relation of the four *αἰτίαι* to Aristotle's notion of substance.

It is plausible to suppose that though the various *aitiai* cover aspects of substances as well as the explanation of substances, the main classification of the four types arises out of reflexion on the nature of an Aristotelian substance. For the purposes of this analysis the following characterization of a (paradigmatic) Aristotelian substance should suffice: *a set of elements with a fixed structure that moves itself toward self-determined goals*. The key factors in this characterization are: element (constituent), structure, agency of motion, and goal. These correspond, roughly—as we shall see—to the well-known “four causes”.

Thus we see a correlation between the factors into which we analyze an Aristotelian substance; its ‘nature’, one might say, and the four main types of explanatory factors (1974, 5).

Moravcsik argues that Aristotle's account of the conditions of adequate explanation, the account of the four causes, makes sense because the four *αἰτίαι* “represent the main factors of what for Aristotle is the primary category of beings, namely substance” (1974, 6). In other words, Aristotle's account of explanation in terms of the four *αἰτίαι*, and thus his account of scientific knowledge and philosophical wisdom, is supported by an ontological claim, namely, his doctrine of substance.

Peter Railton (1989) has argued that all concepts of explanation are based on metaphysical commitments, and this would support the particular case where Aristotle's theory of explanation is based on his metaphysics of substance. Railton in fact makes the strong claim that “it is inconceivable that a notion such as explanation could fail to depend crucially upon one's most general picture of the world and its ways” (1989, 221). He does not, however, provide any argument to support such a strong claim, and at other points in his paper he trades that strong claim for a weaker thesis: “In these two



[examples] there is an evident coincidence between the requirements laid down by a model of explanation, on the one hand, and substantive metaphysical assumptions, on the other” (1989, 223). “To say that one’s background picture of the world is involved in one’s *conception* of explanation is to *suggest* [emphasis added] that one’s intuitions about particular kinds or instances of purported explanation may not constitute a body of neutral data for testing theories of explanation ...” (1989, 224). After having examined more historical case-studies, we will have to address directly the question whether there can be a concept of explanation that carries with it *no* metaphysical or ontological commitments, or whether every concept of explanation *must* contain metaphysical commitments. For now it suffices to point out that there does not appear to be any compelling rationale for Aristotle’s doctrine of the four *αἰτίαι* other than the claim that the reality of substances leads naturally to that doctrine.

We move now to the question of the justification of the ontological claim that there are primary substances, and that they are as Aristotle describes them. Why should we think that Aristotle’s metaphysics of substance (and all that goes along with it) is correct?

### The Threat of Circularity

The answer to the last question is apparently that Aristotle’s metaphysics of substance is the best explanation (better, specifically, than those offered by his predecessors) of change, something we all experience in abundance. The notion of “best explanation” at work here cannot, however, include criteria of explanation embodied in the doctrine of

the four *αἰτίαι* and grounded in Aristotle's notion of substance, for that would be to employ criteria of explanation derived from a metaphysical thesis in the philosophical defense of that same thesis. It would be useless to argue, for example, that the only way to explain (specific instances of) change is in terms of the four *αἰτίαι*, that this is the only acceptable explanation because it is the only one based in the reality of substance, and that we are justified in thinking substance is real because it is the only acceptable way to explain change (in terms of material, efficient, final, or formal causes). So when Aristotle criticizes his predecessors for ignoring the very cause (*αἰτία*) of change (*Metaphysics* 988b27), he should not be referring to his four *αἰτίαι* if his only justification for accepting only those four *αἰτίαι* is his doctrine of substance. Likewise with Plato, justification for the metaphysical doctrine of the Forms should not come from demands for explanation where the criteria for acceptable explanation are themselves grounded in a metaphysics of the Forms. The explanatory criterion that led Plato to the Forms was, we saw, the demand that explanation be in terms of what is *best* for a certain thing in a certain respect.

Allowing one to justify a metaphysical or ontological thesis with an argument to the best explanation, where that argument employs standards of explanation depending for *their* justification on the thesis under dispute, would allow one to justify any metaphysical or ontological thesis. I could claim that, because gremlins are at work in all mechanical and electrical equipment, the only appropriate way to explain mechanical or electrical phenomena (the break-down of a machine, for example) is to explain how and why (for what purpose) the gremlins brought it about. When pressed to defend the

claim that there are gremlins, it would be unacceptable for me to say that they offer the best (or only) explanation of mechanical and electrical phenomena *if* I am ruling out all other contenders for explanation simply on the basis of a prior commitment to the standards of explanation I just laid out above (e.g., reference to purposes).

It would be quite bizarre for one actually to do this. Typically, one would offer *other* criteria for explanatory goodness if one were to defend the thesis that there are gremlins with an inference to an explanation. Perhaps one would say that gremlins offer a better explanation than the electron theory because various electrical phenomena (let's suppose) exhibit goal-directed behavior, and electrons cannot consciously pursue goals whereas gremlins can. This explanation is, of course, ridiculous, but it shows clearly how we move away from the assumption that there are gremlins and move toward a discussion of what aspects of experienced phenomena call for explanation and what *kind* of explanation they call for. We move away from definitions of 'explanation' based on ontological or metaphysical assumptions, and we turn to the question of the *source* of our intuitions about adequate explanation that are at work in an inference to an explanation as the justification of an ontological or metaphysical thesis.

### Aristotle's Arguments for Substance based on Explanatory Considerations

What standards of explanation led Aristotle to his metaphysics of substance? How did Aristotle determine what called for explanation and what kind of explanation it called for? I think we get at least the beginning of an answer, perhaps even a full answer, by looking at Jaakko Hintikka's thesis that "for Aristotle the typical sentences used in

expressing human knowledge or opinion are not among those Quine calls *eternal sentences* (or, even among *standing sentences*) but among those Quine calls *occasion sentences*.... In particular, the sentences Aristotle is apt to have in mind are *temporally indefinite*; they depend on the time of their utterance” (Hintikka 1974, 51). What does this mean, and how does it offer support for Aristotle’s notion of substance?

First, Hintikka’s thesis is that Aristotle did not use anything like the modern notion of a *proposition*. This means that he did not claim that the *meaning* expressed by an uttered or written sentence is a separate abstract entity that is eternal and unchanging. Modern logicians and semanticists (non-Quineans; Hintikka supplies the Quinean counterparts) often use the notion of a proposition in the following way: When I utter “I am [tensed verb] here,” I express with those words through that utterance the proposition, “Dwayne Hudson Mulder is [tenseless or disjunctive verb] in his office at 2:30 pm on the 28th of November, 1995.” That is the *meaning* of the words I uttered. That proposition is eternal and unchanging in the sense that, if it is true, it is eternally true. Whenever anyone expresses that proposition, typically with a tensed verb, they express a true proposition. The truth-value of a proposition cannot change. One can express the same proposition with different words. One can express it in different languages, with different word order, etc. Further, one can express *different* propositions with the *same sentence*. When I utter on Tuesday morning at home “I am here,” I express a different proposition from when I utter on Wednesday afternoon in the classroom “I am here.” For Aristotle, however, the criterion for differentiating meanings was different. He did not employ the notion of a proposition to differentiate

meanings; rather, he took meanings to be mental entities, namely thoughts (Hintikka 1974, 53). If I am expressing the same *thought* with the words “I am here” on Tuesday morning at home and on Wednesday afternoon in the classroom, then the two sentences are the same, or, more precisely, the two utterances have the same meaning. To a modern logician, they cannot have the same meaning because they express different propositions.

This might not appear to cause any philosophically significant problems if we consider only the sentence, “I am here,” because that sentence is true whenever it is uttered. (I assume one cannot be, in a given moment, somewhere other than where one is speaking, although this might some day be possible after the fashion of Daniel Dennett’s (1978) “Where Am I?” In that case one might truthfully utter, “I’m over there in that vat.”) Other sentences, however, produce strange results. The sentence, “Socrates is sitting,” for example, would express a truth when uttered while Socrates sits, and it would express a falsehood when uttered while Socrates does not sit. Given Aristotle’s criterion for the identity of meanings, that sentence has the *same* meaning when Socrates sits and when he does not sit, assuming it expresses the same thought in either instance (Aristotle would think it does; see Hintikka 1974, 53). Thus, the same sentence (meaning) can change truth-values as the result of changes in the world only (the concrete, non-linguistic world).

In general, *temporally indefinite* sentences have this trait (although neither *all* such sentences nor *only* such sentences have this trait; “I am here” is temporally indefinite but does not, barring Dennett-style counterexamples, have this trait, and “I am at place *p* at

time  $t$ " is not temporally indefinite but does change truth-value with a change in speakers due to the variable reference of the first-person pronoun). Temporally indefinite sentences are simply sentences that do not explicitly refer to a temporal index for what they assert to be the case. "Socrates sits" is temporally indefinite. "Socrates sits at high noon on such-and-such a day of such-and-such a year" is not temporally indefinite. Hintikka's thesis, then, is that Aristotle took temporally indefinite sentences to be the standard vehicle of expression for knowledge and opinion. He did not see them as being deficient in any way, whereas a modern logician or semanticist clearly would.

How does this broadly logical point relate to the metaphysical doctrine of substances? In conjunction with an epistemological doctrine, it leads to something like Aristotle's notions of substance and form. Aristotle, along with most other Greek philosophers, thought that one could not have genuine knowledge of things susceptible of change *per se*. We can have scientific knowledge of changeable things, according to Aristotle, but *how* we can have such knowledge requires explanation. Aristotle accepted the view that if everything were in a state of Heraclitean flux, nothing would be knowable. Yet things are knowable. So the question is: what exists that is not in a state of Heraclitean flux? Aristotle, Plato, and others thought that it could be nothing perceptible. Substance and form fill the bill.

Hintikka claims that "the assumption we are dealing with ... helps us to understand ... the widespread Greek doctrine that we can have genuine knowledge only of what is eternal or at the very least forever changeless" (1974, 58). Because temporally indefinite sentences can change truth-value while maintaining the same

meaning, purported knowledge expressed in temporally indefinite sentences is threatened if the object of that purported knowledge is changeable. For example, one might see Socrates sitting and form the belief that Socrates is sitting. One then can be said to *know* that Socrates is sitting. If, however, one leaves the room, Socrates stands up, and one continues to hold the belief that Socrates is sitting, one's knowledge has dissolved into mere false belief.

Today, philosophers would say that we are dealing with two different beliefs: (a) the belief that Socrates sits at time  $t_1$ , and (b) the belief that Socrates sits at time  $t_2$ . Belief (a) is justified (for  $S$ ) by adequate evidence, is true, and is therefore (let's say) knowledge. It continues to count as knowledge at time  $t_2$  also (and all subsequent times as long as the justification is not overridden for  $S$  and the belief continues to be held by  $S$ ). Belief (b) is false and is therefore not knowledge. It never was knowledge either (nor will it ever be). Aristotle analyzes the situation differently. For him, a single sentence-meaning, or belief, has changed from being true to being false while remaining the same sentence-meaning. Only the facts regarding Socrates' position have changed. So for Aristotle a single belief can be a piece of knowledge at one time and become non-knowledge at some later time. One who holds the belief can be completely unaware that it has changed from true to false (unless one remains in constant perceptual contact with the object of that belief). So, belief about changeable things, even if it is knowledge, is not fixed or *secure* knowledge. Given Aristotle's assumptions, the only way to remedy this vulnerability of knowledge to cease to be knowledge without our knowing it is for the object of belief to be something changeless.

Aristotle took temporally indefinite sentences to be standard expressions of belief. He held, therefore, that our knowledge of things changeable extends only as far as our current direct apprehension of those things. One can know, for example, that Socrates is sitting only as long as one *sees* Socrates sitting.<sup>7</sup> When things changeable are out of range of present experience, it is *possible* that they are no longer as we last experienced them to be; Socrates might have gotten up. Yet our scientific knowledge seems to extend far beyond the narrow band of present experience. Aristotle might very well have endorsed an argument of the following sort: We have knowledge of things beyond the scope of present experience; the only way we can have such knowledge is if there are changeless things that can function as the objects of our beliefs; therefore, there are changeless things that can function as the objects of our beliefs. This looks like a transcendental argument for an ontological claim based on the actuality of, and necessary conditions of, a certain type of knowledge.

Whether Aristotle actually considered such an argument need not worry us here. The argument itself depends on the truth of a substantial epistemological claim—the claim that, in an important sense, one cannot know a changeable object when one is not perceiving it. That is, the primary objects of knowledge, if there are any, must be changeless. This is a thesis a contemporary epistemologist, employing the notion of a proposition, can easily reject (using Aristotle’s notion of an object of belief, a notion that

---

<sup>7</sup> I am ignoring, for the sake of simplicity, the justificatory role of testimony. If we take reliable testimony into account, this merely extends the boundaries of our knowledge of the changeable to our own present experiences plus the present experiences of those whose testimony is presently available to us.



excludes propositions as being the objects of belief). Given this epistemological assumption, based on the linguistic position above regarding temporally indefinite sentences, the existence of genuine knowledge regarding changeable things presents a problem demanding an explanation. The postulation of a changeless metaphysical reality, such as Plato's Forms or Aristotle's secondary substances, offers just such an explanation. This, then, is another argument one could offer in defense of the metaphysical claim that there are changeless objects that are not perceptual objects. The argument depends crucially on assumptions about what needs explanation and how to explain it, and those assumptions are necessitated by deeper epistemological positions. The epistemological positions are in turn the consequence of what might seem at first to be insignificant epistemological and linguistic conventions.

In sum, we have found in Aristotle and in Plato an important pattern of interaction between standards of explanation and existence claims. The pattern is clearer in Aristotle. First, a doctrine of explanation can function in arguments for existence claims (ontological claims) to justify inferential steps. One can argue that a thing exists from the premise that the best explanation of an observed fact postulates that thing. This is a sense of 'explanation' closely associated with causation, or at least with the (possibly) broader notion of *bringing something about*. For Aristotle, a full explanation must cite all four types of cause, and a commitment to a causal claim is a commitment to the existence of a cause or an agent to bring something about. For Plato the Forms are causes. In this way, explanation supports existence claims, such as the claim that God exists (as a certain sort of cause or agent of change), or the claim that atoms exist.

Second, a doctrine of what is required for full explanation can *receive* support from existence claims, claims better described as metaphysical rather than ontological. They are claims not about particular existents but about the nature of reality at a fundamental level. We saw, for example, how Aristotle's metaphysics of substance supported his doctrine of explanation in terms of the four causes. A general conception of the world will put constraints on (while not uniquely determining) what can count as an explanation of something that happens *in the world*.

Third, one's general conception of the world and its ways itself receives support through an appeal to explanation; it is generally a response to the explanatory need for ordering and making sense of experience. Aristotle explicitly recommended his metaphysics of substance as a way of making sense of the world of experience that was superior on explanatory grounds to all known competitors. This is a sense of 'explanation' with an unclear connection to causation, however. Inferring from the occurrence of table-experiences to the existence of a table *causing* those experiences is closely akin to inferring the existence of a pedestrian as the cause of tracks in the snow. It is something different, however, to infer from the *general* character of experience to a materialist metaphysics, for example, or an idealist metaphysics, or an Aristotelian metaphysics. One might argue, for example, that since experience presents us with two fundamentally different kinds of phenomena, the mental and the physical, a dualistic metaphysics is the only acceptable explanation of the world as presented in experience (i.e., monism cannot adequately explain the character of our experience of the world).

Fourth and last, standards of explanation used to evaluate metaphysical positions seem to receive yet further support, but in an unusual and still fairly unclear way. Aristotle's conception of what aspects of experience need explanation and what counts as a better or worse metaphysical explanation seems to have depended in some way on apparently unrelated (and *prima facie* neutral) epistemological and semantic beliefs. If one uses a modern notion of a proposition, one will not agree with Aristotle about what aspects of experience require explanation and about what sort of explanation is best for those aspects of experience that do require explanation. At this level, we seem to be open to many and varied types of influence, not only from epistemology and semantics and logic, but also from religion, morals, social conventions, political convictions, and so on.

One well-known and much-discussed aspect of the revolution of modern science is the abandonment of many Aristotelian-Scholastic explanatory projects for others. Before Newton, deviation from circular motion in celestial mechanics required a special explanation, whereas circular motion required no special explanation. After Newton, circular motion and every other deviation from motion in a straight line requires a special explanation, whereas movement in a straight line requires no special explanation. There have already been many studies of the function of religious, social and political influences in bringing about such revolutions in a general scientific outlook.<sup>8</sup> We should note a general openness to many different types of influence over a person's thinking about what requires explanation and what counts as explanation at this fundamental level.

---

<sup>8</sup> See, for example, E. A. Burtt 1932 and H. Butterfield 1957.

It is easy to think that this is determined only by science itself, but it is actually pre-scientific (logically prior to scientific reasoning). In Aristotle it appeared to be a peculiarly Greek combination of Heracliteanism about the world of sense experience and the use of temporally indefinite sentences in epistemology that led him to feel the peculiarly Greek explanatory needs that he (and Plato and the atomists, among others) felt.

### Rationalists on Adequate Explanation

Plato and Aristotle present one with many materials for constructing arguments having the form of inference to an explanation for their metaphysical conclusions. They even present substantial evidence that they themselves arrived at their distinctive metaphysical doctrines on the basis of assumptions about adequate explanation of experience. Their official statements might emphasize other methods of coming to believe in the truth of their metaphysical systems, but concerns about giving explanations of problems presented in experience are never far from view. Plato, especially, emphasizes intellectual intuition of the Forms, in the metaphor of the cave, for example. It is difficult to shake the feeling, however, that such intuition comes into focus, or even comes into being, only after one has accepted the existence of the Forms on the basis of their explanatory power.

The situation seems different with more recent metaphysicians. Many philosophers since the demise of the ancient Greek schools, such as Scholastics and Rationalists, who have aspired to construct metaphysical systems have developed techniques constituting what we now call *deductive metaphysics*, and deduction seems to

have no basis in criteria of adequate explanation. Spinoza claimed to reach all his conclusions deductively, after the model of Euclid's *Elements*. According to Stuart Hampshire, "[N]o other philosopher ever insisted more uncompromisingly that all problems, whether metaphysical, moral or scientific, must be formulated and solved as purely intellectual problems, as if they were theorems in geometry" (Hampshire 1951, 24). Deductive metaphysics starts with self-evident principles and proceeds from these by deduction, or demonstrative reasoning, alone. The self-evident principles might be non-existential. The ontological argument<sup>9</sup> begins with the principle, "that which exists actually and mentally is greater than that which exists only mentally," as Aquinas puts it (*Summa Theologica*, Question 2, Article 1, Objection 2; Aquinas 1948, 21). The self-evident principles might also be existential. Descartes began with a premise supposed to be both self-evident and existential, "*cogito, sum*".<sup>10</sup> None of this appears to make any use of concepts of explanation.

The attempt to find hidden assumptions about explanation in deductive metaphysics is threatened by the fact that inference to the best explanation is *inductive*. Inference to an explanation can be deductive, however, if one supplies a premise stating

---

<sup>9</sup> This example, and the distinction between arguments that start with self-evident existential premises and those that start with self-evident non-existential premises, comes from Passmore 1961, 2.

<sup>10</sup> I am assuming here that Descartes did not *infer* "I exist" from "I think." If he began only with the premise, *cogito*, and inferred *sum*, his starting-point would apparently be non-existential. On the interpretation of Descartes' famous phrase, see Hintikka 1962.

that there is only one possible explanans for the explanandum mentioned in the premises.

For example,

- 1) The refrigerator runs only when the light switch is on.
- 2) The only possible explanation for (1) is that the light switch is connected to the refrigerator's socket.
- 3) Therefore, the light switch is connected to the refrigerator's socket.

This is a valid deductive argument. Premise (2) really expresses a conditional: If the refrigerator runs only when the light switch is on, then the light switch is connected to the refrigerator's socket. It is a conditional, however, that we base on explanatory demands. The argument then takes the form of a standard *modus ponens*. Inference to the Only Explanation is deductive; Inference to the Best Explanation is inductive.

Why am I looking for assumptions about explanation in the work of deductive metaphysicians? Because I believe that even here, where philosophers try to begin with the most obvious truths and proceed only by steps that are obviously truth-preserving, philosophers are motivated by a need to explain something that appears problematic, and the way they proceed is determined in important respects by what they are willing to accept as an adequate response to the problems they perceive. I agree with F. C. Copleston, a philosopher acutely aware of contemporary attacks on metaphysics yet still strongly sympathetic to the task of rationalist metaphysics:

The metaphysician would claim, I think, that when he affirms the existence of transcendent being (or of metaphenomenal beings like 'spiritual souls'), he is not simply making a gratuitous assertion but is answering a 'problem' which arises out of reflection on empirical reality. (Copleston 1950, 70)

For what value can a metaphysical hypothesis have if it does not account for something or explain something or render something intelligible, ‘something’ being some set of facts about or some aspect of empirical reality? (*ibid.*, 76)

I shall examine deductive metaphysics also because I have found that many philosophers of this sort have much of interest to say about what needs explaining and how properly to explain it. Deductive metaphysicians, more than any others, present ideas about explanation that seem strange to contemporary philosophers with empiricist tendencies. Our ultimate objective is to explore how one can go about defending one’s assumptions about what calls for explanation and what constitutes a proper explanation. People generally do not think much about defending what they believe until it is challenged. Examining positions most different from our own can make us especially aware of the need to argue in support of our own positions, because it is exceptionally clear to us that those positions different from our own need some defense. (Since we do not accept them, we certainly do not take them to be obvious, needing no argument.) Once again, one is reminded of Carnap’s comment: “The nature of one’s opponents often determines the way in which one’s views are expressed.” Typically one bothers to defend a thesis only when one feels a specific challenge to it. So I am trying to challenge as many assumptions about explanation as possible by setting up as many opposing views about explanation as possible.

Metaphysicians since the time of Aristotle have often been greatly concerned with the problem of contingent, or finite, being as such. This interest has by no means disappeared, although it has subsided considerably. Richard Swinburne claims in *The Existence of God*: “That there should be anything (logically contingent) at all is

overwhelmingly strange, when we think about it. But there is something logically contingent. The issue is whether the world and its operations are the stopping-point, or whether we *must* go beyond the world to find the stopping-point of explanation..." (1979, 107-8, emphasis added). Philosophers have often felt that finite being as such somehow requires an explanation, and the attempt to explain it leads to their metaphysical systems. F. C. Copleston explains, "[I]f it is once admitted that the existence of finite things gives rise to a problem, namely the problem of explaining or rendering intelligible their existence, the answer, if any answer is forthcoming, can hardly be anything else but of a metaphysical character, because to admit the problem of the existence of the finite is to admit the problem of the Absolute" (Copleston 1950, 74).

Leibniz provides a very good example of this kind of thinking. In *A Specimen of Discoveries About Marvellous Secrets* of 1686 he writes, "If there were no necessary being, there would be no contingent being" (1973, 76). Clearly, this is supposed to be coupled with what Leibniz and many other metaphysicians think is an obvious premise: there is contingent being. These two premises together lead to the distinctly metaphysical conclusion that there is necessary being. This is a simple *modus tollens*, but there is good reason to describe it as an inference to the only explanation (of contingent being) because of the way Leibniz supports the first premise; he claims it is known to be true because it is required by the demands of adequate explanation.

[A] reason must be given why contingent things exist rather than not exist. But there would be no such reason unless there were a being which is in itself, that is, a being the reason for whose existence is contained in its own essence, so that there is no need for a reason outside it. Even if one went on to infinity in giving reasons for contingent things, yet outside their series (in which there is not a sufficient reason) a reason for the whole series must be found. From this it also



follows that the necessary being is one in number, and all things potentially, since it is the ultimate reason of things, in so far as they contain realities or perfections. And since the full reason for a thing is the aggregate of all primitive requisites (which do not need other requisites) it is evident that the causes of all things can be reduced to the attributes of God. (Leibniz 1973, 76-7)

Leibniz mentions the principle of sufficient reason in this passage, and it is indeed this principle that makes him feel the need to explain what he classifies as contingent being. In the *Monadology* of 1714 he elaborates this principle, “Our reasonings are based on two great principles: the *principle of contradiction* ... and the *principle of sufficient reason*, by virtue of which we consider that no fact can be real or existing and no proposition can be true unless there is a sufficient reason, why it should be thus and not otherwise, even though in most cases these reasons cannot be known to us” (1973, 184). Leibniz is quite explicit in stating that this principle, taken by itself, is adequate for reaching metaphysical conclusions: “Now by this principle alone, to wit, that there must be a sufficient reason why things are thus rather than otherwise, I prove the existence of the Divinity, and all the rest of metaphysics or natural theology, and even in some manner those physical principles which are independent of mathematics, that is to say, the principles of dynamics or of force” (Leibniz’s second letter to Clarke, *circa* 1716; 1973, 207).<sup>11</sup>

Leibniz thinks that for every fact there must be a reason sufficient to make it so. It is difficult to imagine a counter-instance to this principle, i.e., a fact for which there

---

<sup>11</sup> The contemporary version of this argument in Richard Swinburne 1979 is more modest; he claims not to prove the existence of God from contingent being, but only to confirm to some extent the hypothesis that God exists. Leibniz’s claim to produce a *proof* is, of course, what makes him a deductive metaphysician.

is no reason sufficient to make it so. Such a fact would be of an event that is truly spontaneous, uncaused. Some believe that there are sub-atomic events that are spontaneous in this sense. Setting these cases aside for now, we can accept Leibniz's principle (a principle that has clear affinities with his determinism). For every fact, there is a cause, a sufficient reason.

He also thinks, however, that contingent being presents a problem—demands an explanation—and the only acceptable explanation is in terms of a being which is not itself contingent, i.e., a necessary being. It is just this constraint on acceptable explanation of what is presented to us in ordinary empirical experience that leads to a metaphysics of necessary being beyond the scope of any possible (empirical) experience. What is a contingent being, and how does it demand of us an explanation in terms of non-contingent being (*why* must an adequate explanation refer to non-contingent being)? How is a contingent being distinguished from a non-contingent being? These notions are rooted in fundamental modal notions. A contingent being is a being that *could have not been, or could have been otherwise than it is*.

Ignoring here religious or mystical experience, experience presents us only with contingent beings. Necessary being dwells not in the empirically accessible world. Yet it is not given in experience either that a being might have not been or could have been otherwise than it is. Consider an example: I perceive someone wearing a blue shirt. Of course I have the belief that they could have been wearing another shirt at this time. I do not, however, perceive this alternative possibility; I have no alternate residences in other possible worlds. I do not, for example, have a faint shadow experience of them

wearing various other shirts of different colors alongside or “behind” the vivid image of them wearing a blue shirt. I believe that they could have been wearing a different shirt at *this* time because I have in fact seen them at *other* times without this blue shirt on, with different shirts on. Even if I have never seen this person wearing any other clothing, I know that people *do* wear different clothing at different times. Further, there is, I believe, no basis in my experience for postulating a *law* physically necessitating a person to wear any particular shirt at a particular time. I know of reasons one might have to wear certain clothing at certain times, but not as a matter of physical (or any other kind of) necessity, rather as a matter of imposed conventions freely accepted. So it appears that the basis for the distinction between contingent and necessary being is not given directly in experience. The necessary beings themselves are not accessible through experience, and the identification of things in experience as contingent is based not on what is perceptually given but on *laws* that delimit what is possible from what is impossible.

If we *grant* that there is a necessary being that can cause non-necessary things but is not itself caused by anything else, then it is relatively easy to understand that a complete explanation of any contingent fact must ultimately lead to a necessary being that *explains* something else but does not require an explanation in terms of anything else. If one admits Leibniz’s metaphysics, then one must admit that any complete explanation of a fact necessarily refers to a necessary being. One might rely on a regress argument that finds the terminus for the regress of explanatory factors in something that does not need explanation in terms of something else. In this way the metaphysics supports the

notion of explanation. But we again face the threat of a vicious circle in our philosophical arguments *if* we use that very notion of explanation to support the metaphysics.

One cannot cogently argue as follows:

a) The only way adequately to explain any contingent event presented in empirical experience is to refer it to a non-contingent meta-empirical entity; therefore such a non-contingent meta-empirical entity exists.

b) The reason adequate (complete) explanation must refer to a non-contingent meta-empirical entity is that any other explanation would terminate in a contingent entity which itself demands further explanation (any explanation referring only to contingent beings leaves out the true explanatory factor which can explain other things but is itself not in need of further explanation).

Proposition (a) presents an inference from the demands of adequate explanation to a metaphysical conclusion. Proposition (b) argues for criteria of adequate explanation on the basis of a thinly veiled metaphysical assumption, i.e., that there exists a non-contingent explanatory factor which can explain without needing to be explained. This is a vicious circularity.

If it is explanatory demands that support the metaphysical conclusion, where can we look for justification of the relevant notion of explanation without assuming the metaphysical conclusion? I believe Leibniz (and others) try to ground it in an analogy between explaining one fact with another and explaining *all* facts. It seems intuitively clear that whenever one fact could have been otherwise there is an explanation of why

it is as it is and not otherwise. When alternatives are possible, there is an explanation (whether anyone considers it or not) why the alternative possibilities did not actualize. Now, one concedes that there might be an infinite regress of these contingent facts, but whether the totality of facts is finite or infinite, one can imagine the totality itself being other than it in fact is. So it seems that the totality of contingent facts has to be explained, and given the principle that the explanation of something contingent must be something *else*, the explanation of the totality of contingent facts cannot itself be among them, i.e., it cannot be a *contingent* fact.

Another principle that is at work here is the principle that one cannot explain in terms of the unexplained. If I explain one contingent fact in terms of another contingent fact, its proximate cause, that is itself unexplained (unexplained in fact, although there is an explanation for it), my explanation of the first contingent fact cannot be considered complete; it leaves something yet to be explained. It seems that this principle is behind Leibniz's statement: "But there would be no such reason [for the existence of contingent things] unless there were a being which is in itself, that is, a being the reason for whose existence is contained in its own essence, so that there is no need for a reason outside it." If there is an infinite regress of contingent things to be explained, the task of explanation will be infinite and nothing will ever be explained, unless explanation can be grounded in something self-explanatory, "containing within itself the reason for its own existence," as Copleston also puts it in a debate with Bertrand Russell (1948, 406).

These two principles are conceptual commitments. One holds (a) that, just as one fact that might have been otherwise must have an explanation outside itself, the totality

of all facts in the world, because *it* could have been otherwise, must have an explanation outside itself. One also holds (b) that explaining one contingent fact in terms of another contingent fact cannot be complete because it leaves us with something requiring further explanation. There are some reasons to reject both of these principles, but we will save our criticism for later. Now we just want to try to understand what might lead one to accept these principles.

The first one seems plausible because it seems plausible to refer to the fact of the world, the fact of the existence of the world as a whole, even if it happens to be infinite. We can refer to specific facts, such as the fact that Andy Rooney has overgrown eyebrows, and to facts of broader scope, such as the fact that life has evolved on the planet Earth. We can get still broader in scope. There is the fact that very many galaxies formed in the universe, and the fact that they tend to cluster in groups rather than being spaced out evenly. It seems natural, through an extrapolation of the gradual increase in breadth of facts, to suppose that at the end of the line of greater and greater breadth there is the broadest fact of all, the fact that the universe exists and that it exists just as it is in its entirety.

The second principle seems natural as well, and there are two ways to approach it. First there is what I will call *the subjective approach*. We are all familiar, I suppose, with the experience of asking for an explanation of something, being told something that is equally puzzling, and feeling that we haven't got any explanation at all. I once heard a comedian relate a story of being pulled over for speeding by the police. The police officer asked him, "Why were you driving so fast?" "I had the gas pedal to the floor,"

he replied. One wants to know why he was speeding and is told that the depressed gas pedal caused the speeding. The depressed pedal is indeed an explanatory factor, but one does not know why the pedal was depressed. The depressed pedal explains the speeding, but as long as one has no explanation for the depressed pedal one still lacks an explanation of the speeding. Explanation in terms of the unexplained does indeed seem to be no explanation at all.

Perhaps a better example would include an explanation that is not *understood*. Many people feel baffled when confronted with complicated and obscure scientific explanations. One might wonder, for example, why a helium-filled balloon moves forward in an accelerating cabin. If they are told that it is because of Einstein's Principle of Equivalence and told what this principle states, they might feel that they received no explanation at all because the purported explanation referred to something itself needing explanation. The psychological sense of satisfaction associated with receiving an explanation and understanding is impossible without the explanation terminating in something that does not itself stand in need of explanation in the given context. The subjective experience of getting an explanation must be grounded in something that is already explained or for some other reason requires no explanation. Explanations can be built only on other explanations, and the ultimate foundation for this seems to be the self-explanatory. We understand certain things immediately, it seems, and such things must serve as the starting-point or foundation for explanations of anything else.

The second approach to this principle I will call *the objective approach*. The first approach took explanations to depend for their existence on being considered by

someone, and it took the subjective sense of understanding to be an important criterion for determining whether some account is in fact explanatory. In short, we said that an explanation in terms of the unexplained fails to be an explanation at all because it is incapable of generating the sense of understanding that typically accompanies successful explanation. In the objective sense, we do not suppose that explanations depend for their existence on being considered by someone. In other words, there can *be* an explanation for a fact although no one has ever thought of or considered it, i.e., although it has never produced a sense of understanding for anyone. In this sense, explanations are facts. The fact that an alternating electrical current is running through the wiring of the desk lamp explains the fact that the lamp is giving off light, even if no one is thinking about these facts. The principle that explanation in terms of the unexplained is incomplete means, first, that for any fact mentioned in an explanation there must be a fact that explains it; in other words, there are no unexplained facts. The principle is meant to say something more, however, in so far as it refers to “the unexplained.” It says that when one actually *does* come to consider these facts, and tries to *give an account* of the fact-to-be-explained in terms of the explanatory fact, that account will be incomplete if it terminates in what we can call a “brute fact,” a fact whose explanation has not been included in the explanatory account. The account is incomplete just because an explanation for the fact objectively exists but has been left out of the explanatory account. The incompleteness, on this interpretation, has nothing to do with a lack of understanding on someone’s part. It has rather to do with an objective deficiency of an account of some part of reality.



What have we found so far? Let's call the metaphysics of contingent and necessary beings the Leibnizian metaphysics (although I think F. C. Copleston gives a clearer, contemporary expression of the distinctive style of reasoning I wish to focus on, both in his debates with Ayer and with Russell and in his 1950). This metaphysics is the basis of one version of the argument from contingency for the existence of God, an argument relying on considerations of what existential commitments adequate explanation requires. The relationships between existence claims and concepts of explanation is somewhat confused here; they are not as easily delimited as they were for the case of Aristotle and Plato. We can, however, try to isolate conceptually the grounding of a doctrine of explanation on metaphysical theses and the grounding of metaphysical commitments by way of explanatory considerations.

First, the Leibnizian metaphysics supports the claim that explanation of phenomena in general must be in terms of a transcendent reality. This resembles the way the Aristotelian metaphysics supports Aristotle's doctrine of adequate explanation as presented in his account of the four causes. If we grant that the universe is a realm of contingent being and there is a necessary being outside this realm, it is easy to see that the necessary being is the ultimate explanation of the universe and that explanation terminating in something contingent is in some way deficient. An explanation might have some *usefulness* even when it does not terminate in the necessary being as ultimate cause, but it will not be a *complete* explanation even of the particular explanandum at hand. Local, incomplete, explanations can serve certain purposes. Explanation in the fullest sense of the term, however, must eventually lead to the ultimate explanatory factor,

something that transcends the universe of contingent beings. It is also very natural, given this metaphysics, to assert that the entire universe of contingent things stands in need of explanation, for one can easily see where the explanation would come from.

But now why should one accept the Leibnizian metaphysics in the first place? Is the Leibnizian metaphysics recommended by (or even demanded by) explanatory considerations? Yes it is, in a way. One might try to begin free of commitments about the existence of God and argue that God exists on the basis of the explanatory power of such a hypothesis. Apparently, the demands of adequate explanation lead to the metaphysical conclusion that God exists. The following quotes are illustrative:

[T]he theist claims that the various phenomena which he cites in evidence of theism require explanation, and he claims that theism allows us to understand why these occur, and is itself a much more natural stopping-point for explanation than are the original phenomena (Richard Swinburne 1979, 90).

Well, my position is the affirmative position that [God] actually exists, and that His existence can be proved philosophically.... So, I should say, in order to explain existence, we must come to a being which contains within itself the reason for its own existence, that is to say, which cannot not exist (F. C. Copleston, in Russell and Copleston 1948, 405-406).

[T]he existence of phenomena in general requires some explanation, and I should say explanation in terms of a transcendent reality (F. C. Copleston, in Ayer and Copleston 1949, 145).

The quote given earlier from Leibniz's *A Specimen of Discoveries* expresses exactly the same line of reasoning. Leibniz expresses similar ideas elsewhere, e.g., in his *Résumé of Metaphysics* of 1697.

There is a reason in Nature why something should exist rather than nothing. This is a consequence of the great principle that nothing happens without a reason, and also that there must be a reason why this thing exists rather than another. This reason must be in some real entity, or cause.... This entity must be necessary; otherwise a cause must again be sought outside it for the fact that it exists rather

than does not exist, which is contrary to the hypothesis. This entity is the ultimate reason for things, and is usually called by the one word 'God'. (1973, 145)

This seems innocent enough—arguments for existence claims based on the demands of explanation, just as one might argue that a pedestrian has been this way since that hypothesis best explains the observed footprints in the snow. But we must beware of the threat of circularity. We must ask why explanation of something contingent in terms of something else that is contingent is unsatisfactory. As Copleston puts it, “[I]f we proceed to infinity in that sense [explaining contingent things in terms of contingent things], then there’s no explanation of existence at all” (Russell and Copleston 1948, 406). Leibniz himself expresses it with a useful analogy in *On the Ultimate Origination of Things* of 1697.

For neither in any one single thing, nor in the whole aggregate and series of things, can there be found the sufficient reason of existence. Let us suppose the book of the elements of geometry to have been eternal, one copy always having been written down from an earlier one; it is evident that, even though a reason can be given for the present book out of a past one, nevertheless out of any number of books taken in order going backwards we shall never come upon a full reason; since we might always wonder why there should have been such books from all time—why there were books at all, and why they were written in this manner. What is true of the books is true also of the different states of the world; for what follows is in some way copied from what precedes (although according to certain laws of change). And so, however far you go back to earlier states, you will never find in those states a full reason why there should be any world rather than none, and why it should be such as it is. (1973, 136)

One can rightfully question these claims. That is, one can question the second principle of explanation I identified above, that explanation in terms of the unexplained is no explanation at all. For I can explain an event simply with the information *that* a sufficient proximate cause occurred without adding *why* that proximate cause occurred.

I *can* explain why the car was travelling at a high speed with the claim *that* the accelerator was fully depressed, without needing to explain further *why* the accelerator was fully depressed. The feeling that one has no explanation in that earlier example comes not from a problem with explaining in terms of the unexplained, but from a problem of ambiguity regarding what is to be explained. When one asks “Why were you going so fast?” in that situation, one typically is asking why *you* were driving the car so fast (i.e., depressing the accelerator), not merely what *caused* the car to go so fast. There is in principle no problem with explaining in terms of the unexplained. One needs to understand *that* a cause has occurred but not necessarily *why* the identified cause occurred.

Leibniz or Copleston might fall back on the first principle I identified earlier, that in addition to particular contingent things or events the entire universe of contingent things and events requires an explanation. They might say that the problem with an infinite regress of explanations is not that each particular explanation in the chain is somehow deficient, but that the infinite series of explanations cannot in principle explain the universe as a whole (or, as it is sometimes put, explain why there is something rather than nothing). We can question this principle as well, however. That is, we can question the legitimacy of asking for (or the meaningfulness of talk about) an explanation for the universe as a whole. Russell and Ayer both attack Copleston on this point. Russell responds to Copleston: “I don’t think there’s any meaning in it [the question of the existence of the whole universe] at all. I think the word ‘universe’ is a handy word in some connections, but I don’t think it stands for anything that has a meaning.... I

shouldn't say the universe is unintelligible—I think it is without explanation.... [F]or my own part, I do think the notion of the world having an explanation is a mistake" (Russell and Copleston 1948, 408-410). Ayer takes a similar line of attack (Ayer and Copleston 1949):

[T]here's a peculiarity about these 'why' questions (p. 144). [Y]our very notion of an explanation of all phenomena is self-contradictory (145). [I]t's precisely questions like this question about the world as a whole that I think we should rule out (146). [T]hese [metaphysical] answers are given not as explanations of any particular event, but of all events. And I wonder if this notion of an explanation of all events isn't itself faulty (149). My objection to the kind of statements that we've agreed to call metaphysical is that they don't explain.... I don't want to say that it isn't an accurate explanation. What I want to say is that it isn't an explanation at all (167).

Clearly there is room to reject the first principle, to deny that there is this sort of *explanatory need* with respect to the universe as a whole, or the question of why there is something rather than nothing. Yet very many philosophers have championed just this explanatory need as a central philosophical concern, and many continue to do so (see, e.g., Parfit 1992).

Whence this felt explanatory need? In the case of Aristotle, we found that the felt explanatory need, for which his metaphysics of substance was an answer, came from an assemblage of epistemological, logical, and quasi-metaphysical (Heraclitean) principles. This was acceptable, for it moved the discussion forward to the topic of what reasons Aristotle or anyone else might have for accepting those principles. The case before us now seems to be significantly different. The only apparent source of the felt explanatory need we've been discussing is acceptance of what I've chosen to call the Leibnizian metaphysics. If one accepts the Leibnizian metaphysics, then the two principles of

explanation make sense and are defensible. But this brings us full-circle to where we started, and it's a vicious circle.

If one accepts the Leibnizian metaphysics, then the distinctive explanatory principles are well-motivated, but they seem unmotivated if one rejects the metaphysics. If one accepts the distinctive explanatory principles, then the metaphysics appears well-supported, but it seems to be unsupported without the distinctive appeals to explanation. Hence, it is illegitimate to accept both the metaphysics and the doctrine of explanation solely on the basis of their mutual support, for this is just the sort of vicious circularity of philosophical reasoning that we considered but successfully avoided in our examination of Aristotle.

I consider this example of the interplay between existence claims and principles of explanation to be an exemplary case of the danger of a vicious circularity in ontological reasoning. It generates purported proofs for metaphysical existence claims, but they are ultimately without any independent support and hence no proofs at all.

I have claimed that the metaphysical theorizing of Plato, Aristotle, and Leibniz has been deeply influenced by implicit convictions about the nature of explanation, although I do not think that these philosophers would have characterized their metaphysical reasoning in just the way I have. Theorizing about what exists has, in our own century, often relied much more explicitly on considerations of explanation. The question of what exactly explanation is has also received much more direct attention than it ever had before. I turn in the next chapter to a brief survey of some of the more

prominent uses of abductive inference, or “inference to the best explanation,” in various areas of contemporary ontology, and I shall then consider general questions about the necessary place of a doctrine of explanation in all reasoning about what objectively exists.

## CHAPTER 4

### EXPLANATION IN CONTEMPORARY ONTOLOGY

#### Examples of Explanation in Contemporary Ontological Reasoning

We have looked at some prominent metaphysicians of the past and found in their work substantive assumptions about the nature of explanation. We have made some suggestions of how such assumptions about explanation might have guided their metaphysical thinking or even supported their metaphysical arguments to some extent.

Turning to twentieth-century philosophy, we find explanation playing a more obvious role in arguments regarding existence, i.e., in ontological reasoning. Inference to the best explanation appears to many to be the most promising candidate for justifying claims to existence. Although there is no universally accepted and clear distinction between metaphysics and ontology, I will use the term 'metaphysical reasoning' to designate the supporting of doctrines of objective existence with arguments claiming either that some truths of existence can be known *a priori* or that some truths of existence can be *deduced* from self-evident, indubitable starting-points (possibly including truths of existence themselves). I will apply the term 'ontological reasoning' to the supporting of doctrines of objective existence with only inductive arguments or deductive arguments from premises known *a posteriori*. The distinctive method of metaphysics, then, is based on *a priori* knowledge and deduction; the distinctive method of ontology



is based on *a posteriori* knowledge and induction.<sup>1</sup> The twentieth century has seen a widespread rejection of metaphysical method as a means of justifying most types of existence claims. Ontological method lingers on, and even flourishes. We shall examine the nature of explanation as it figures in the justification of ontological claims.

There is, apparently, some historical precedent for the use of inference to the best explanation as a justification for ontological conclusions. J. L. Mackie suggests that an inference of this sort served to justify realism regarding material objects for Locke. Discussing the problem of justifying physical-object realism, Mackie claims that the most forceful basis for such a justification comes from the argument “that the real existence of material things outside us is a well-confirmed outline hypothesis, that it explains the experiences we have better than any alternative hypothesis would, in particular better than the minimal hypothesis that there are just these experiences and nothing else” (Mackie 1976, 64). He says that common-sense realism has simplicity to recommend it, “simplicity of the sort that I would call *the elimination of unexplained coincidence...*” (p. 66). And Mackie explicitly recommends these arguments based on best explanation as an interpretation of Locke (pp. 70-71; see also John Locke’s *Essay* [1975], Book IV, Chapter XI, especially §§4-7).

---

<sup>1</sup> This follows somewhat Carnap’s characterization in a 1957 note added to his “Elimination of Metaphysics Through Logical Analysis”: “[I use the term ‘metaphysics’] in this paper, as usually in Europe, for the field of alleged knowledge of the essence of things which transcends the realm of empirically founded, inductive science. Metaphysics in this sense includes systems like those of Fichte, Schelling, Hegel, Bergson, Heidegger. But it does not include endeavors towards a synthesis and generalization of the results of the various sciences” (1932, 80). My distinction here is more at *methods of defense* of existence claims than the nature of the claims themselves.

More recently, in the 1840s, J. S. Mill and William Whewell debated the acceptability of inference to the best explanation as a form of inductive support for scientific hypotheses (interpreted realistically).<sup>2</sup> Whewell supports inference to the best explanation: “We have here spoken of the prediction of facts *of the same kind* as those from which our rule was collected. But the evidence in favour of our induction is of a much higher and more forcible character when it enables us to explain and determine cases of a *kind different* from those which were contemplated in the formation of our hypothesis” (Whewell 1967, 65, quoted in Achinstein 1992, 350). Mill attacks inference to the best explanation. “Most thinkers of any degree of sobriety allow, that an hypothesis of this kind is not to be received as probably true because it accounts for all the known phenomena, since this is a condition sometimes fulfilled tolerably well by two conflicting hypotheses; while there are probably many others which are equally possible, but which, for want of anything analogous in our experience, our minds are unfitted to conceive” (Mill 1858, 328, quoted in Achinstein 1992, 351). We will return to the topic of inference to an explanation as support for scientific theories. Mill does not reject the claim, however, that science provides explanations, as Duhem did later on. Duhem rejected the possibility of explanation in science, for he thought that explanation was essentially metaphysical, appealing to entities beyond any possible experience. Mill, on the other hand, developed a concept of explanation that he thought should be acceptable to a scientifically-minded empiricist philosophy. In this respect and many others, Mill’s

---

<sup>2</sup> See Achinstein 1992.

work resembles the later work of Carl Hempel on scientific explanation, and Hempel notes that his work is an elaboration of Mill's ideas.

Early analytic philosophers gave accounts of the inductive inferences that are supposed to take one from percepts to actual existents. Bertrand Russell, for example, suggests that considerations of simplicity and of better explanation lead inductively to the conclusion that ordinary physical objects exist. "There is no logical impossibility in the supposition that the whole of life is a dream, in which we ourselves create all the objects that come before us. But although this is not logically impossible, there is no reason whatever to suppose that it is true; and it is, in fact, a less simple hypothesis, viewed as a means of accounting for the facts of our own life, than the common-sense hypothesis that there really are objects independent of us, whose action on us causes our sensations" (1912, 22-23). If Russell is concerned in this passage with the discernment of the truth regarding the objective existence of physical objects, then he is asserting some connection between the explanatory power of a hypothesis and the likelihood of its being true.

One could view the Logical Atomism of Russell and the early Wittgenstein as being motivated by concerns with explanation. They took it as their task to explain the strength of the new logic developed by Peano, Frege, Russell and Whitehead, among others (i.e., its superiority over Aristotelian logic). They were particularly interested, it seems, in explaining why language is truth-functional throughout. The general project was to *explain* the appropriateness of the new logic and thereby to *justify* its claim exclusively to embody the correct laws of thought. As J. O. Urmson put it in *Philosophical Analysis*, "Many metaphysical systems which endeavour to give a general

account of the world are clearly modelled on some less high-flying discipline.... The shortest account of logical atomism that can be given is that the world has the structure of Russell's mathematical logic" (1956, 6). As evidence of this, Urmson cites the following quote from Russell:

The kind of philosophy that I wish to advocate, which I call Logical Atomism, is one which has forced itself upon me in the course of thinking about the philosophy of mathematics, although I should find it hard to say exactly how far there is a definite logical connexion between the two.... I shall try to set forth ... a kind of logical doctrine which seems to me to result from the philosophy of mathematics—not exactly logically, but as what emerges as one reflects: a certain kind of logical doctrine, and on the basis of this a certain kind of metaphysic. (Russell 1918, 178)

The key is to understand the relationship Russell describes when he says that the metaphysical doctrine of Logical Atomism was *based on* a logical doctrine. He would himself admit, I believe, that the logical doctrine does not lead "logically" to the metaphysics (just as the philosophy of mathematics does not lead "logically" to the logical doctrine), but the metaphysics results from *reflection* on the logical doctrine. Yet, I believe he would agree that the logic *does* follow "logically" from the metaphysics (e.g., if the world has such-and-such features, then necessarily universal instantiation is valid). The inference from the logical doctrine to the philosophy of Logical Atomism looks somewhat like that form of nondeductive inference we call inference to the best explanation. There is some reason to suppose that Russell's "reflection" on the Fregean logic and the philosophy of mathematics led him to adopt Logical Atomism as the best explanation (and justification) of the former.

Russell claims, in his Introduction to Wittgenstein's *Tractatus Logico-Philosophicus*, that the question with which Wittgenstein is concerned is: "[W]hat

relation must one fact (such as a sentence) have to another in order to be *capable* of being a symbol for that other?" (Wittgenstein 1921, ix). To answer this question is to explain how meaningful language is possible with an account of how the world represented in language is. Russell characterizes the inference from the structure of an ideal logical language to the structure of the world as follows: "In order that a certain sentence should assert a certain fact there must, however the language may be constructed, be something in common between the structure of the sentence and the structure of the fact. This is perhaps the most fundamental thesis of Mr. Wittgenstein's theory" (Wittgenstein 1921, x). The common structure is given, of course, in the picture theory of meaning. Urmson addresses the question of "how the metaphysics [of Logical Atomism] was connected with analytic practice and why the metaphysics was accepted, in particular why the identity of structure of at least an ideal language and reality was so strongly insisted upon." He answers, in part, that "the metaphysics commended itself largely as the rationale of the view of language and analysis with which it was conjoined" (Urmson 1956, 22). This again suggests that the metaphysics was adopted as an explanatory foundation for the logic and the method of philosophical analysis.

The logical positivists had nothing to do with arguments to the best explanation for metaphysical theses, because they would have nothing to do with metaphysics. For the logical positivist, metaphysical statements, statements about a transcendent external world, are neither true nor false; they are meaningless. Carnap wrote, "In the domain of *metaphysics*, including all philosophy of value and normative theory, logical analysis yields the negative result *that the alleged statements in this domain are entirely*

*meaningless*.... Our thesis, now, is that logical analysis reveals the alleged statements of metaphysics to be pseudo-statements” (Carnap 1932, 60-61). Elsewhere he wrote, “In the realism controversy, science can take neither an affirmative nor a negative position since the question has no meaning” (Carnap 1928, 333). Moritz Schlick expressed the same attitude. “[T]he ‘problem of the reality of the external world’ is a meaningless pseudo-problem” (Schlick 1932, 86). “The denial of the existence of a transcendent external world would be just as much a metaphysical statement as its affirmation. Hence the consistent empiricist does not deny the transcendent world, but shows that both its denial and affirmation are meaningless” (p. 107).

The logical positivists apparently did allow for the meaningfulness of some existential claims, as long as these claims satisfy the verifiability criterion of meaning. Schlick claims that “consistent empiricism does *not* deny the existence of an external world; it merely points out the empirical meaning of this existential proposition.” He even writes that, “logical positivism and realism are not in opposition; whoever acknowledges [the positivist’s] fundamental principle must be an empirical realist” (Schlick 1932, 107). Rudolf Carnap, in “Empiricism, Semantics, and Ontology” (1950), admitted the meaningfulness of ontological statements. He distinguished what he called internal questions from external questions. Internal questions are about the existence of individuals of a certain acknowledged kind. They can be answered “either by purely logical methods or by empirical methods, depending upon whether the [linguistic] framework [in which they are expressed] is a logical or a factual one” (1950, 206). He adds, “The concept of reality occurring in these internal questions is an empirical,

scientific, non-metaphysical concept” (*ibid.*). It is a non-metaphysical concept, but it is a concept of reality. So let’s call it an ontological concept, following Carnap’s lead. External questions concern not a single entity of a given kind, or even a large group of entities, but a “system of entities as a whole.” An internal question might ask if there is a prime number between 1,207 and 1,250 or if there is a corkscrew in the kitchen; an external question might ask if numbers exist, or if the physical world exists. External questions begin to look somewhat like metaphysical questions, but Carnap no longer rejects them as meaningless. Rather, he says that they require practical decisions rather than theoretical answers. “Those who raise the question of the reality of the thing world itself have perhaps in mind not a theoretical question as their formulation seems to suggest, but rather a practical question, a matter of a practical decision concerning the structure of our language. We have to make the choice whether or not to accept and use the forms of expression in the framework in question” (1950, 207).

Turning to ontology, the problem of *justification* forces itself on us immediately. Even if we grant that the “empirical meaning” of an existential claim exhausts its meaning, we must admit that the meaning of an existential claim goes beyond the empirically given (in the sense of what someone has *actually* experienced, regardless of counterfactual possibilities of experience). Even Schlick admitted this in “Positivism and Realism” (1932). When I assert that there is a table before me, I mean more than merely that I am currently presented with certain visual images. I mean also that if I *were* to move to the other side of the table, I *would* have certain other visual images. Schlick says,

Strictly speaking, the meaning of a proposition about physical objects would be exhausted only by an indefinitely large number of possible verifications, and we gather from this that such a proposition can in the last analysis never be shown to be absolutely true. It is indeed generally recognized that even the most certain propositions of science are always to be taken as hypotheses, which remain open to further refinement and improvement. This has certain consequences for the logical nature of such propositions... (1932, 91).

The consequences for the logical nature of such propositions have to do with deep problems in inductive logic and in the nature of laws, and they quickly became the focus of much interest.

What I do when I make existential claims is to accept various *laws* governing not only what experiences I do have but also what experiences I (counter-factually) would have had under different circumstances and what experiences I (subjunctively) would have if different circumstances were to occur. Schlick mentions that “propositions concerning bodies are transformable into equivalent propositions concerning the occurrence of sensations in accordance with laws” (1932, 107). The problem facing any such transformation is that any given set of sensations is in accord with an indefinite number of laws. How does one decide which law to adopt when attempting to project from occurrent sensations to future sensations? Explanation and explanatory coherence do not offer a complete answer, but they provide an important part of a very compelling answer. The key to getting beyond the empirically given and to justifying an existential claim, it seems, can be nothing other than inferring to an explanation.

Having thrown out metaphysics (theorizing about a *transcendent* world) and having insisted that the meaning of existential (ontological) claims be exhausted empirically, positivists focussed on problems of underdetermination. The fundamental



problem is that any given set of experiences underdetermines governing laws. This means that, while a given set of experiences will exclude some laws, it will not exclude all laws but one. This is, of course, just one manifestation of the problem of induction discovered by Hume. Thus, Carnap's extensive later work on inductive logic addresses itself to such problems of underdetermination.

The problem is that a strict empiricist is intensely uneasy about "going beyond the given in experience." Experience, however, proves insufficient to decide many of the questions of science, questions that require the selection of individual laws to the exclusion of all competitors. Empiricists have a great concern to preserve the best science and account for it, so the positivists reluctantly turned to the search for criteria not found immediately in experience to account for theory selection in science. They began investigating what we can most appropriately label "theoretical virtues," those features of theories independent of their fit with experiential data that can either recommend or count against them. Among such theoretical virtues, I would like to focus on simplicity, explanation, and explanatory coherence. Simplicity is often taken to be an explanatory virtue—the simpler hypothesis is, *ceteris paribus*, the better explanation. Explanation is how a theory relates to the data already collected, and two competing hypotheses can both account for all the available data while one nonetheless provides a *better* explanation of the data. Explanatory coherence concerns how well a hypothesis explains or is explained by other, previously accepted and independently supported hypotheses.

Inference to the Best Explanation as the Basic Form of Nondeductive Inference

As attention focused on theoretical virtues, philosophers began to look more closely at inferences based on those virtues. In 1965 Gilbert Harman published “The Inference to the Best Explanation,” where he argued that “in cases where it appears that a warranted inference is an instance of enumerative induction, the inference should be described as a special case of another sort of inference, which I shall call ‘the inference to the best explanation’” (1965, 88).<sup>3</sup> Harman not only facilitated the identification of inferences of this sort and unified discussion of it, he also argued that this is “the basic form of nondeductive inference” and argued that it is a *warranted* form of inference.

I am most interested in Harman’s claim that the inference to the best explanation is the basic form of nondeductive inference. If this claim is true, it supports my assertion that anyone who rejects metaphysical method (as I defined it above) but accepts ontology will *have* to rely on inferences of this sort and will thus have to employ a particular concept (or particular concepts) of explanation in supporting any claim to objective existence. Although I looked for assumptions about adequate explanation in some metaphysicians of the past, I am perfectly willing to admit that some metaphysicians worked from what they took to be self-evident starting-points through deduction to substantive conclusions regarding objective existence without invoking

---

<sup>3</sup> See also Harman 1968a and 1968b, and Lipton 1991 for more defense of this position. R. A. Fumerton 1980 argues against Harman’s thesis. Fumerton wants “to cast doubt on the claim that there is a legitimate process of reasoning to the best explanation which can serve as an alternative to either straightforward inductive reasoning or a combination of inductive and deductive reasoning (1980, 590). My arguments will not, I believe, depend on either Harman’s or Fumerton’s thesis.

considerations of adequate explanation at any point (Descartes, for example, appears to have proceeded in this way). Ontologists, on the other hand, have available to them no demonstrative methods. They must rely on nondeductive inferences, and if Harman is right, this means they must support their inferences with particular notion(s) of explanation and criteria for goodness of explanation. As Wesley Salmon puts it, “[A]pplication of this principle of inference [to the best explanation] obviously presupposes some explication of *explanation*” (1989, 8).

Harman characterizes the inference to the best explanation as the inference “from the premise that a given hypothesis would provide a ‘better’ explanation for the evidence than would any other hypothesis, to the conclusion that the given hypothesis is true” (p. 89). Harman does not present much of an argument for his claim that inference to the best explanation is the basic form of nondeductive inference. He makes a two-part claim: (a) there are nondeductive inferences that cannot be reduced to enumerative induction, and (b) there are no nondeductive inferences that cannot be set up as an instance of the inference to the best explanation. Regarding the first of these he says,

I do not pretend to have a conclusive proof that such inferences cannot be made out to be complicated uses of enumerative induction. But I do think that the burden of proof here shifts to the shoulders of those who would defend [enumerative] induction in this matter, and I am confident that any attempt to account for these inferences as [enumerative] inductions will fail. Therefore, I assert that even if one permits himself the use of enumerative induction, he will still need to avail himself of at least one other form of nondeductive inference. (1965, 90)

He presents something of an argument for the second claim, that enumerative induction is “superfluous.” He claims that if we describe nondeductive inferences as instances of

the inference to the best explanation, then we can account for which are warranted and which are not.

[O]ne is warranted in making [the inference from “All observed *A*’s are *B*’s” to “All *A*’s are *B*’s”] whenever the hypothesis that all *A*’s are *B*’s is (in the light of all the evidence) a better, simpler, more plausible (and so forth) hypothesis than is the hypothesis, say, that someone is biasing the observed sample in order to make us think that all *A*’s are *B*’s. On the other hand, as soon as the total evidence makes some other, competing hypothesis plausible, one may not infer from the past correlation in the observed sample to a complete correlation in the total population. (1965, 91)

He concludes that “inferences which appear to be applications of enumerative induction are better described as instances of the inference to the best explanation.” This claim, if true, clearly would support the claim that I need, that all nondeductive inference relies on concept(s) of explanation. This implies that ontologists should give a lot of attention to the criteria for explanatory goodness they use in arguing for their ontological conclusions.

R. A. Fumerton (1980) has argued against Harman’s thesis. He claims that all inferences we describe as being to the best explanation can be reduced to “straightforward” inductions or as combinations of induction and deduction. The argument I will develop later in this chapter about the connection between explanation and nondeductive inference depends on neither Harman’s nor Fumerton’s thesis.<sup>4</sup> I am

---

<sup>4</sup> I happen to disagree with Fumerton. I believe that there are warranted nondeductive inferences of the form “to the best explanation” that cannot be reduced to enumerative induction. They occur in situations where there is no available information that can serve as the *basis* of a straightforward induction. For example, inferences based on fingerprinting evidence *now* can be recast as straightforward inductions based on a vast base of observed correlations in the past. When fingerprinting was a new technique, however, there was no such observational base, and the conclusions could be defended only on the basis of considerations, based on quite *general* knowledge of the physical

claiming in this section only that Harman's thesis, if true, would support one of my major theses. My support for that thesis will be independent of Harman's claim, however; I shall use Harman's work as a starting point for an investigation of the connection between notions of explanation and nondeductive inference. First, however, I shall survey some prominent examples of nondeductive inferences in contemporary ontology that are *explicitly* based on notions of explanation.

### Inference to the Best Explanation in the Philosophy of Science

When acceptance of Logical Positivism began to wane substantially in the 1960s throughout Anglo-American philosophy, ontology became much more widely discussed. Realism in particular has been not only the subject of much debate but also the recipient of much support from philosophers, support from explanatory arguments. Ontological discussions have been especially common in the philosophy of science, regarding scientific realism, and in ethics, regarding moral realism. In addition, some epistemologists are still discussing ontological arguments used to support claims to knowledge of such mundane things as desks and trees. In the philosophy of religion, according to Robert Prevost, "a number of philosophers, while admitting of no compelling deductive argument for God's existence, consider religious belief to be justified on the basis of its power to explain some phenomenon or phenomena. Theism, on this account, is an explanatory theory, or hypothesis, whose acceptability is measured by its explanatory power" (Prevost 1990, 1). In the philosophy of mind, philosophers

---

world, about the probable causes of observed fingerprints on, say, a murder weapon.

have dealt with arguments for other minds and with arguments regarding the ontological status, or nature, of the mind and mental entities and properties. Philosophers of mind have also used an explanatory argument to support a so-called language of thought.<sup>5</sup> Finally, in the philosophy of language, proponents of possible-worlds semantics have defended accepting its metaphysical implications on the basis that it is the only satisfactory explanation of intentional phenomena. I will examine the use of explanatory principles in some of these fields.

Clark Glymour presents a strong thesis regarding the place of inference to the best explanation in scientific thinking.

One way to argue to a theory is to show that it provides a good explanation of a body of phenomena and, indeed, that it provides a better explanation than does any available alternative theory. This pattern of argument is not bounded by time or by subject matter. One can find such arguments in sociology, in psychometrics, in chemistry, and in astronomy, in the time of Copernicus and in the most recent scientific journals. The goodness of explanations is a ubiquitous criterion; in every scientific subject it forms one of the principal standards by which we decide what to believe. (1984, 173)

Glymour provides instances to support his far-reaching claims, from Copernicus, Dalton, Einstein, and Spearman. The only thing we need to add to his claims is that he means “arguing to a theory” and “deciding what to believe” in a realist sense as opposed, for example, to an instrumentalist interpretation.

Hilary Putnam was largely responsible for bringing ontological discussions back into fashion. According to Jarrett Leplin, “Hilary Putnam seems to have inaugurated a new era of interest in realism with his declaration that realism is the only philosophy that

---

<sup>5</sup> See Fodor 1975.

does not make the success of science a miracle” (Leplin 1984, 1). The miracle argument for scientific realism can take two forms. First, one can argue that, in order to explain the success of scientific theories (the accuracy of their predictions, etc.), we must say that scientific theories are either true or approximately true; we infer that they are true (or nearly true) because their being true (or nearly true) offers the best explanation of their empirical success. This is a standard argument to the best explanation, and this appears to be the point of some versions of the miracle argument. These arguments presume that claiming the success of science to be a miracle is not an acceptable explanation of the success of science. Those who offer the miracle argument *don't* believe in miracles. Everything *except* realism is said to be impossible because alternatives to realism can explain the success of science only as a miracle, and this is clearly not acceptable (we could say that, at the other extreme, some theists would say that the *only* explanation for anything is that it is a miracle, i.e., an act of God). So Putnam writes, for example, that “the typical realist argument against idealism is that it makes the success of science a *miracle*” (Putnam 1978, 18). If this argument were directed against a theistic idealist such as Berkeley its ineffectiveness would be obvious.

The second form of the miracle argument is an argument not directly for the truth or verisimilitude of successful scientific theories but for the acceptability of inference to the best explanation as an essential part of scientific methodology; scientific method itself is defended on the basis that its overall convergence on truth is the best explanation of its overall empirical success. One can express this argument as follows: If inference to the best explanation as a basis for scientific method did not tend to lead to the truth, the

overall success of science would be a miracle; therefore, scientific method tends toward the truth. In this argument, one is not concerned with arguing for the truth of any one particular theory on the basis that it is successful; one argues for the acceptability of a method (IBE) for justifying scientific theories on the basis that this method has been generally successful and the best explanation of its success is that it is a truth-conducive form of inference. Richard Boyd offers one version of such a position. "I have offered a defense of scientific realism against empiricist antirealism which proceeds by proposing that a realistic account of scientific theories is a component in the only scientifically plausible explanation for the instrumental reliability of scientific methodology" (1984, 58).

Elsewhere, Boyd gives a clear picture of how and why many philosophers of science have rejected the verificationist rejection of metaphysics and taken up discussions of experience-transcendent reality, specifically unobservables. Verificationist attempts to come up with a rational reconstruction of actual scientific practice without any metaphysical elements ran into apparently insurmountable difficulties. "Scientists appear to be studying unobservables as well as observables, and it is not easy to give a rational reconstruction of their practice to eliminate their apparent concern for metaphysics. That is some argument for realism ..." (1991, 14). Boyd goes on to give an argument for realism based on considerations of how adequately to explain the success of scientific methods.

As realists have examined these questions, a central theme has emerged that represents an attempt to articulate in philosophically appropriate detail ways in which the theory-dependence of scientific practices might support realism. According to this central theme, the case for scientific realism is best put by



arguing that the methods of science are so theory-dependent that it is impossible to explain *scientifically* the *instrumental* success of scientific methods except on the hypothesis that in mature sciences the background theories that determine methods are approximately true of unobservable (as well as of observable) entities, and that the operation of the methods they determine tends to produce subsequent improvements in such approximations to the truth. Scientific realism is thus to be defended as a scientific hypothesis—an empirical claim about the way in which scientific methods work in the world. (1991, 14)

This is another expression of the argument that the best (or only) way to explain the successful track record of inference to the best explanation in science is to claim that it is truth-conducive, assuming that it is part of the theory-dependence of science that it argues for the truth of theories in part, at least, by evaluating their explanatory power.

#### Inference to the Best Explanation in Ethics

Explanation has played a large part in the recent revival of moral realism. There are many different versions of moral realism currently in circulation (for summary accounts of the diversity of views, and for references to major exponents of those views, see Sayre-McCord 1988a; Railton 1986, 163-165; and Arrington 1989, ch. 4). What they have in common is the claim that there are moral facts in an objective sense.<sup>6</sup> There is much disagreement about the nature of those moral facts. On some accounts, they are ontologically independent from non-moral facts, on some they supervene on non-moral

---

<sup>6</sup> Sayre-McCord 1988a gives a definition of realism that is more semantic than ontological. He claims that realism about a domain of things implies minimally that there are meaningful propositions about things in that domain and that some meaningful propositions about things in that domain are literally true (p. 5). This definition seems too weak, however, since there are alternatives to the correspondence notion of truth such that Sayre-McCord's conditions could be satisfied without satisfying the ontological claim that there are objective things or facts existing independently of any conceiver.

facts, and on some they are reducible to non-moral facts. There is also disagreement over whether we can *know* moral facts, and if so, how we come to know them.

A moral realist is committed to cognitivism, the view that moral judgments are capable of being either true or false (as opposed, for example, to the view called emotivism that moral statements are cognitively meaningless, that they lack truth-values; emotivism regarding ethics was supported by some Logical Positivists). It is the objective moral facts in virtue of which moral judgments are either true or false. Cognitivists are divided, however, into those who give an error theory and those who give a success theory of ethics (see Sayre-McCord 1988a, 9-10). An error theory of ethics admits that moral judgments are capable of being either true or false, but says that they are all in fact false because there is no moral reality that they correctly describe. An anti-realist can be a cognitivist by advancing an error theory. The realist must advance a success theory: not only are moral judgments capable of being either true or false, but there are also in fact some moral judgments (or better, propositions, for they need not actually be believed or have been believed by anyone) that are literally true. That is, there are moral facts which will make some moral propositions true, i.e., the propositions correctly describing the moral facts.

Gilbert Harman (1977) has prompted a debate about whether (objectivist) moral realism is necessary in order to explain our common moral intuitions and judgments. J. L. Mackie has also used considerations of adequate explanation in order to defend his error theory, “[T]he argument from relativity [against moral realism] has some force simply because the actual variations in the moral codes are more readily explained by the

hypothesis that they reflect ways of life than by the hypothesis that they express perceptions, most of them seriously inadequate and badly distorted, of objective values” (Mackie 1977, 37). Harman attacks moral realism by noting that “you do not seem to need to make assumptions about any moral facts to explain the occurrence of ... so-called moral observations.... In the moral case, it would seem that you need only make assumptions about the psychology or moral sensibility of the person making the moral observation. In the scientific case, theory is tested against the world” (1977, 121). Many moral realists have tried to respond to Harman’s argument. Nicholas Sturgeon replies directly, “My own view is that Harman is quite wrong, not in thinking that the explanatory role of our beliefs is important to their justification, but in thinking that moral beliefs play no such role.... I shall argue ... that moral beliefs commonly play the explanatory role Harman denies them” (1985, 233-234; see also David Zimmerman 1985 and Sayre-McCord 1988b). There is general agreement now that moral theories require at least explanatory potency to be acceptable.

### Inference to the Best Explanation in Recent Epistemology

Some recent epistemologists have used arguments from explanatory considerations as a source of epistemic justification for belief in the physical objects of ordinary perception and even for belief in the unobservable entities described by science. Realists of various sorts are rallying behind inference to the best explanation as their best chance at rebutting the skeptic with justification for ontological claims.

Alan Goldman uses a form of inference to the best explanation at every epistemic level. “I shall provide an argument ... that the use of the best-explanation principle tends to be truth preserving, even when inferences to best explanations cross epistemic levels from belief to experience, from experience to its objects, and from observed objects to the unobserved or unobservable” (1988, 11). For those interested in the role of explanatory criteria in ontological reasoning, the inferences from experience to its objects and from observed objects to unobserved objects will be most significant, of course. Goldman says elsewhere, “The defense of empirical foundations, the validation of beliefs about appearances, depend[s] ... on an inference to the best explanation. In my view that inference provides the *only* satisfactory answer to skeptical challenges, challenges intelligible even at the foundational level” (1988, 203).

Peter Carruthers claims “It is clear ... that the hypothesis of reliable perception [of a physical world] provides overwhelmingly the best available explanation of the course of our experience, and of the existence of our perceptual beliefs” (1992, 170). He goes on to provide an argument that inference to the best explanation is itself a reliable method of belief-formation, and thus capable of justifying belief in the physical world as the best explanation of our experience. Put roughly, his argument is that the concept of best explanation is innate, and is the product of an evolutionary process that selected it because it is a guide to the truth. That is, the best explanation of why we have an innate idea of best explanation is that it has survival value, and the best account of why it has survival value is that it leads us in our inferences to generally true beliefs (see Carruthers 1992, ch. 12).

William Lycan defends a position he calls explanationism. “I am a rabid explanationist, eager to defend the claim that *every* justified form of ampliative reasoning is a special case of explanatory inference” (1988, 178). Lycan continues the line of argument Gilbert Harman (1965) had started. Lycan claims specifically that “explanatory inference is fundamental to philosophical thought and that deduction and ordinary induction ... are either abbreviated forms of explanatory inference or merely tools used in the tightening of explanatory coherence” (*ibid.*, 112). This is what I was trying to get at in a more descriptive, less normative way in my treatments of Plato, Aristotle, and Leibniz. Lycan goes further and claims that explanatory inference is fundamental not only to philosophical thought but to every sort of thinking. “*We are always and everywhere stuck in the business of making comparisons of plausibility, and such comparisons are made only by weighing explanatory virtues*” (*ibid.*, 125, author’s emphasis). This statement is meant to apply to ampliative inference only, and if it is true, it entails that all ontological reasoning is necessarily based on explanatory inference.

Paul Moser (1985, 1989) is an explanationist foundationalist, and he endorses a form of inference to the best explanation to justify observation beliefs (about external objects) on the basis of immediately justified beliefs about phenomenological contents. In *Empirical Justification*, he sets up a general epistemological problem for all realists of any sort, so it is worth quoting him at length.

On my intuitionist account of immediate justification, only given-beliefs are immediately justified. These are beliefs about certain phenomenological content such as ostensibly presented perceptual objects, properties, and psychological states, including belief, perceptual and sensation states. The truth of such a belief does not entail the existence of any external perceptual object or property. The truth of an observation belief, however, does entail the existence of some external

perceptual object or property. And on my account such a belief is inferentially justified at best; that is, it must be justified on the basis of certain immediately justified given-beliefs. Obviously, if I am to solve the regress problem concerning particular observation beliefs, I must show how those beliefs can be epistemically justified on the basis of certain immediately justified given-beliefs. In short, I must find some way to justify inferences from foundational beliefs to certain nonfoundational observation beliefs. Otherwise, no observation beliefs will be epistemically justified, and some form of epistemological skepticism will be reasonable. In that case, the epistemic regress problem will provide a powerful argument not for foundationalism, but for epistemological skepticism of some form. (1985, 187-188)

Moser then examines three prominent methods of justifying belief in external objects on the basis of beliefs about phenomenological content. He finds Phenomenalist Deductivism and Chisholm's Critical Cognitivism flawed, and ends up endorsing an inductive inference to the best explanation of phenomenological content as a successful way of justifying belief in external objects and thus rebutting the skeptic (1985, 198-205). He also provides an interesting reply to skeptical challenges of inductive inference to the best explanation. Such a skeptical move appears to undermine itself, for it too makes appeals to consideration of best explanation. "[I]t is doubtful that any relevant kind of epistemological skepticism can itself be justified, or shown to be reasonable, if we reject inference to the best explanation. For the relevant skeptical objections derive their force from the alleged fact that they support a skeptical rather than a nonskeptical *explanatory account* of the relevant epistemological data" (p. 203).

In *Knowledge and Evidence* Moser makes much the same use of inference to the best explanation as the most basic element in the justification of beliefs about external objects (see especially Moser 1989, 88-107). In a section describing how propositions, including existential propositions, can be made probable to some extent by the subjective

contents of experience, Moser makes use of an explanatory criterion for conferring probability. This is not a point of merely some interest and minor significance, rather it is the keystone of his entire epistemological system. Without a link between personal, subjective experience and objective existence, knowledge of the external world is impossible. Without a link between nonconceptual contents of experience and propositions, knowledge of any sort is impossible; for knowledge requires believing, believing requires an object of believing, and “an object of one’s believing is a propositional truth-bearer” (Moser 1989, 13). Explanation provides these links. “One’s subjective nonconceptual contents can make a proposition, *P*, evidentially probable to some extent for one in virtue of those contents’ *being explained for one* by *P* in the sense that *P* is an essential part of an explanation for one of why those contents exist, or equivalently, why those contents occur as they do.... Propositions can be essential parts of an explanation for one, since they can provide one with answers to explanation-seeking why-questions” (1989, 91-92). Moser refines this preliminary statement of what he calls unconditional probability-makers, but the essential function of explanatory considerations remains the same (see 106 for a refined statement).

#### Inference to the Best Explanation in the Philosophy of Mind

I will end this very brief survey of recent ontology with a quick look at the philosophy of mind and cognitive psychology, for many arguments in this area regarding the thesis of physicalism have been based on considerations of how adequately to explain our experiences. These arguments involve a unique kind of difficulty, for there is often an

ambiguity about what experiences one is trying to explain and about what type of explanation one seeks. One can invoke considerations of how adequately to explain the experience an experimenter has of the behavior of experimental subjects in an experiment in cognitive psychology. Clearly, this sort of explanation falls under the heading of scientific explanation generally. One could postulate some mental entity as the best explanation of experimental results in a manner analogous to the postulation of theoretical entities in, say, quantum mechanics to explain experimental results. One could likewise argue against the postulation of certain mental entities (or the entire class of things called mental entities) on the basis that they are explanatorily superfluous just as one would argue against the postulation of an ether as an explanation of how light is propagated.

Paul Churchland is a prominent representative of this first type of approach. He considers the philosophy of mind to be heavily influenced by the philosophy of science; in a sense he replaces at least much of the philosophy of mind with the philosophy of the science(s) of the mind/brain. As explanation is an important topic in the philosophy of science, so it becomes a topic of importance in the philosophy of mind.

The nature of explanation in general and of psychological explanations in particular became an issue central to philosophers of mind because the conceptual framework with which we conceive of ourselves (as creatures with beliefs, desires, and other mental states) came to be seen as a common-sense explanatory *theory*—‘folk psychology’—a theory that enables each of us to predict and explain the behavior of the other humans that surround us. This made an issue in turn of the status of theoretical entities (Are mental states real?), the nature of cross-theoretical identities (Are mental states identical with brain states?), and the criteria for intertheoretic reduction (Could folk psychology itself eventually be explained in turn by a deeper and more neurobiological theory of human nature?)” (Churchland 1989, xi).



Churchland develops a parallel structure between the predictive and explanatory role of the theories of the indisputably physical sciences such as physics and chemistry, and the predictive and explanatory role of theories of the mind. He eventually rejects folk psychology as a theory of the mind and develops his own. In each case, however, in the philosophy of science and in the philosophy of mind, we are concerned with the theoretical explanation of a certain type of experience. In the former case it is experience of the natural world; in the latter it is experience of, as Churchland says, “the behavior of the other humans that surround us.” It is crucial to notice that this is experience that has already been explained to a considerable extent with the ontological thesis that it is experience *of* some objective entity such as a cloud chamber, a sample of copper, or a biological organism of the species *homo sapiens*.

On the other hand, philosophers of mind can try to explain “the subjective character of experience” from the unique “point of view” connected with the experience. Thomas Nagel (1974) uses these phrases and is concerned with explaining conscious experience (consciousness) in a sense different from that of explaining a limited set of observations of subjects in an experimental setting. We could say that Nagel is concerned with explanation of experience at a different level from that of the cognitive scientist. He argues against physicalism, not on the basis that it fails to account for the results of experiments with human subjects (i.e., experience of the behavior of others), but on the basis that it does nothing more than explain *the behavior of others* (in one sense of ‘explain’). Specifically, it does not explain consciousness, i.e., *my own* consciousness as I experience it.

If physicalism is to be defended, the phenomenological features must themselves be given a physical account. But when we examine their subjective character it seems that such a result is impossible. The reason is that every subjective phenomenon is essentially connected with a single point of view, and it seems inevitable that an objective, physical theory will abandon that point of view (Nagel 1974, 167).

It is not clear what sort of explanation Nagel has in mind here. In fact, he writes, “We have at present no conception of what an explanation of the physical nature of a mental phenomenon would be” (p. 166). It is clear, however, that he will not be satisfied with a neurophysiological account of what sorts of brain-events correspond to reports of phenomenological contents (and other relevant behavior), even though such an account would include Thomas Nagel himself. An account of what physical brain-states or brain-events are associated in a law-like manner with *his* subjective conscious experiences cannot explain, for Nagel, the subjective character of *his* experience.

Nagel and philosophers of a similar mind are in the minority these days. Since the time behaviorism rose to prominence in the 1930s, philosophers of mind have overwhelmingly preferred a scientific, experimental approach to a primarily introspective approach. One task of the philosophy of mind has been to explain the behavior of intelligent beings, beings that are conscious. The sense of ‘explain’ has generally been scientific. Even after behaviorism fell to widespread criticisms (especially as a semantic thesis), the conviction that the goal of the philosophy of mind was to offer a theoretical explanation of conscious behavior (and hence of consciousness) remained.

### The Explanation of Experience *via* Existence Claims: A Broad Sketch

I have tried so far to exhibit metaphysical and ontological theses as being primarily explanatory in the sense that they explain, or make understandable, the subjective contents of experience. This form of explanation is really quite simple. I can explain why the figure of a person I saw just seconds ago disappeared into thin air with the ontological claim that my experience was *not* an experience of (caused by) a real person; it was an illusion. I can explain why I experience the desk lamp exactly as I experienced it hours ago when I was last in the room with the ontological claim that I experience a desk lamp that is a physical object having an independent existence that is sustained continuously while I am not having any experience of it. I can explain why I seemed to be floating a few feet off the ground a few short hours ago with the ontological claim that I was actually lying in bed sleeping and the experience of floating was caused not by actual floating but the distinctive brain activity of dreaming. That is, I explain my experience as being caused not through the normal mechanisms of perception but by similar activity in my brain, itself a physical object, without the relevant type of bodily activity.

Having explained one's experience with a general ontology and an account of what brought the experience about in that ontology, one can embellish one's inventory of the furniture of the world by offering explanations of particular experienced objects or events with other *postulated* objects or events as causes. One begins with an ontological story about what brought about a certain experience, then one proceeds to give causal accounts within the relevant ontological framework of particular objects of

experience. We obviously need to clarify our use of ‘explanation’ to clarify the relationships between explanation and existence claims in contemporary philosophy. Instead of examining individually each of the uses of explanation in the ontological arguments mentioned above, I will first examine contemporary analyses of explanation, analyses that started with and developed from the classic “Studies in the Logic of Explanation” by Carl Hempel and Paul Oppenheim (1948). We will try to see whether there is a concept of explanation adequate to the tasks of contemporary ontology, and whether there are any vicious justificatory circles forming between explanatory justifications of existence claims and justifications of analyses of ‘explanation.’ Before turning to that task, however, I would like to undertake a more general analysis of metaphysical and ontological reasoning in order to answer the question of whether substantive assumptions about adequate explanation *necessarily* function in such reasoning. In other words: Is it possible to argue for a metaphysical or ontological claim without presupposing specific criteria of explanatory goodness?

### Explanation in Arguments Regarding Existence

I believe it is appropriate to focus our attention on *inferences* to metaphysical or ontological conclusions. Even though people probably form existential beliefs immediately on the basis of having appropriate experiences, it is always possible to form an intermediate proposition describing (categorizing) the experience without making any existential assertions. Further, it seems unlikely that one can justify an existential claim directly on the basis of experience, if that experience is not first categorized. How can

one justify the claim "There is a computer before me" with an experience unless that experience has first been described as "computer-ish"? One might claim that the existential proposition also does the work of the descriptive proposition, so that one can justify it immediately on the basis of having experiences that lend themselves to the relevant descriptions. If this is the case, however, then we can at least isolate the descriptive from the existential proposition in principle in every case. So I can recommend isolating them simply for the sake of clarity. We should distinguish the application of concepts to experience, on the one hand, from the inference from conceptualized experience to existential claims, on the other.

Another reason for focussing on inferences is that we are interested primarily in how people defend, or justify, the claims they make about reality. Plato might claim to believe in the Forms simply because he intuits them through a rational faculty. G. E. Moore might claim to believe that something is good simply because he intuits through a nonsensory faculty the nonnatural property of goodness in that thing. These claims do not carry much weight philosophically, however, for the faculties of intuition apparently are not shared by many people. Believing that a computer is before one simply on the basis of a visual intuition of it carries more weight philosophically only because a large majority (although not all) of us have this faculty of intuition. Still, in such cases it can be useful for a philosopher to imagine how one *would* defend even the ordinary perceptual beliefs if the common sensory experiences *were not* widely shared. Justificatory problems plague metaphysical arguments based on starting-points that are claimed to be self-evident.

An existential proposition can be supported, or justified, by other propositions either deductively or inductively. If an existential proposition is supported deductively by another proposition, then that supporting proposition will contain an existential element that needs justification in its turn. Since deduction is nonampliative, the supporting proposition will have to “contain” somehow the existential proposition we are interested in defending. More specifically, we cannot deductively infer an existential claim from a nonexistential claim. If we look only at deductive arguments, we will not escape from the sorts of existential claims we are interested in defending to reach something more basic (for a similar line of argument, see for example Lycan 1988, 115-122).

An existential proposition can be justified on the basis of propositions that are epistemically more basic only through an ampliative, or inductive, inference. If I can show that all ampliative inference (henceforth, *inductive inference*) harbors assumptions about the conditions of adequate explanation, I will have shown that all plausible metaphysical or ontological arguments rely on assumptions about the conditions for adequate explanation. The following is an argument, inspired by Harman (1965, 90-91),<sup>7</sup> for the thesis that all inductive inference relies on considerations of best explanation.

- (1) Let  $P$  be the conjunction of all premises in an inductive argument,  $A$ ; let  $C$  be the conclusion to  $A$ .

---

<sup>7</sup> Harman argues merely that “if ... we think of an enumerative induction as an inference to the best explanation, we can explain when a person is and when he is not warranted in making the inference from ‘All observed  $A$ ’s are  $B$ ’s’ to ‘All  $A$ ’s are  $B$ ’s.’” I argue that all inductive inferences, whether warranted or not, rely on substantive presuppositions about the best explanation of their premises.

(2)  $A$  is warranted only if there are no known defeaters for  $A$  (i.e., falsifiers of  $C$ , propositions warranting a falsifier of  $C$ , or premises bringing the probability of  $C$  below  $a$ , where warrant requires probability of  $a$  or greater).

(3)  $A$  thus *presupposes* the negation of all potential defeaters.

(4) If  $E$  is a proposition which would provide a better explanation of  $P$  than either (i)  $C$  or (ii) an explanation that entails  $C$  or (iii) an explanation that makes  $C$  more probable than each of its probabilistic competitors, then  $E$  is a potential defeater of  $A$ .

(5) Therefore,  $A$  presupposes that  $E$  is false.

(6) Therefore,  $A$  presupposes either (i) that  $C$  is the best explanation of  $P$  or (ii) that the best explanation of  $P$  entails  $C$  or (iii) that the best explanation of  $P$  makes  $C$  more probable than each of  $C$ 's probabilistic competitors.

(7) Therefore, any inductive inference presupposes either (i) that the hypothesis presented in its conclusion is the best explanation of the information presented in its premises or (ii) that the best explanation of the information presented in its premises entails its conclusion or (iii) that the best explanation of the information presented in its premises makes its conclusion more probable than each of its probabilistic competitors.

(2) is true analytically if we set  $a$  as the minimum probability conferred by  $P$  on  $C$  required for inductive warrant. We allow for disagreements about what  $a$  should be. For any theory of inductive warrant, however, any proposition which, if added to the

argument as a premise, would reduce the probability of *C* below the value that theory of inductive warrant assigns for *a*, is a defeater of the argument *A*. This premise is based simply on the fact that the addition of new premises can undermine an inductive argument.

(3) uses the term “presuppose” in the sense of depending on the *truth* of a proposition without explicitly asserting it. “To presuppose something is to take it for granted in a way that contrasts with asserting it” (Soames 1989, 553). Scott Soames discusses three different accounts of what it is for one proposition to presuppose another. I am here claiming that *an argument* presupposes a set of propositions, so my use of the term does not fit exactly into any of Soames’ three accounts (nor can my account be *reduced* to any of Soames’ accounts). My use of the term does, however, bear some resemblance to those that Soames discusses. Soames says that the truth of a presupposed proposition is a necessary condition either (a) for the proposition that presupposes it to be either true or false or (b) for the sentence that presupposes it to be meaningful (Soames 1989, 556 and 562). I claim that for an argument to presuppose a set of propositions is for it to presuppose each proposition in that set individually, and for an argument to presuppose a proposition is for the deductive validity or inductive goodness of that argument to *depend on* the truth of that proposition. This is distinct from and cannot be reduced to the presuppositions necessary for the constituent parts of the argument to be (a) either true or false propositions or (b) meaningful sentences.

The dependence that an argument has on its presuppositions can be of two sorts. Deductive arguments depend on their presuppositions as a *necessary* condition for their



validity; this sort of presupposition, therefore, resembles rather closely the presupposition Soames discusses. For example, the argument-form from  $(x)(Fx)$  to  $(\exists x)(Fx)$  presupposes that at least one thing exists. If it is false that at least one thing exists, then the premise of every instance of this argument-form will be true and the conclusion will be false, rendering the argument-form invalid. In other words, this argument-form is valid only in possible worlds where at least one thing exists. Another example comes from Traditional logic. Traditional logicians (following Aristotle) apparently accepted the inference from "All  $S$  is  $P$ " to its subaltern "Some  $S$  is  $P$ ." If one reads this inference as  $(x)(Sx \supset Px) \ \& \ (\exists x)(Sx), \ \therefore \ (\exists x)(Sx \ \& \ Px)$ , then this inference does *not presuppose* that something is  $S$ , for it asserts it in the premise. If, however, one reads the inference as  $(x)(Sx \supset Px), \ \therefore \ (\exists x)(Sx \ \& \ Px)$ , then this inference presupposes that something is  $S$  in exactly the same way that  $(x)(Fx), \ \therefore \ (\exists x)(Fx)$  presupposes that something exists; if something is  $S$ , then the inference is valid, and if nothing is  $S$ , then the inference is invalid.<sup>8</sup>

Inductive arguments do not depend on their presuppositions in this way; the truth of their presuppositions is not a necessary condition for their inductive goodness. An inductive argument can be a good argument even if one or more of its presuppositions is false, because an inductive argument can be a good argument even if its premises are all true and its conclusion is false. But an inductive argument presupposes certain propositions in virtue of asserting that its premises *give some support* to its conclusion.

---

<sup>8</sup> For more on the existential assumptions of traditional deductive logic, see Mulder 1996.

An inductive argument, whether actually warranted or not, presents itself in some sense as being warranted. That is, an inductive argument asserts, or makes the claim, that its premises give at least a certain designated degree of support (the degree demanded by inductive warrant) to its conclusion. Consider the inductive argument from “All observed swans are white” to “All swans are white.” This argument, in virtue of asserting that the lack of observed non-white swans gives at least some designated degree of support to the conclusion that all swans are white, presupposes that  $q$ , where  $q$  = “If some swans are not white, then probably some observed swans are not white.” We must remember, however, that presupposing *contrasts* with asserting. The argument does not assert anything about what is *probably* the case; it asserts simply that since all observed swans are white, all swans are white. This assertion presupposes  $q$ ; it presupposes that the world is such that, if there were non-white swans, some of them would probably show up in our sample. The presupposition would be false if the world were not actually this way. For example, if the mechanism for selecting swans for our observational sample systematically excluded non-white swans, our presupposition would be false and the argument would not be inductively warranted. The claim that the inference is inductively warranted is contingent on the world’s being such that there is no systematic exclusion of non-white swans from our sample; that is, it is contingent on the truth of the presupposition  $q$ .<sup>9</sup>

---

<sup>9</sup> I am using an externalist account of inductive warrant; I am describing what it is for an inductive argument (not a person) objectively to bear a certain relationship called “presupposing” to a proposition. If I were giving an internalist account of inductive warrant for a person, I would have to talk not only of the way the world is but also a subject’s being aware of the relevant features of the world.

When (3) says that the inductive argument  $A$  presupposes the negation of all potential defeaters, it means that  $A$  presupposes that it is not the case that there is anything about the world that would render the inference unwarranted. A potential defeater is any proposition which, if true, would when added as a premise to the argument render the probability of the conclusion less than  $a$ . The swan argument presupposes the *negation* of “if a swan is not white, then it is not observed,” because this proposition is a potential defeater of the argument.

(4) says that if there is a better potential explanation of  $P$  than the explanation supporting  $C$ , then it is a potential defeater of  $A$ .

Since the conclusion of an inductive argument might not itself be an explanation at all, we need to account for various ways the conclusion can relate to potential explanations of the information presented in the premises. First, the conclusion might itself be the best explanation of the premises; the conclusion that all swans are white might be (contrary to fact) the best explanation of why all observed swans are white. Second, the conclusion might be entailed by the best explanation. There may be (contrary to fact) a genetic account of why all swans must be white. Third, the conclusion might be made more probable than any of its probabilistic competitors by the best explanation. There may be (contrary to fact) a genetic account of why it is very unlikely that there be any non-white swans.

Let's call explanation  $E$  of (4) an “alternative explanation.” An alternative explanation neither is  $C$ , entails  $C$ , nor makes  $C$  more probable than each of its probabilistic competitors.  $E$ , if true, would be a defeater of  $A$ . This is because all

alternative explanations are potential defeaters of  $A$ . Consider the following alternative explanation of  $P$ : All swans in the sample were purposely selected for whiteness. It is *because* this explanation is *alternative* to  $C$  that it is a defeater of  $A$ ; if it is true, it undermines the inductive warrant for the inference from  $P$  to  $C$ . By definition, an alternative explanation does not make  $C$  more probable than each of its probabilistic competitors; that is, it makes at least one probabilistic competitor to  $C$  (call it  $C'$ ) more probable than or equally probable to  $C$ . Hence, either  $C'$  would be the inductively warranted conclusion or else neither  $C'$  nor  $C$  would be warranted, if they are equally probable.

(5) follows from (3) and (4).  $A$  presupposes the negation of all potential defeaters, and  $E$  is a potential defeater; therefore,  $A$  presupposes the negation of  $E$ .

(6) follows as a generalization of (5). If (5) is true for all alternative explanations (and it is because we put no restrictions on  $E$  when we introduced it as an alternative explanation), then  $A$  presupposes that there is *no* better alternative explanation. And this simply amounts to the claim that the best explanation either is  $C$ , entails  $C$ , or makes  $C$  more probable than any probabilistic competitor.

(7) is simply a generalization of (6), for we put no restrictions on  $A$  when we introduced it. It reaches the conclusion that any inductive argument presupposes that the best explanation of its premises supports its conclusion. It can support the conclusion either by being the conclusion itself, entailing the conclusion, or making the conclusion more probable than any of its probabilistic competitors.

Taking these three possible relations together, saying that an inductive argument presupposes one of them is saying *more* than just that the argument presupposes that its conclusion is *consistent* with the best explanation of the premises. The conclusion can be consistent with the best explanation without itself being, or being entailed by, or being made optimally probable by the best explanation. An inductive argument presupposes something stronger than mere consistency between its conclusion and the best explanation of its premises. If all the swans of our observational sample were purposely selected for whiteness, then this selection process would be the best explanation of why all observed swans are white. The conclusion that all swans are white is consistent with this explanation, but this explanation, if correct, does not *support* the conclusion that all swans are white; on the contrary, it *undermines* that conclusion and hence defeats the inductive argument to that conclusion.

As I mentioned earlier in this chapter, I believe that my thesis is independent of both Harman's thesis that inference to the best explanation is the basic form of nondeductive inference and Fumerton's thesis that all inferences to best explanations reduce to straightforward inductions or combinations of ordinary induction and deduction. My thesis is that all nondeductive inferences *presuppose* certain substantive, nonanalytic standards for explanation and thus presuppose some more or less detailed concept of explanation. This presupposition of nondeductive inference is just one part, a part having special significance here, of the broader presuppositions about *how the world is* that every nondeductive inference must make. Different nondeductive inferences can betray, by their very structure, various presuppositions about how the world is. Such arguments

betray their worldly presuppositions because under some assumptions about how the world is, their premises *will not* give anything close to adequate support for their conclusions, while under other assumptions they will give *strong* support to their conclusions.

The upshot of all this is that every ampliative (nondeductive, or inductive) argument *depends on*, in a clear and specific sense, a set of criteria for explanatory adequacy and goodness, and it is for this reason that any inductive inference can be reconstructed as an explicit inference to an explanation. It is not *merely* that every justified form of ampliative reasoning is a case of explanatory inference, as Lycan says. Nor is it merely the case that introducing explanatory considerations in a reconstruction of an inductive inference as an inference to the best explanation allows us to distinguish warranted from unwarranted inductive inferences, as Harman says. My thesis does not depend at all on any particular claims about justification or warrant for inductive inferences. My thesis is that every ampliative inference, whether justified or not, depends on certain assumptions about explanatory adequacy and goodness, through a mechanism I have called “presupposition.”

This chapter has examined the prominent role of clear and explicit references to explanation in contemporary ontological reasoning. It has also presented an argument that all nondeductive inference, the basis of all ontological reasoning, presupposes some specific criteria of explanation. Consequently, all ontological reasoning must be based on some more or less specific *concept* of explanation, and it typically invokes that concept clearly in justifying ontological claims as the best explanation of some set of

experiences or observations. This has been a continuation of the work of the previous chapter, which examined the reliance on standards of explanation in traditional metaphysical reasoning.

The time has come to examine recent philosophical attempts to analyse concepts of explanation. My concern will be to find out what role, if any, there is for prior ontological or metaphysical commitments in such analyses. This will be the second major topic of this dissertation: the possible influence commitments to existence claims have on our understanding of what explanation is and how to evaluate explanations.

## CHAPTER 5

### HEMPEL'S ANALYSES OF EXPLANATION

Modern analyses of explanation began with the publication, in 1948, of "Studies in the Logic of Explanation" by Carl G. Hempel and Paul Oppenheim. Philosophers examined notions of explanation before this; the philosophical study of explanation goes back at least as far as Aristotle, and even Plato clearly had some interesting things to say about how properly to formulate explanations. More recently, John Stuart Mill deals with explanation in *A System of Logic* (1858); he anticipated much of Hempel's model of deductive explanation. I will, however, agree with J. Alberto Coffa, "Explanation theory abandoned its pre-theoretical stage and became a respectable branch of philosophical inquiry when, in the late forties, Hempel began to develop his model of deductive-nomological (D-N) explanation" (1974, 56). Virtually all philosophers working in this area concur; Wesley Salmon writes, "The 1948 Hempel-Oppenheim article marks the division between the prehistory and the history of modern discussions of scientific explanation" (1989, 10; see also his 1984, 21).<sup>1</sup>

---

<sup>1</sup> Salmon 1989 provides a thorough survey of discussions of scientific explanation from 1948 to the late-1980s. It also includes a chronological bibliography of virtually all philosophical work on scientific explanation, and much on other types of explanation, up to 1989.



### Deductive Explanation

In the 1948 “Studies” article, Hempel and Oppenheim describe what is known as the deductive-nomological (D-N) model of the scientific explanation of particular events.<sup>2</sup> They do not deal with nondeductive (statistical) explanations (see p. 251) or with explanations of general regularities (see footnote 33), although Hempel takes up these topics in later publications (his fullest account of scientific explanation is his 1965). A D-N explanation of a particular phenomenon is a valid deductive argument having a description of that phenomenon as its conclusion. The premises of the argument contain at least one general law or theory that is essential to the argument, and the premises may also contain statements of specific antecedent conditions. Thus one can explain, for example, why a sample of copper conducts electricity by deducing the statement that it does from the antecedent conditions (that it is connected to an electrical source under appropriate conditions, etc.) and the general law that all copper connected to an electrical source (under certain conditions, etc.) will conduct electricity.

In an argument that constitutes an explanation, the premises make up the *explanans* and the conclusion is the *explanandum* (in general, the explanandum is what gets explained, the explanans is what does the explaining). A general law that appears in an explanation is sometimes called a covering law; hence the D-N model can be called *a covering-law model of explanation*. We should refrain from calling the D-N model of

---

<sup>2</sup> I should say it *came* to be known as deductive-nomological explanation, or the deductive-nomological model of explanation, for Hempel and Oppenheim never use the term ‘deductive-nomological’ in the 1948 article, nor do they refer to a “model” of scientific explanation.

explanation *the* covering-law model, however, because there are other models that can also be called covering-law models. There is a version of the D-N model less detailed than Hempel-Oppenheim (1948) in Popper (1935 [1959] sect. 12, 1945 ch. 25), in Braithwaite (1953), and in Nagel (1961).

Hempel and Oppenheim distinguish a *potential* explanation (having a potential explanans) from an actual explanation, or simply an explanation. An argument satisfying the conditions set out above is a potential explanation. A potential explanation that has true premises is an actual explanation, or just an explanation (see p. 248, condition R4). More formally, the authors claim, for a model language *L*, that the statements of the antecedent conditions must be true and that the general statements must be theories in an actual explanation (1948, 273 (7.6)), but they stipulate that a general statement must be *true* in order to be a theory (“...every theory is true.” p. 272; I believe this stipulation goes against ordinary usage).

According to this account, an explanation of a particular phenomenon and a prediction of a particular phenomenon have the same logical form. If the particular phenomenon mentioned in the conclusion has already occurred, a deductive argument fulfilling the above conditions is an explanation of it. If the phenomenon has not yet occurred, the same argument supports a prediction of it. The authors use such a capacity to support predictions (“predictive import”) as a litmus test, so to speak, for adequate explanations. They criticize other conceptions of explanation on the basis that they allow for explanations that have no predictive import (pp. 249 and 255).

We encounter already in this last thesis a point of considerable controversy for commentators on Hempel's theories of explanation. I will deal with the controversies later in this chapter and in the next, after I have presented a more complete picture of Hempel's work on explanation. At present, I would like merely to raise the question that will be the focus at the end of this chapter and in later chapters: what is the source or basis of Hempel's and others' firmly held intuitions about what counts as an explanation and what does not? At this point, we must begin to wonder why Hempel feels so strongly that all explanations *must* have predictive import. Others will dispute this claim. Other disputed intuitions about explanation will also surface as we examine Hempel's work further.

Hempel and Oppenheim present a formal analysis of explanation for a model language  $L$ . I shall not go into the full details of this analysis here, but I should mention some novel points that come out of this analysis. They rely on the notion of a theory rather than that of a law in this formulation, because all laws are universal in form but some explanations contain no universal statements. A D-N explanation must contain at least one statement that is essentially generalized, however, and all theories are essentially generalized (fundamental theories are purely generalized as well), according to Hempel and Oppenheim. They claim initially that a potential explanans must contain at least one statement that is essentially generalized and may contain singular sentences. The explanandum sentence must be derivable from the conjunction of the essentially generalized sentences and the singular sentences but not from the singular sentences alone. An explanation is a potential explanation that has all true premises.

Hempel and Oppenheim find a problem with this initial formulation, however. It allows us to construct arguments that exhibit an unacceptable kind of self-explanation (one other type of partial self-explanation *cannot* be ruled out in a satisfactory way, they say; see pp. 274-276). We are left to wonder what their basis for distinguishing acceptable from unacceptable explanations is. More accurately, this basis for discrimination is left at an unanalysed intuitive level, and we will later focus on clarifying the basis for these sorts of discriminations. The following argument is an example of the self-explanation they find unacceptable (p. 276):

- (*T*) All metals are good conductors of heat.
- (*C*) If it is true that (If the Eiffel Tower is metal, then it is a good conductor of heat), then Mt. Everest is snow-capped.
- (*E*) Therefore, Mt. Everest is snow-capped.

This argument satisfies the conditions laid down so far. *E* is a logical consequence of *T* & *C* but not of *C* alone. *T* is essentially generalized, and *C* is singular. In more informal language, the explanans contains a general law (*T*) that is necessary for the deduction, it contains a statement of particular antecedent conditions (*C*), the argument is valid, and the argument is sound (the premises are true). This is nonetheless not a satisfactory explanation of why Mt. Everest is snow-capped, they claim. In order to rule out this sort of argument, Hempel and Oppenheim add an additional requirement

to the effect that “the assumption that  $T$  is true must not imply that verification of  $C$  necessitates verification of  $E$ ” (p. 277; see their final formulation 7.8 on pp. 277-8).<sup>3</sup>

One of the most difficult topics in a theory of explanation is the concept of a law. The central difficulty involving laws is the difficulty of finding a principled method for identifying lawlike sentences. A lawlike sentence is simply a sentence that has all the characteristics of a law, with the possible exception of truth. It is, however, quite difficult to find a principle for distinguishing lawlike sentences from other types of universal generalizations, even though it feels for the most part easy to recognize a lawlike sentence when we see one.

Consider some examples. “All gold is malleable” is a lawlike sentence, and it happens to be a law if adequately qualified. “All apples in this basket are red,” by contrast, is not a lawlike sentence. How does one distinguish the one from the other? They cannot be distinguished in virtue of their logical form, for they have the same logical form, universal form; while universal form is necessary for laws, it is not a sufficient condition.

Hempel and Oppenheim note that one feature distinguishing the two is that the lawlike sentence can (potentially) *explain* particular instances that fall under it while other

---

<sup>3</sup> Hempel and Oppenheim deal with a number of topics that I will pass over here. For example, they offer an analysis of teleological explanation in terms of explanation based on antecedent conditions and laws connecting the antecedent conditions and the phenomenon to be explained. They argue that their model applies not only to the physical sciences but also outside of them. They dispute the claim that there are phenomena that are “emergent” in the sense of being completely novel and hence unexplainable. Finally, they outline a method of measuring the systematic predictive or explanatory power of a theory.

universal generalizations cannot (1948, 266). We can explain why a particular sample of gold is malleable in part with the law that all gold is malleable, but we cannot explain why a particular apple is red (rather than green) by saying that it is in the basket and all apples in the basket are red. This is why it is essential for a theory of explanation to have an account of the nature of laws. We cannot use explanatoriness to help identify laws, however, for we are using the notion of a law to characterize explanation.

Hempel and Oppenheim consider two ways of distinguishing lawlike sentences from others (pp. 266-267). First, the sentence that is not lawlike has a limited scope; the lawlike sentence has unlimited scope. Yet there are laws that are limited. Kepler's laws of planetary motion, for example, are explicitly limited in application to the planets of our solar system. However, Kepler's laws are *derivative*; they are universal, but they follow from more comprehensive laws and ultimately from what are called *fundamental* laws (see p. 267). In this case, the law of universal gravitation is the fundamental law from which Kepler's laws ultimately derive (approximately). So Hempel and Oppenheim admit that there can be laws with limited scope, but they must all be derivable from fundamental laws. Fundamental laws (and fundamental lawlike sentences) must have unlimited scope.

The second condition for a lawlike sentence also applies only to fundamental lawlike sentences. A fundamental lawlike sentence may contain only purely qualitative predicates. A predicate is purely qualitative if and only if "a statement of its meaning does not require reference to any one particular object or spatio-temporal location" (1948, 268). The authors admit that there is a serious difficulty in characterizing

qualitative predicates precisely, but assume that the notion will be intuitively clear enough to readers given a brief list of qualitative predicates and non-qualitative predicates (pp. 268-9). From the condition of unlimited scope and the restriction to purely qualitative predicates together, it follows that a law may not contain designations of particular objects.

Hempel and Oppenheim say no more about the nature of laws in this article. They refer to a sentence satisfying both conditions above as *purely universal* (p. 272). Any sentence that is purely universal is a fundamental lawlike sentence; if it is also true, it is a law. Derivative lawlike sentences are derivable from fundamental lawlike sentences and are essentially, although not purely, universal.

The account of lawlike sentences given here is not wholly satisfactory; the difficulties involved in giving an account of the nature of laws are much deeper than the treatment in "Studies in the Logic of Explanation" would suggest. In particular, the conditions mentioned fail to exclude many statements that are not lawlike. For example, the statement 'No gold sphere has a mass greater than 100,000 kg' satisfies Hempel's conditions, but it is only *accidentally* true rather than being a law. The difficulty of characterizing lawlike sentences is made clearer by comparing this last statement, which is not lawlike, with 'No enriched uranium sphere has a mass greater than 100,000 kg,' which is not only lawlike but a law (Salmon 1989, 15). The notion of a purely qualitative predicate also came under special attack.<sup>4</sup>

---

<sup>4</sup> This difficulty was heightened by Goodman's Grue-Bleen problem in his 1955. Goodman claims, and most philosophers now agree, that qualitateness of predicates is a relative matter.

In the 1950s and early 1960s, Hempel worked on resolving many of the problems in the original paper. He also developed an account of scientific explanations that are not D-N explanations, explanations that rely essentially on laws that are statistical rather than universal (see his 1962). The culmination of his work is presented in “Aspects of Scientific Explanation” (1965). He begins with a more careful account of what it is he is trying to *model* (he introduces the term ‘model’ on p. 335).

It is important to identify clearly what one is trying to model, for it is impossible to evaluate a model unless one knows what it is supposed to be a model of. The task of designating what is to be modelled also forces the crucial question we shall address later of how one initially makes the discrimination of what counts as an instance of the thing we aim to model. Explanations, Hempel tells us, are always answers to questions that can be formulated as why-questions. There are, however, some why-questions whose answers are not explanations. Some why-questions (*reason-seeking, or epistemic, why-questions*) ask for evidence or epistemic justification of a belief and can be formulated as “Why should it be believed that *p*?” *Explanation-seeking why-questions* can always be formulated as “Why is it the case that *p*?” (1965, 334-5). Hempel claims that “any adequate answer to an explanation-seeking question ‘Why is it the case that *p*?’ must also provide a potential answer to the corresponding epistemic question ‘What grounds are there for believing that *p*?’” (p. 335). This is apparently another firmly held conviction about explanation that Hempel leaves at an intuitive level.

The account of D-N explanation is very much the same as in the 1948 article, but he does give special emphasis to a certain feature of this model (a feature shared by the



statistical models to come). According to Hempel, an explanation makes its explanandum understandable by making it something rationally to be expected, given the explanans (pp. 336-7). This is just a corollary of the thesis of the structural identity of explanation and prediction, or the claim that all explanations must have predictive import. If the information in the explanans could justify a prediction of the phenomenon described in the explanandum had it not yet occurred, then that information renders that phenomenon something rationally to be expected, at least to a high degree, given the explanans. It is important to notice, however, that this view commits Hempel to what has been called an *epistemic* interpretation of explanation. Explanations essentially present information on the basis of which one should have rationally expected, could have predicted, and can potentially justify belief in the occurrence of the explanandum phenomenon. It focuses more on sentences and potentially justificatory inferences than on events or facts.

This seems initially plausible. However, it remains to be seen whether rational expectation is sufficient for understanding and whether it is necessary for understanding, i.e., whether there are no ways to achieve understanding without the element of rational expectation. Again, we see Hempel here expressing strong intuitions about the conditions of acceptable explanation, and these intuitions are debatable and will in fact be contested by critics.

In "Aspects," Hempel introduces some minor refinements on the D-N model. One of the most significant additions to the 1948 article, however, is the more

sophisticated treatment of the nature of laws. Hempel begins with an admission of the difficulty of giving an adequate account of laws:

The characterization of laws as true lawlike sentences raises the important and intriguing problem of giving a clear characterization of lawlike sentences without, in turn, using the concept of law. This problem has proved to be highly recalcitrant, and I will make here only a few observations on certain aspects of it that are relevant also to the analysis of scientific explanation (1965, 338).

After reviewing the point brought out in “Studies” that lawlike sentences cannot be distinguished by their logical form alone, he adds a new consideration, one raised by Nelson Goodman in the intervening years. Goodman<sup>5</sup> had pointed out that laws can (inferentially) support counterfactual and subjunctive conditionals while nonlaws cannot. “Thus the law about the expansion of gases can serve to support statements such as ‘If the oxygen in this cylinder had been heated (were heated) under constant pressure then it would have expanded (would expand)’; whereas the [nonlawlike] statement [that all members of the Greenbury School Board for 1964 are bald] lends no support at all to the subjunctive conditional ‘If Robert Crocker were a member of the Greenbury School Board for 1964 then he would be bald’” (Hempel 1965, 339). Returning to an earlier example, we can see that ‘No gold sphere has a mass greater than 100,000 kg’ cannot support the subjunctive conditional that if two gold spheres greater than 50,000 kg each were to be brought together they would not form a gold sphere greater than 100,000 kg in mass; the law about the enriched uranium, on the other hand, *does* support the subjunctive conditional that if one brought together samples of enriched uranium having

---

<sup>5</sup> See the section entitled “The Problem of Law” in his 1955, any edition.

a combined mass over 100,000 kg one would *not* thereby form a sphere of enriched uranium greater than 100,000 kg in mass.

Hempel repeats that laws can explain particular instances they cover while other universal generalizations cannot, but again we cannot use the notion of explanatory power to characterize the notion of law, for we are using the notion of law to characterize the notion of explanation. In addition, the comments about counterfactual and subjunctive conditional statements “shed light on the concept of lawlikeness, [but] they afford no satisfactory explication of it” because such statements “present notorious philosophical difficulties” (p. 339). In particular, it is extremely difficult to give a characterization of what statements will support such conditionals without using the notion of a law, and we have seen that we need the notion of being able to support subjunctive and counterfactual conditionals in our characterization of laws. Hempel presents some further discussion intended to clarify the notions of lawlike sentence and law but concludes that he does not have “a fully satisfactory general characterization of lawlike sentences and thus of laws” (p. 343).

Hempel expands on the “Studies” account of D-N explanation also by dealing not only with the explanation of particular phenomena but also with the explanation of general uniformities expressed in laws. In the 1948 “Studies,” the authors considered only those explanations having a sentence describing a particular phenomenon as the explanandum, or conclusion. In “Aspects,” Hempel considers explanations of uniformities expressed in laws, that is, explanations that have a law as the explanandum sentence.

Aside from the nature of the explanandum, the structure of D-N explanations of uniformities remains mostly the same as that of D-N explanations of particular phenomena. One important difference is that, according to Hempel, one can have what he calls “approximative D-N explanation” of laws, that is, explanations in which the explanans does not strictly entail the explanandum. He gives the example of explaining Galileo’s law of falling bodies by way of Newton’s law of gravitation. Galileo’s law claims that the acceleration of a freely falling body near the earth is constant (ignoring air resistance), whereas Newton’s law of gravitation implies that the acceleration remains only approximately constant, for it increases slightly as the distance between the falling body and the center of the earth decreases. Since Galileo’s law is found to hold only to a high degree of approximation, Hempel says that Newton’s law provides for an approximative explanation of Galileo’s law.

We are left to wonder, however, what has happened to Hempel’s criteria concerning rational expectation and predictive import. I do not see how knowledge of Newton’s laws would lead one rationally to expect Galileo to have formulated his laws for falling bodies or how it would justify a prediction of someone formulating those laws. It seems to me that this is another example of Hempel’s relying on unanalysed intuitions about what is and is not an explanation, intuitions that tell him that scientists clearly *do* explain some laws in terms of “deeper” laws or theories.

According to Hempel, D-N explanations of laws, whether approximative or not, must derive the explanandum-law from *more comprehensive* laws that explain more than just the explanandum-law and thereby exhibit the uniformities expressed in the

explanandum-law as being special cases of deeper uniformities expressed in the explanans-law(s). In this way they generally deepen our understanding of the world. Deriving the explanandum law from more comprehensive law(s) appears to be a necessary condition of the explanation of uniformities. One cannot explain a law  $L$  simply by deducing it from the stronger law  $L \& L^*$  where  $L^*$  is an unrelated law (for example, one cannot explain Kepler's laws by deducing them from the conjunction of Kepler's laws and Boyle's law); one must deduce it from a higher-level, more comprehensive law in order to explain it. Hempel presents no account, however, either in "Studies" or in "Aspects," of how adequately to distinguish levels of comprehensiveness in this sense.

Explanations having the form D-N include, but are not restricted to, causal explanations. According to Hempel, any causal claim will imply some law to the effect that whenever the cause occurs (some event of a certain kind) then the effect will occur (some event of another kind). Such a law can be part of the explanans in a D-N explanation of the effect, together with the specific conditions of the particular cause involved. Thus, any causal explanation can be cast in the D-N form of explanation.

The D-N model includes non-causal explanations as well, however. One type consists of explanations that are based on so-called laws of coexistence (p. 352). An example (supplied by Hempel) of a law of coexistence is the law correlating the length and the period of a pendulum. With this law, one can construct a D-N explanation of the period of a pendulum based on its length. However, one would not say that the length of the pendulum *caused* its period, Hempel says.

It is interesting to note in connection with this last example that Hempel claims we might not be inclined to take the period of the pendulum to explain its length, even though we could deduce the length of a pendulum from its period and the law correlating them in a D-N fashion (p. 352). Many writers have subsequently given similar examples of arguments that fit the D-N model but fail to explain the purported explanandum (e.g., the flagpole and its shadow), claiming they provide counterexamples to Hempel's model.<sup>6</sup> Hempel apparently does not see them that way, for he writes, "the common-sense conception of explanation appears to provide no clear grounds on which to decide whether a given argument that deductively subsumes an occurrence under laws is to qualify as an explanation" (p. 353). This statement seems to indicate that Hempel admits the possibility of arguments fitting his D-N schema for explanation of particular events without being explanatory according to the common-sense conception, and does not regard these as counterexamples to his model. Further, we have already mentioned that Hempel admits that the deductive subsumption of a law under a stronger law is not necessarily an explanation.

---

<sup>6</sup> Wesley Salmon has recently repeated this assessment of the situation in his 1992. He writes "Much of the criticism [of Hempel and Oppenheim 1948] was brought into sharp focus by means of counterexamples that have, themselves, become classic. These examples fall into two broad categories. The first consists of arguments that fulfill all of the requirements for D-N explanation, yet patently fail to qualify as bona fide explanations. They show that the requirements set forth by Hempel and Oppenheim are not *sufficient* to determine what constitutes an acceptable scientific explanation. The second consists of examples of allegedly bona fide explanation that fail to fulfill the Hempel-Oppenheim requirements. They are meant to show that it is not *necessary* to fulfill those requirements in order to have correct explanations. We must treat this second category with care, for Hempel and Oppenheim never asserted that all correct explanations fit the D-N pattern" (Salmon 1992, 21).

The counterexamples are obviously effective against the universal claim that all arguments fitting the D-N schema and having only true premises are adequate explanations. It is unclear, however, whether Hempel's D-N model is intended to make this universal claim (this point, I believe, is acknowledged in none of Hempel's commentators). This raises the important question of what Hempel's models are intended to assert about explanation (to be examined in more detail later in this chapter). This question can be answered only by giving a full account of the method of philosophical analysis of a concept. If Hempel's models are meant to take ordinary concepts of explanation only as a point of departure and aim at articulating new, more detailed concepts of explanation, they have the character of *stipulations* of definitions for newly invented terms such as 'D-N explanation', 'D-S explanation', and 'I-S explanation'. As stipulations, they *cannot* be susceptible to counterexamples.

Hempel claims explicitly in "Studies" that his formal account of explanation for a model language  $L$  (given in (7.8) and (7.6)) is intended as a definition, giving necessary and sufficient conditions, for 'explanation' (p. 273 ff.). I do not know how to take this claim, however. Hempel admits in "Studies" as well as in "Aspects" that there are explanations that are not deductive; his "definition" of D-N explanation cannot be meant to give necessary conditions for an explanation, even a scientific explanation, for that would exclude all statistical explanations.

His mention of the pendulum explanation raises two possibilities concerning the claim to give sufficient conditions. First, he can say that one can construct a D-N explanation of the length of a pendulum from its period and the appropriate law, in spite

of the rather clear fact that this violates “the common-sense concept” of explanation. Second, he can say that one can construct an argument fitting his D-N model for the conclusion that the pendulum has a specific length that nonetheless fails to be an adequate explanation; his D-N model does not present sufficient conditions for a scientific explanation. Since Hempel explicitly raises examples of D-N arguments that are not common-sense explanations and explanations that are not D-N arguments, it is impossible to maintain that Hempel intended to give a definition capturing “the common-sense conception” of explanation. Counterexamples derived from a common-sense notion of explanation hence seem inappropriate, and it is hard to see what other basis for a counterexample to Hempel would be appropriate.

In “Aspects,” Hempel provides a more thorough treatment of the relationship between explanation and prediction than he did in “Studies.” This is a crucial topic, for it is, as already noted, intimately connected to his epistemic interpretation of explanation. The thesis of the structural identity of explanation and prediction can be broken down into two sub-theses: (i) “every adequate explanation is potentially a prediction”, and (ii) “every adequate prediction is potentially an explanation” (p. 367).

Hempel supports the first sub-thesis. We have looked only at D-N explanations so far, and it is clear that every adequate D-N explanation is potentially a prediction. Any D-N explanation would support a prediction of the event described in the explanandum sentence if it had not yet occurred, because the explanans sentences logically entail the explanandum sentence. Hempel goes beyond this, however, anticipating his treatment of non-deductive explanations. In some explanations, the



explanans sentences do not deductively entail the explanandum sentence, they merely inductively support it. Hempel claims that in these cases too, every adequate explanation is a potential prediction. He supports this claim with the “*general condition of adequacy for any rationally acceptable explanation of a particular event*”:

Any rationally acceptable answer to the question ‘Why did event *X* occur?’ must offer information which shows that *X* was to be expected.... Thus, the explanatory information must provide good grounds for believing that *X* did in fact occur; otherwise, that information would give us no adequate reason for saying: “That explains it—that does show why *X* occurred.” And an explanatory account that satisfies this condition constitutes, of course, a potential prediction in the sense that it could have served to predict the occurrence of *X* (deductively or with more or less high probability) if the information contained in the explanans had been available at a suitable earlier time (1965, 367-368).

Hempel holds that any adequate explanation must render its explanandum, if not deductively certain given the explanans, at least highly likely given the explanans. He says that the explanans of an adequate explanation must confer nomic expectability on its explanandum. This implies that all adequate explanations are potentially predictive, because nomic expectability is sufficient to support a prediction.

Hempel claims that while it is a necessary condition of an adequate explanation that it would, under appropriate conditions, provide a prediction of the explanandum phenomenon, predictive import is not sufficient for having an explanation. Some predictions cannot serve as explanations (1965, 368). Thus he rejects the second sub-thesis above. Hempel claims that what distinguishes predictions which can be explanatory from those that cannot is that the latter “involve no laws or theoretical principles, no explanatory statements that make a general claim” and the former do. Reliance on general principles is necessary for explanations but not for predictions.

I believe Hempel is wrong in his assessment of what distinguishes potentially explanatory predictions from non-explanatory predictions. There are clearly predictions *based on laws* that are still not explanatory. One could predict the extent of the tides at the time of the next full moon based on a law relating the position of the moon and the tides without thereby having the basis for an explanation of the extent of the tides. What is it, then, that distinguishes explanatory predictions from nonexplanatory predictions? One's answer to this question will constitute an important ingredient in one's general account of what explanation is, that is, in one's philosophy of explanation.

Hempel is right, of course, that not every prediction is explanatory, but is he also right that every adequate explanation *is* potentially a prediction? There have been dissenters. Since Hempel clearly commits himself to a universal claim about scientific explanation here, counterexamples are an appropriate response, and the dissenters have objected with counterexamples from science. There are many that have the same general thrust; the paresis-syphilis counterexample is the most well-known. According to this counterexample, one can explain why Jones got paresis by saying that he has syphilis and only people with syphilis can get paresis. One could not have predicted that Jones would get paresis just on the basis of knowing that he has syphilis, however, because only a small fraction of people with syphilis actually get paresis. So the explanation of Jones' paresis on the basis of his syphilis is not potentially predictive; thus it constitutes a counterexample to Hempel's universal claim.

Hempel responds to this example by saying that one cannot adequately explain why Jones contracted paresis simply by saying that he has syphilis and only those with

syphilis can get paresis. He says that the syphilis explains the paresis no more than Smith's buying a sweepstakes ticket explains his winning (pp. 369-70)! In either case, we lack an explanation because we can still ask why Jones got paresis or why Smith won, for it was possible, even highly probable, that Jones not get paresis and that Smith not win. This example has gained considerable significance, for as we shall see later, Wesley Salmon and others have used this example and similar examples to criticize Hempel's account of statistical explanations (or, more accurately, to criticize his general account of adequate explanation). The point for which Hempel comes under attack is present already here, however, in his universal claim that all adequate explanations are potentially predictive. We seem to have here a bedrock disagreement of intuitions about what is and is not an adequate explanation. It is just this sort of disagreement, a direct conflict of strong intuitions, that will become our central topic of discussion later, for we want to know, or perhaps *explain*, why people have the intuitions they do about explanation and why these intuitions can be different for different people.

For now, we need only to notice that Hempel defends his universal claim that all explanations are potentially predictive with the claim that any argument that is not potentially predictive cannot be an adequate explanation. This is not especially illuminating; it is merely contraposition (cf. Hempel's response to N. R. Hanson's example of beta-particle emission from a radioactive substance, which Hanson claims is fully explainable *ex post facto* but not predictable, pp. 407-408). But Hempel appears to exhibit a commitment to the latter claim very much like the commitment one would have to a statement one considers to be analytically true. Those who disagree, and

believe that syphilis explains paresis, likewise exhibit a very strong commitment to the claim that this is an adequate explanation. These same people would surely say that Smith's buying a sweepstakes ticket does explain his winning, or at least that that's all the explanation there is to be had.

A little further on (pp. 373-4) Hempel returns to this example to present what he thinks is an adequate account of what is going on. His comments indicate that he would say that we can explain why Jones contracted paresis with the statement that he had syphilis *only* if it is coupled with a claim to the effect that Jones is one of those people (in the minority) who has the *disposition* to develop paresis when exposed to (untreated) syphilis. The fact that Jones did contract paresis is evidence that he is in this group. In other words, there must be some factor, although we don't know what it is, that Jones has in common with others who get (would get) paresis when they have (if they were to have) untreated syphilis. This factor is not shared by those who do not get (would not get) paresis.

Hempel exposes here some of his *deterministic ontological assumptions*. He seems to be working with something like the assumption that the reference class for an explanans-fact can, objectively if not in practice, be narrowed to bring the probability of the explanandum fact indefinitely close to one and even eventually to one; this seems to be a consequence of his belief that for any low probability event there is a subclass of its reference class in which it will be converted to a high probability event. I will point out later that assumptions such as these, broadly metaphysical assumptions, play a large

part in shaping the crucial sorts of intuitions about explanation that I have been highlighting so far in this exposition.

There is another objection to Hempel's universal claim that all adequate scientific explanations are potentially predictive, an objection interesting not for its force against Hempel's claim (since it has none<sup>7</sup>) but for its bearing on the topic of *inference to an explanation* (the key to justifying ontological claims we examined in the previous chapter). Hempel claims that it might be objected that "sometimes the only ground we have for asserting some essential statement in the explanans lies in the knowledge that the explanandum event did in fact occur" (1965, 370-1).

In an inference to an explanation, the knowledge that an event has occurred is taken as evidence supporting a statement that is an essential part of an adequate explanation of the event. Hempel calls arguments "in which the information or assumption that *E* [the explanandum] is true provides an indispensable part of the only available evidential support for one of the explanans statements" a *self-evidencing argument* (p. 372).

Hempel provides a clear example. One explains the absorption spectrum of a particular star in part with an account of the star's chemical composition; it has this

---

<sup>7</sup> Hempel is claiming that all explanations can serve, in situations where the explanans statements are known to be true and the explanandum event has not yet occurred, to predict the explanandum event (or justify the belief that it will occur). The alleged objection here only asserts that one is sometimes not in such a situation; one is sometimes in a situation of knowing some essential part of the explanans only on the basis of knowing independently that the explanandum event has in fact occurred. This has no bearing on the counterfactual question—would one have been able to predict the explanandum event, if it had not yet occurred, only on the basis of the information presented in the explanans sentences?

spectral analysis *because* it is composed of hydrogen, helium, etc. in such-and-such proportions. The only evidence we have, however, that the star has such a composition is its absorption spectrum. The D-N explanation under consideration here is as follows:

- (1) The laws of chemistry, optics, spectroscopy, etc.
- (2) The particular chemical composition of star *S*
- (3) Therefore, the spectral absorption pattern of star *S*.

Although Hempel calls this a self-evidencing explanation, it is part of a clear example of an inference to (the best) explanation when used to justify belief in (2) on the basis of (3) and (1).

The term 'self-evidencing explanation' is perhaps not the best choice, for it leads to worries of circularity, which Hempel is quick to dismiss. One will face a dangerous sort of epistemic circularity only if one uses the argument both to justify belief in the relevant part of the explanans (2) and to justify belief in the explanandum (3). In other words, one will be in trouble if one believes (a) that the composition of the star is such-and-such *only* because the assumption of such a composition would (best) explain a particular absorption spectrum for the star, and (b) that the spectral analysis is such-and-such *not* on the basis of independent observation and test but *only* because the star has such-and-such a composition which entails (together with the laws of chemical spectroscopy) just this spectral pattern.

It is perfectly acceptable to use either justification on its own, however. If there were some independent means of identifying the chemical make-up of the star, one could use this information to justify belief that the spectral analysis will be such-and-such even

when that analysis has not been performed. Alternatively, one could justify belief that the composition of the star is such-and-such on the basis of an independently established spectral analysis by way of a standard inference to an explanation.

The argument (1)-(3) does not itself present an inference to an explanation. It is a standard deduction, and according to Hempel it's a deduction that constitutes a D-N explanation of a certain spectral analysis. The inference to an explanation that we have been considering relates to the argument (1)-(3) as follows:

- (I) [(1) & (2)]  $\supset$  (3) (from the validity of argument (1)-(3))
- (II) (3) The spectral analysis of star *S* is such-and-such (established observationally)
- (III) (1) The laws of optics, spectroscopy, etc. (established by independent empirical methods)
- (IV) Therefore, (2) star *S* has chemical composition (because this is the (best, only) adequate explanation of (3), given (1)).

It is important to decide which of the options (best, only) mentioned in (IV) is appropriate, for use of 'best' will render the inference inductive (since it suggests there are plausible alternative explanations) and use of 'only' will render it deductive. If it is deductive, however, it is still different from the deductive argument (1)-(3), for it has (2) as its conclusion.

The deductive version of this inference to an explanation would actually need an additional premise (corresponding to the parenthetical remark in (IV)) in order to avoid being enthymematic:

- (I') (1) The laws of chemistry, optics, spectroscopy, etc.
- (II') (3) The spectral analysis of star *S* is such-and-such
- (III') [(1) & (3)]  $\supset$  (2) ((2) is the only acceptable explanation of (3), given (1))
- (IV') Therefore, (2) star *S* has such-and-such a chemical composition.

For maximal perspicuity, the inductive inference to the best explanation should be expanded as well to include premises stating that the probability conferred on (3) by (2), given (1), is maximal, i.e., greater than the probability conferred by any alternative explanation. We will forego such elaborations here. This section of Hempel's article does, in any case, allow us to make at least some progress in clarifying the logical structure of inference to the best explanation.

The topic of statistical explanation has crept into the exposition at various points, because Hempel has been making various statements about scientific explanation in general. It is time now to take a more direct look at the nature of statistical explanation.

### Statistical Explanation

A statistical explanation is "any explanation that makes essential use of at least one law or theoretical principle of statistical form" (Hempel 1965, 380); we thus need to know what a statistical law is. Statistical laws have much in common with universal laws. Both support counterfactual and subjunctive conditionals. "Implications in the form of



counterfactual and subjunctive conditionals are thus hallmarks of lawlike sentences both of strictly universal and of statistical form” (p. 378). Statistical laws are also subject to something like the condition of unlimited scope; the reference class of a statistical law must be *potentially* infinite, i.e., unlimited.

According to one interpretation, statistical laws differ from universal laws by attributing a characteristic not to all members of a certain class, but to something less than all members of a class. A universal law asserts that all members of a class have a certain characteristic; a statistical law asserts that a specified proportion, but not all, of the members of a class have a particular characteristic (see Hempel 1965, 379). Statistical laws can be expressed alternatively by saying that an arbitrarily chosen member of the reference class has a certain *probability* of having a particular feature. The basic form of a statistical law is expressed as a probability,

$$pr(G, F) = r.$$

Hempel says, “Broadly speaking, this statement asserts that in the long run the proportion of those instances of *F* which are also instances of *G* is approximately *r*” (p. 376).

Hempel considers two types of statistical explanation. First, one can explain one statistical law by deductively subsuming it under more comprehensive statistical law(s). The deduction will be based on the principles of mathematical probability. This is called deductive-statistical (D-S) explanation. Second, one can explain a particular event by nondeductive subsumption under at least one statistical law. Because the subsumption

in this case is nondeductive, this type of explanation is called inductive-statistical (I-S) explanation.

The idea of I-S explanation is initially quite simple, even though it leads to some difficult problems. As an example, consider that we can explain why someone recovered from a streptococcal infection by saying that she was given penicillin and that penicillin cures streptococcal infection in a high percentage of cases. The explanatory argument is inductive (or probabilistic), for the conclusion could be false even when the premises are all true, because the patient could be one of those few patients that are not cured with penicillin. It is only highly probable that she be cured. The explanation is statistical, because the law that *most* people with strep recover when given penicillin is statistical.

The general structure of an I-S explanation of basic form is:

$$pr(G, F) = r$$

$$\underline{Fa} \quad [r]$$

$$Ga$$

The first premise states that the probability that something is *G*, given only that it is *F*, is equal to *r*. The second premise asserts of a particular thing, *a*, that it has feature *F*. The conclusion states that *a* has feature *G*, and the double line with [*r*] behind it indicates that the premises inductively support, or confirm, the conclusion to a degree of *r*. The only restriction that Hempel places on the argument is that it be a strong inductive argument, i.e., that it confer a probability “fairly close to 1” on its conclusion (p. 390). I-S explanation seems unproblematic at first, but there is a problem lurking in this type

of explanation: the problem of what is called the ambiguity of inductive-statistical explanation.

This problem is the consequence of a well-known difference between deductive and inductive arguments. A valid deductive argument cannot be made invalid by the addition of any new premise. An inductive argument, on the other hand, can be drastically changed by the addition of new premises. A very strong inductive argument can become very weak with the addition of a single premise. In general, the strength of an inductive argument has no bearing whatsoever on how strong it will be when new premises are added.

The effect this has on statistical explanation is to open up the possibility that, for a given statistical explanation, there is information that could be added in a premise that would render the explanandum sentence highly unlikely, thereby rendering the argument an explanation for the original explanandum sentence's *negation*. For example, if Smith actually recovers from strep infection when treated with penicillin, one can explain Smith's recovery with the information that she was treated with penicillin and that most people with strep recover when given penicillin. But imagine that she in fact has a type of streptococcal infection that is resistant to penicillin and hence does not recover upon treatment. Since it is a statistical law that almost none of the people with the penicillin-resistant infection recover from penicillin treatment, one can explain the fact that Smith did not recover with the information that she had a penicillin-resistant strain of the infection and very few patients with this strain recover with the penicillin treatment.

The problem is that if Smith recovers we can apparently (I-S) explain recovery simply with the information that she was treated with penicillin. If Smith has a resistant strain and does not recover, we can apparently (I-S) explain that too. The problem is that if Smith has a resistant strain there are two equally good arguments apparently explaining contradictory statements. As Hempel puts it, "Any statistical explanation for the occurrence of an event must seem suspect if there is the possibility of a logically and empirically equally sound probabilistic account for its nonoccurrence" (1965, 395). This is another one of those strong intuitions about explanation that we shall scrutinize later.

I do not share Hempel's suspicion of a statistical explanation if there is the *possibility* of an equally good explanation of the negation of the original explanandum, for the explanation one chooses will depend and ought to depend on the information one actually has, including the information that the explanandum event actually occurred. When there is a patient with a strep infection, it is legitimate to feel confident that we will be able to explain what will happen when she takes penicillin whether she recovers or not. As we have already seen, the occurrence of an explanandum event can sometimes be the only evidence one has for a part of the explanans. Given a patient with a strep infection who is treated with penicillin, we can say that if she recovers we will have evidence that the penicillin was effective and can explain the recovery by way of the penicillin. Further, we can say that if she does not recover, this is evidence that the infection is a penicillin-resistant strain, and we can then explain why the penicillin did not work. These will be self-evidencing explanations, in Hempel's terminology.

I do not believe that there is a problem simply because we can construct an explanation for Smith's recovery and an explanation for Smith's nonrecovery. The real problem of ambiguity, I believe, is that there can be a case of a patient with a penicillin-resistant infection who nonetheless recovers due to the penicillin treatment, and we can give an I-S explanation of this simply with the information that she had a strep infection, leaving out the fact that it was a penicillin-resistant strain. *This* presents a problem, at least for anyone who shares Hempel's general intuitions about explanation, because one *should not be able* to explain *this* case of recovery, unless one has *more specific* information about how the penicillin was effective in this case *even though* (i.e., taking into account the fact that) the infection was of a penicillin-resistant variety. I believe Hempel would say the problem is that we can construct an argument for the conclusion that Smith recovers that satisfies all the conditions of I-S explanation laid out so far without in fact adequately explaining the recovery. An adequate explanation must explain why the infection, which is *in fact* penicillin-resistant, responded to penicillin treatment in this case. It must identify a *narrower* reference class of which the patient is a member and for which recovery after penicillin treatment is highly probable.

One might object that the problem, in this form, is not threatening, because, given the patient's recovery, we probably don't *know* that we are dealing with a resistant strain, and when it comes to giving explanations, it is what we know (or justifiably believe) that matters, not what is objectively the case beyond our knowledge. All we know is that it is a case of strep infection, that the patient was given penicillin, and that the patient then recovered. Given our knowledge situation, then, the explanation is acceptable.

Unfortunately, the ambiguity of I-S explanation poses a problem even when we restrict our attention to a particular knowledge situation. This situation is known as the *epistemic ambiguity of I-S explanation*.

The epistemic ambiguity of I-S explanation results from the fact that one can often construct statistical explanations for contradictory statements even when one restricts oneself to the use of justified beliefs for the explanans (this is the case even if we assume that the set of all one's justified beliefs at a given time is logically consistent and closed under deduction). We get an illustration of this situation if we merely imagine that we have a patient whom we justifiably believe to have a penicillin-resistant strain of streptococcal infection. Before we administer the penicillin, we can say that whatever happens, whether the patient recovers or not, we will be able to explain it on the basis of our justified beliefs. If the patient does not recover, we can explain it on the basis of the belief that it was a penicillin-resistant strain; if the patient recovers, we can leave out of consideration the belief that the infection is penicillin-resistant and explain the recovery simply on the basis of the belief that it is a strep infection and that most strep infections are cured with penicillin. This ambiguity, like the former, presents a problem for Hempel; "It is disquieting that we should be able to say: No matter whether we are informed that the event in question ... did occur or that it did not occur, we can produce an explanation of the reported outcome in either case; and an explanation, moreover, whose premises are scientifically established [justified] statements that confer a high logical probability upon the reported outcome" (1965, 396).

With either type of ambiguity, the problem is that the I-S model qualifies as explanations some arguments that are not adequate explanations, at least according to some of Hempel's firmly held *intuitions*. In the former case, the model admits an argument that fails to be explanatory because it fails to identify the reference class that "objectively" explains the explanandum phenomenon; it disregards certain information that is relevant to the probability of the explanandum. In the latter case, the model admits an argument that fails to be explanatory because it disregards certain information *justifiably believed* that is relevant to the probability of the explanandum phenomenon. In our example, the disputed "explanations" ignore the statement (whether justifiably believed or not believed but true) that Smith has a penicillin-resistant strain of the infection, and this is a problem because that statement is relevant to the inductive argument under consideration (i.e., it will, if included in the argument, affect the probability conferred upon the conclusion). An analogous problem confronts prediction, as we can see if we consider that one can construct both a very strong prediction that Smith will recover with the penicillin and a very strong prediction that she will not recover, drawing only from our logically consistent and deductively closed set of justified beliefs.

Hempel's solution to the problem of ambiguity is related to Carnap's solution to the corresponding foundational problem in inductive logic.<sup>8</sup> Carnap imposed a requirement of total evidence for the calculation of the degree of confirmation for a given statement relative to a knowledge situation. The solution to the problem of ambiguity

---

<sup>8</sup> For an important distinction between the two, see Hempel 1968, 120-1.

is the requirement of maximal specificity for I-S explanation (pp. 399-400). It requires that the reference class used in the statistical law(s) in the explanans be maximally specific, and Hempel defines maximal specificity in reference to a body of accepted statements,  $K$ , as follows. We are given a proposed explanation of basic form:

$$pr(G, F) = r$$

$$\underline{Fa} \quad [r]$$

$Ga$

The conjunction of all the statements in  $K$  and the premises of an I-S explanation must be such that if they imply that  $a$  is a member of a class  $F_1$ , and that  $F_1$  is a subclass of  $F$ , then they must also imply that  $pr(G, F_1) = r$  (unless the probability of  $G$  given  $F_1$  is determined by a theorem of mathematical probability theory). This would mean that the explanation of Smith's recovery by way of Smith's membership in the class of all strep patients is unacceptable, because we know of a more specific class that is a subclass of the class of strep patients (i.e., the class of patients with *penicillin-resistant* strep) of which Smith is a member and in which the probability for Smith's recovery is different from what it is in the original reference class. As Hempel puts it, "The general idea thus suggested comes to this: In formulating or appraising an I-S explanation, we should take into account all that information provided by  $K$  which is of potential *explanatory* relevance to the explanandum; i.e., all pertinent statistical laws, and such particular facts as might be connected, by the statistical laws, with the explanandum event" (pp. 400-1).

The requirement of maximal specificity solves the problem of the epistemic ambiguity of I-S explanation, but does not solve the general problem of the ambiguity



of I-S explanation, because it makes essential reference to a set of accepted statements *K*. According to Hempel, this shows that “*the concept of statistical explanation for particular events is essentially relative to a given knowledge situation....* We will refer to this characteristic as the *epistemic relativity of statistical explanation*” (p. 402). To be more precise, we should call it the epistemic relativity of *inductive-statistical* explanation, because the phenomenon of ambiguity does not affect deductive explanations, whether nomological or statistical. We can require that anyone offering an I-S explanation use as a reference class in the relevant statistical laws only a class that is maximally specific. “[I]t remains the case [however] that for a given statistical argument with true premises and a high associated probability, there may exist a rival one with equally true premises and with a high associated probability, whose conclusion contradicts that of the first argument” (p. 402).

Why are we unable to come up with an “objective” version of the requirement of maximal specificity (even if we, as cognitively limited humans, in principle cannot know if it is satisfied in any actual case)? Why can’t we escape the epistemic relativization that Hempel calls “essential” with a requirement that there exist no subclass of the reference class in which the probability of the explanandum is different than it is in the reference class? It appears that Hempel is assuming (although he does not explicitly say this) what we will call thesis (S): There will *always* be a subclass of the type described for the reference class of any I-S explanation in which the probability of

the explanandum event is less than one and greater than zero.<sup>9</sup> This assumption rules out the possibility of an objective version of the requirement of maximal specificity, and hence it is one way to get the *essential* epistemic relativity of I-S explanation. Further, since the negation of (S) implies that epistemic relativity is *not* essential, the assumption of (S) is the *only* way to get the *essential* epistemic relativity of I-S explanation!

As an example of what (S) implies, consider the class of all people who have a streptococcal infection and are treated with penicillin. The probability of recovery is very high. First, there is the subclass consisting of patients with a penicillin-resistant strain, and in this class the probability of recovery is very low. Further, however, (S) implies that there is a subclass of this subclass consisting of all people with a resistant strain having an additional factor *X*; in this class the probability of recovery is different (let's say, very high). Hempel's comments on the proper analysis of the syphilis-paresis example (pp. 373-4) also express the commitment that there must be some additional factor (he calls it there a "disposition") that can designate a subclass in which the probability of contracting paresis is high (he also uses there an example involving skin cancer). Hempel seems to say that this is universally the case.

This assumption (S) implies the conclusion that there will always be a subclass, designatable without reference to *F*, in which the probability of the explanandum event is equal to one (and that there will always be a subclass in which the probability is zero).

---

<sup>9</sup> We can restate thesis (S) more precisely as follows: For every class *F* and every class *G* such that  $pr(G,F)=r$  (where  $0 < r < 1$ ), there exists a subclass of *F*, *F*<sub>1</sub>, such that  $pr(G,F_1) \neq r$  (where the probability of *G* given *F*<sub>1</sub> is *not* determined by a theorem of mathematical probability theory).

Assuming that the classes we are dealing with have a finite number of members, the series of subclasses of subclasses of subclasses... cannot go on indefinitely. Every time we have a subclass giving a probability strictly between zero and one, we can move to the next subclass giving a different probability, but this series must end somewhere. It cannot end with a class giving a probability strictly between zero and one, for (S) entails that there will be yet another relevant subclass of this class. So the series can end only with a class giving a probability of zero or one to  $G$ . When we are dealing with I-S explanations, we will naturally stay away from low probabilities. So it seems that Hempel is committed to the (objective) existence of a series of stronger and stronger I-S explanations of an event that eventually terminates in an explanation conferring a degree of confirmation of one on its explanandum (even if our "epistemic situation" forces us to stop somewhere before this). Eventually, there will be a class all of whose members are  $G$ , even if the class has only one member.

The assumption that for every  $G$  there will be a different feature (or set of features),  $F_n$ , such that  $pr(G, F_n)=1$  appears to be supportable only on the assumption of determinism. Indeterminism implies there is some  $G$  such that  $0 < pr(G, F) < 1$  and there is no subclass of  $F$ ,  $F_m$ , such that  $pr(G, F) \neq pr(G, F_m)$ .

The *essential* epistemic relativity of I-S explanation follows from an assumption of determinism. Another way of making the same point is that epistemic relativity is essential to I-S explanation if and only if all (scientific) predicates can, at least in principle, be defined purely extensionally (i.e., all predicates are fully determinate—bivalence holds for all sentences of the form  $Fx$ ). If there is a predicate

that cannot, even in principle, be defined extensionally, then there are some things that will have that property only with a degree of probability. It follows that there will be some classes that violate principle (S) with respect to that property and hence violate the claim that epistemic relativity is essential to I-S explanation. On the other hand, if epistemic relativity is not essential, then there will be a class  $F$  and a feature  $G$  such that  $0 < pr(G, F) < 1$  and there is no subclass of  $F$ ,  $F_1$ , such that  $pr(G, F_1) \neq pr(G, F)$ . From this it follows that there will be at least one predicate that has at least one instance that is merely probabilistic, i.e., extensional definition is impossible for this predicate.

Now we need merely note that the thesis of extensionality of predicates functions here as an *ontological* assumption about what it is for an object to have a property. This ontological assumption is intimately connected to the ontological assumption of determinism, mentioned above. They are just two sides of the same coin. The main point to keep in mind here is that it appears that the only way Hempel can defend his position on I-S explanation (actually it goes back to his general criterion for the adequacy of explanation *per se*) is with appeal to an ontological (metaphysical) assumption, either of determinism or of the extensionality of scientific predicates.

Wesley Salmon claims that Hempel, in private communications, has said to him that he is not committed to any form of determinism. He claims to support the essential epistemic relativity of I-S explanation with a semantic thesis: the very concept of objective homogeneity in a non-universal reference class is meaningless (Salmon 1984, 53). In other words, Hempel claims that it is *meaningless* to assert that there is a statistical law  $pr(G/F) = x$  ( $0 < x < 1$ ) where there is *no* subclass  $F \& F_1$  such that  $pr(G/F)$

$\neq pr(G/F \& F_1)$  and  $pr(G/F \& F_1)$  is not determined by a theorem of mathematical probability theory. I will return to this topic when I take up Salmon's account of objective homogeneity and claim that the concept is meaningful. For now, I can only say that the only apparent justification Hempel might have for his semantic thesis depends on an assumption of determinism; as Salmon says, the charge seems to be at bottom not a charge of meaninglessness but a charge that the concept is vacuous, and this is just a return to determinism.

We have found here what seems to be a metaphysical commitment shaping in profound ways an analysis of explanation. Peter Railton (1989) also claims to find deterministic assumptions in Hempel's analyses. This begins to answer the central question I raised near the beginning of this chapter, the question regarding the origins or sources of theorists' fundamental intuitions about what is and what is not an explanation. Hempel exhibits deeply entrenched intuitions about the requirement of predictive import for an explanation, for example. These sorts of intuitions appear to function, in virtue of their entrenchment, as criteria for *identifying* explanations, the very things one aims to model in a theory of explanation. We should turn to a closer examination of this activity that is central to theorizing about explanation. We will try to make more progress in finding the source(s) of the original criteria for *identifying* the data over which theorists of explanation theorize.

### Modelling Explanation

After having presented all his models of scientific explanation, Hempel clearly asserts a necessary condition for an adequate scientific explanation. “[A]ll adequate scientific explanations and their everyday counterparts claim or presuppose at least implicitly the deductive or inductive subsumability of whatever is to be explained under general laws or theoretical principles.... And the covering-law models represent, as far as I can see, the basic logical structure of the principal modes of such explanatory subsumption” (1965, 425). Hempel presents subsumability under a general law as a necessary condition for any adequate explanation, but he clearly does not present any one model as containing by itself necessary conditions for adequate explanation.<sup>10</sup>

Does he claim that his models give *sufficient* conditions for adequate explanation?

One outstanding problem that Hempel acknowledges speaks decisively against the thesis

---

<sup>10</sup> Very surprisingly, Michael Friedman claims that the D-N model (or, at least, a deductive model) gives necessary conditions for an explanation: “According to the D-N model, a description of one phenomenon can explain a description of a second phenomenon only if the first description entails the second” (1974, 198). This statement is doubly inaccurate, for not only is *entailment* not necessary for an adequate explanation (“partial entailment” is sometimes sufficient), covering laws are necessary, according to Hempel, for explanation, and Friedman fails to mention the general laws. Hempel nowhere talks of the description of one phenomenon on its own entailing a description of another phenomenon. Indeed, it is difficult to see how the description of one phenomenon might, without the aid of a law, entail the description of *another* phenomenon, although it might easily entail different descriptions of the *same* phenomenon. Friedman’s statement is puzzling also because he expresses it as a claim about what the D-N model asserts rather than about what Hempel (or anyone else) asserts, and it seems that models don’t assert anything (specifically, the D-N model of explanation does not assert anything about the necessary or sufficient conditions for explanation). People assert things *of* models, or, better, people assert that some things *are* models of other things, and this can mean a great variety of things in different contexts, because there are many different types of modelling.

that the D-N and D-S models contain sufficient conditions for explanation of a universal law and explanation of a statistical law, respectively. The derivation of  $L$  from  $L \& L^*$  is not necessarily explanatory, although it fits the D-N model if  $L$  is universal and the D-S model if  $L$  is statistical.

Are the models of the explanation of particular events meant to give sufficient conditions for adequate explanation? Some commentators have said as much. Wesley Salmon, for example, says that “if a particular fact is successfully subsumed under a lawful generalization, it is, on [the ‘received view’ of Hempel and others], completely explained” (1984, 21). The pendulum example still counts against the claim that the D-N model presents sufficient conditions. We can multiply examples having the same force. It is unclear whether Hempel would bite the bullet and say, for example, that the period of a pendulum D-N explains its length (even though the *everyday* concept of explanation, which we have perhaps left behind, might not allow this) or admit that there are arguments that fit the D-N schema but are not explanations (i.e., that the D-N model does not provide sufficient conditions for an adequate explanation).

Hempel’s concern with the phenomena of ambiguity suggests very strongly that he wants the I-S model to present sufficient conditions for an adequate explanation. The phenomena of ambiguity constitute a *problem* only if it is unacceptable to have arguments fitting the I-S schema without being adequate explanations. Yet the fact that Hempel leaves unresolved the original problem of ambiguity (rendering epistemic relativization essential) indicates that he does not put forward the I-S model of explanation for

particular events as presenting a set of sufficient conditions for an *objectively* adequate explanation.

In short, it is clear that neither the D-S model nor the D-N model for the explanation of laws presents a set of sufficient conditions, and Hempel acknowledges this. Hempel *appears* to acknowledge that the D-N model for explanation of particular events does not present a set of sufficient conditions, and in leaving the problem of the “objective” ambiguity of I-S explanation unresolved he seems to acknowledge that the I-S model does not present sufficient conditions for objectively adequate explanation.

#### Hempel’s Method: Conceptual Analysis

The difficulty of answering the questions of the above paragraphs stems from an unclarity in Hempel’s philosophical methodology.<sup>11</sup> We can ask a simple question, “What is Hempel trying to model in his models?” There is, however, no simple answer. In “Studies,” the authors never even speak of models. The opening to “Aspects” leads us to think that Hempel intends to model answers to explanation-seeking why-questions. Yet he does not feel any special allegiance to ordinary language. Wesley Salmon provides some useful comments regarding Hempel’s methodology:

The philosophers who were most instrumental in forging the old consensus—the *logical empiricists*—looked upon the task of philosophy as the construction of *explications* of fundamental concepts. The clearest expression of that goal was given by Rudolf Carnap. The concept we are attempting to explicate—in our case, *scientific explanation*—is known as the *explicandum*. This concept, which is frequently used by scientists and by others who talk about science, is vague

---

<sup>11</sup> Some of the room for disagreement about the significance of counterexamples for Hempel’s models is indicated in Bromberger 1966, esp. 82-86.



and, possibly, ambiguous; the job of the philosopher is to provide a clear and exact concept to replace it. The resulting concept is known as the *explicatum*. The process of explication has two stages: first, the explicandum must be clarified sufficiently for us to know what concept it is that we are trying to explicate; second, an exact explicatum must be precisely articulated (1989, 5).

Salmon says that in this process, remaining faithful to the original concept, the explicandum, is not of the greatest importance. One often has to move away considerably from the original concept in order to attain the requisite level of precision and clarity we are after. There is, in general, a tension between trying to stay true to the original concept, which is what we are after all trying to explicate, and improving upon it with a more precise concept. This tension brings it about that one will have to decide on individual points of divergence whether it is more important to remain close to the original concept or to move away from it for the sake of clarity. It is not at all obvious what criteria one ought to use to judge these matters.

With Hempel, we should take special notice of what he considers to be a threatening counterexample to his models and what he does not. A *counterexample* based on the original explicandum (the ordinary-language concept) will be effective only if similarity to the explicandum is of primary importance *at that particular point of the explication*. A *charge of inconsistency or ambiguity* in the explicatum is effective only if precision and clarity are of primary importance at that particular point of the explication. One can respond to a counterexample with the claim that we need to improve on, and hence leave behind, the explicandum at this point. One can respond to a charge of ambiguity by saying that the explicandum is essentially ambiguous in some respect. In examining how Hempel deals with potential counterexamples, we might find

a general principle at work guiding his decisions about what needs to be salvaged of the explicandum and what needs to be scrapped.

The pendulum example presents difficulties of interpretation. He is considering the D-N type argument from the period of a pendulum together with the laws correlating period and length to the length of the pendulum as a conclusion. He does not dispute that the argument has D-N form. Let's take another look at what exactly he says about this case:

A sentence stating the length of a given pendulum, in conjunction with the law, will be much more readily regarded as explaining the pendulum's period than a sentence stating the period, in conjunction with the law, would be considered as explaining the pendulum's length.... [T]he common-sense conception of explanation appears to provide no clear grounds on which to decide whether a given argument that deductively subsumes an occurrence under laws is to qualify as an explanation. (1965, 352-3)

Does this mean that we have to scrap this aspect of the common-sense conception of explanation? And if Hempel is scrapping this aspect of the common-sense conception, is he saying that it is an undeniable consequence of his explicatum that the period does explain the length so that one can *D-N explain* the length by the period? Or is he trying to say that, although the common-sense conception provides no clear criteria for distinguishing explanatory from nonexplanatory D-N arguments, it does tell us that the period of the pendulum *does not* explain the length of the pendulum? Why doesn't he say simply that period will not be taken to explain length? (Perhaps he is anticipating and leaving room for outré examples such as van Fraassen's of the tower and the shadow which do allow, in some cases, for explanation of the height of a tower by the length of its shadow, or of the length of a pendulum by its period.)

Although he does not appear to be greatly vexed by the pendulum “counterexample” (in 1965), he is greatly concerned (in 1948) to exclude arguments of the Mt. Everest type from his D-N model. The Mt. Everest case, remember, was meant to point out a logical flaw in the D-N model as Hempel and Oppenheim first presented it in their formal analysis (1948, 276). Their original formulation allowed for the explanation of any fact by way of any law. If  $L$  is a law (they used “All metals are good conductors of heat”),  $C$  is a deductive consequence of  $L$  (they used “If the Eiffel Tower is metal, then it is a good conductor of heat), and  $Fx$  is a fact (they used “Mt. Everest is snow-capped”), then one can D-N explain the fact by the law as follows:

$$L$$

$$\underline{C \supset Fx}$$

$$Fx$$

In order to rule out such arguments, the authors added another requirement, and claimed that the revised definition gave necessary and sufficient conditions of explanation (pp. 277-8).

Hempel’s concern with the Mt. Everest counterexample is a consequence of the fact that it renders this technical term utterly useless; it allows D-N explanation of anything by anything else, so long as the explanans and explanandum are true. “Counterexamples,” such as the pendulum, that show only that there are *some* (apparent) discrepancies between the technical term and the ordinary concept, are much less worrisome, for they do not necessarily threaten the usefulness of the technical concept (indeed, they may even mean it is more useful for certain purposes). It is apparently for

similar reasons that Hempel and Oppenheim do not try to exclude a certain type of partial self-explanation they discuss (1948, 274-6), for excluding it would exclude everything except purely theoretical explanation, and this would excessively *limit* the usefulness of their technical term.

In the foregoing cases, Hempel does not show much concern for preserving ordinary notions of scientific explanation, but he shows great concern for providing a technically useful explicatum. With his I-S model of explanation, however, he relies much more heavily on ordinary notions or pre-theoretical intuitions to indicate problems with his model. He attacks the epistemic ambiguity of I-S explanation as a serious problem threatening his model. As I have already explained, it seems that the real *problem* of epistemic ambiguity is that we can formulate an I-S form argument that is not, according to our pre-theoretical intuitions, an explanation of its conclusion. We can argue from the fact that Smith has strep and is treated with penicillin and that most strep patients recover with penicillin to the conclusion that she recovers, but if we also *know* that she has a penicillin-resistant strain, then this argument clearly (pre-theoretically) does not explain why Smith recovered. Hence, we appear to have a counterexample to the I-S model. Why doesn't this argument *explain* Smith's recovery? We do not have here a logical flaw as we did before, such that "anything goes." It seems Hempel is relying on some considerations of the ordinary notion of explanation; it is the common-sense concept of explanation that tells us that we don't have an explanation if we have left out information relevant to the outcome of the explanandum.

In sum, it appears that Hempel is concerned primarily with avoiding definitions that are too inclusive or too exclusive to be useful as technical notions of scientific explanation. He also shows substantial concern at some points, however, with ordinary notions of explanation. Much of this concern at crucial points makes sense only if he is working with the substantive ontological or metaphysical assumptions we have identified in connection with determinism.

Assumptions about what exists in the world and how existing things behave shape our intuitions about what counts as an explanation of some occurrence in the world. It remains unclear, however, *how* ontological assumptions influence our thinking about explanation. Certainly, ontology constrains explanation in that it tells us what is and what is not available to serve as explanatory facts and explanatory relationships between facts. There is more, however. Ontological assumptions also significantly constrain the possible ways for some event to “make sense” in the context of an overall view of the world. Ontological assumptions significantly constrain what we feel able to understand, or comprehend, and understanding is intimately connected to explanation. The importance of understanding as a guide to identifying explanations will come out clearly in the next chapter as we examine some more recent debates about scientific explanation. We shall then get a clearer picture of how explanation and understanding are related, and of how understanding is guided by ontological assumptions.

## CHAPTER 6

### CONCEPTS OF EXPLANATION, AND REALITY

The original intention of positivist analyses of explanation was to explicate a concept of explanation that would allow us to say that science does provide explanations but that explanation does not necessarily lead us into metaphysics (see Salmon 1989, 5). It appears now that the classic analyses of Hempel were guided or constrained, at least in part, by ontological, or metaphysical, assumptions<sup>1</sup> (see Salmon 1984, 50-5, and Railton 1989, 222-3). It looks as if one's view of statistical explanation and its relation to D-N explanation will be determined by one's metaphysical commitments. But is it possible to eliminate completely the influence of metaphysical assumptions in an adequate account of scientific explanation? Is it possible to *justify* a concept of explanation without recourse to any metaphysical (or ontological) assumptions? Is it possible to construct an adequate theory of explanation, or a theory of scientific explanation, that does not rest on, or commit one to, substantive metaphysical theories?

---

<sup>1</sup> I want to reintroduce the term 'metaphysical' here. The distinction I introduced earlier between metaphysics and ontology was meant to distinguish different methods of reasoning about what exists. When we turn away from methods of reasoning to the claims about what exists themselves, there is a *very rough* distinction between metaphysical claims and ontological claims. Metaphysical claims tend to be broader, more fundamental, more abstract and more general. Ontological claims tend to be narrower, less fundamental, more concrete and more specific. It seems that determinism is better described as metaphysical.

There are two separate questions here. First, is it possible to articulate and justify a concept of explanation without thereby committing oneself to substantive metaphysical or ontological theories? Second, is it possible to account for some actual explanatory practice (that of science, say, or some other area) without committing oneself to substantive metaphysical or ontological theories? It could be that the answer to the first question is “yes” while the answer to the second question is “no.” There may be some way to articulate and justify a concept of explanation that is completely metaphysically neutral. But it may also be true that there is no way to account for, say, *scientific* explanation while remaining completely metaphysically neutral, because science is, perhaps, essentially metaphysical.

### Two Types of Justification

First we need to explain what it is to “justify” a concept. Justification normally applies to propositions. There are various ways to justify a concept as well, however. Justification of a concept is justification of a particular explication of a concept. Explications of concepts *are* evaluated; there is ample evidence of this in any work of philosophical analysis, including analyses of ‘scientific explanation’. A classic example is the analysis of ‘scientific explanation’ as “reduction of the unfamiliar to the familiar.” There have been many criticisms of this analysis; it has been evaluated almost universally as unsatisfactory (for supporters see, e.g., Bridgman 1927, 37, Holton and Brush 1973, 185, and Dray 1961, 79-80; for criticisms see, e.g., Hempel 1965, 430-31, Salmon 1984, 13-14, and Friedman 1974, 191-192). The criticism is that it fails as an analysis

of 'scientific explanation' because there are many clear cases where science successfully explains by reducing some phenomenon that was familiar to theoretical principles that are unfamiliar and perhaps even quite counter-intuitive (see Hempel and Salmon for many examples).

One way, then, of evaluating, or justifying, an explication of a concept is to measure it against a body of particular judgments as to what does and what does not satisfy that concept. In explicating a concept of scientific explanation, there is a body of standard examples of actual scientific explanations that one must capture. One is, in this case, trying to construct a rigorous intensional definition for a term whose extension is supposedly clear. Counterexamples are appropriate criticisms of such conceptual analyses. If someone offers the intensional definition of 'E':  $(x)[(Ax \& Bx \& Cx) \equiv Ex]$ , then any of the following would refute the definition:  $Aa \& Ba \& Ca \& \sim Ea$ ;  $\sim Bb \& Eb$ ;  $\sim Ac \& \sim Bc \& \sim Cc \& Ec$ ; and so on.

Another way of justifying a concept is to recommend it as useful according to some instrumental standards of utility. When Hempel rejects an explication of the concept of scientific explanation on the basis that it allows for explanation of anything by anything else, he is apparently criticizing that explication primarily for its uselessness rather than its lack of fit with particular judgments about what is explanatory (although it is clearly deficient on these grounds as well). One can likewise recommend an explication of scientific explanation for its theoretical usefulness even though it does not count as explanatory certain cases that seem to common sense to be explanatory (or vice versa).



These two methods for evaluating explications of concepts seem to be the only methods that allow for clear resolution of disputes between competing explications. In particular, one may not appeal to any objective standard, independent both of actual patterns of usage and of instrumental concerns, that can serve for evaluation of an explication. The strategy of appealing to the essence of explanation is bound to be objectionable to anyone who disputes the purported essence, and any defense of the appeal to essences is bound to be question-begging to one who disputes the existence of such essences or our reliable access to them (see Moser 1993, ch. 1). Justification of an explication by the standards of “real definition” is doomed to failure with anyone who does not accept the appeal to essences involved.

In actual theorizing about explanation, there is typically a mixture of the two types of justification. We’ve seen this already in Hempel’s two types of reaction to potential objections. In theories of *scientific* explanation, however, the first type of justification takes precedence, because such theories are supposed to capture a specific aspect of actual scientific practice. Scientific practice gives us a large stockpile of actual scientific explanations. If Newton’s explanation of the tides by way of gravitational forces acting on the oceans is not a case of scientific explanation, then everything is up for grabs. If we can’t explain (approximatively) Kepler’s laws of planetary motion with Newton’s laws, we apparently can’t explain anything. On the other hand, if we can scientifically explain why Jones did not get the job by citing the position of the planets today in relation to where they were on the day of her birth, then our theory is clearly too inclusive.

### Causation in Scientific Explanation

This leads to one specific question: is it possible to justify an explication of scientific explanation without entering into metaphysics? In other words, is it possible to do justice to the actual practice of scientific explanation as it is currently practiced without relying on metaphysical theories. Many recent critics of Hempel apparently would say “no.” Some important lines of criticism have led directly into metaphysics, so that now theorists of explanation draw freely on more basic metaphysical convictions. Peter Railton claims, “Bas van Fraassen’s *The Scientific Image* and Wesley Salmon’s *Scientific Explanation and the Causal Structure of the World*, to mention only two prominent examples, are entirely explicit about placing analysis of explanation in a metaphysical setting, the one irrealist, the other, realist. The accounts they give of explanation, while of course making use of familiar strategies of exempling and counterexempling, are defended on grounds that do not purport to be innocent of metaphysics, and that involve metaphysically driven reinterpretations of intuitions” (Railton 1989, 221).

Causal analyses of explanation are the most obvious examples of how the analysis of a concept of explanation can depend on assumed metaphysical theories. According to David Lewis, for example, “To explain an event is to provide some information about its causal history” (1986, 185). He also argues that giving information about causal histories is *all* there is to explanation (pp. 188-192). Wesley Salmon argues, in *Scientific Explanation and the Causal Structure of the World*, that we can eliminate the inadequacies in Hempel’s account of scientific explanation *only* by making a causal account a *necessary* part of explanation.

Explanation is a two-tiered affair. At the most basic level, it is necessary, for purposes of explanation, to subsume the event-to-be-explained under an appropriate set of statistical relevance relations, much as was required under the S-R [statistical-relevance] model. At the second level, it seems to me, the statistical relevance relations that are invoked at the first level must be explained in terms of *causal* relations. The explanation, on this view, is incomplete until the causal components of the second level have been provided. This constitutes a sharp divergence from the approach of Hempel, who explicitly rejects the demand for causal laws (in *Aspects*, 352-354). (Salmon 1984, 22)

The aim of a scientific explanation, according to the ontic conception, is to fit the event-to-be-explained into a discernible pattern. This pattern is constituted by regularities in nature—regularities to which we often refer as laws of nature.... [S]ome regularities have explanatory power, while others constitute precisely the kinds of natural phenomena that demand explanation.... The relationships that exist in the world and provide the basis for scientific explanations are causal relations.” (1984, 121)

Causation is a distinctly metaphysical topic, and ever since Hume, empiricists have been careful to avoid any concept of causation other than the Humean analysis. Saying that explanation consists essentially of identifying causes requires one to give a full account of causation, to tell us what a causal connection is and how to distinguish it from an accidental correlation (or a correlation due to a common cause). This latter task seems to force us into metaphysics, beyond the realm of any possible experience. Hume tells us that when we experience an instance of causation, there is nothing in our experience that corresponds to the element of necessary connection, the element that is supposed to distinguish cases of causation from cases of mere (accidental) constant conjunction.<sup>2</sup> But this distinction is necessary for the purposes of explanation; there must be an objective distinction between genuine cases of causation and non-causal constant

---

<sup>2</sup> I will deal later with the additional complexities involved in notions of probabilistic causality.

correlation in order for causation to provide the key ingredient of explanation. The Humean analysis of causation, relying essentially on subjective psychological facts, does not give us what we need for an adequate causal account of explanation. It seems that we must go beyond experience and postulate a necessary connection (or, more likely, a connection that is not *necessary* but is nonetheless real) to distinguish causation from constant conjunction, and this makes it problematic how we identify, in practice, cases of causation.

Salmon takes seriously the challenge posed by Hume for anyone trying to give an account of causation. He admits, quite correctly, that “intellectual integrity demands that we squarely face Hume’s incisive critique of causal relations and come to terms with the profound problems he raised” (1984, 135). Accordingly, Salmon gives an account of causation that is supposed to support his theory of scientific explanation and to be acceptable to an empiricist philosopher. (He thinks it would be “extreme” to abandon empiricism; see his footnote 4 on p. 137.) I take this to mean that Salmon thinks he can avoid the metaphysics that the early positivists were so intent on eliminating, while still giving an account of scientific explanation that is adequate to actual scientific practice. I believe, however, that his account of causation gets us into some distinctively metaphysical territory.

Salmon chooses to deal primarily with causal processes (instances of causal propagation) rather than causal events (instances of causal production). His reasons for this need not concern us here (see pp. 138-141 for his comments). In order to characterize causal processes, he has to distinguish them from non-causal processes (he

has a few general comments distinguishing processes from non-processes on pp. 139-140). He refers to non-causal processes somewhat unfortunately as *pseudo-processes* (p. 141), even though they are, on Salmon's own account, actual processes. So he presents the contrast between causal processes and non-causal processes as a contrast between causal processes and "pseudo-processes": "Causal processes are those that are capable of transmitting signals; pseudo-processes are incapable of doing so" (p. 141). He elaborates on what it is to be able to transmit a signal with the idea of transmission of a *mark*. "The basic method for distinguishing causal processes from pseudo-processes is the criterion of mark transmission. A causal process is capable of transmitting a mark; a pseudo-process is not" (p. 142). He provides examples of what he means by transmitting a mark, as well as paradigm examples of a causal process and a pseudo-process.

There are problems with this account of causal processes, problems for empiricists anyway. Salmon says, "In order to qualify as causal, a process need not actually be transmitting a mark; the requirement is that it be capable of doing so" (p. 147). This mention of *ability* is illicit for a Humean empiricist. It is clear that a thing has the ability to do what it actually does. Actuality implies possibility; doing implies the ability to do. Attribution of an ability in other cases, in cases where the ability has not been manifested, are the interesting cases (see the previous quote from Salmon), and they are the problematic cases. What is it that one attributes when one attributes the ability to transmit a mark to something that has not yet transmitted a mark? It seems one is attributing a "mysterious power," the sort of thing Hume explicitly warns against.

Salmon responds by saying that attribution of an ability to transmit a mark is nothing like that, and that it is not the sort of thing that is objectionable to an empiricist. “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 experiments” (p. 147). He goes on to say that experiments give us clear cases of the manifestation of an ability to transmit a mark. When he says that ability to transmit a mark is a matter of performing certain experiments, he means that such an ability is made manifest by means of experiment. We see that a beam of light has the ability to transmit a mark (and is therefore a causal process) when we see it actually transmit a mark.

Salmon’s earlier comment, however, gives a certain importance to cases where a mark is not being transmitted. How do we attribute an ability where it is not manifested? It appears that Salmon is suggesting that experiment can provide a straightforward answer. Presumably, we can attribute the ability to transmit a mark to any beam of light on the basis of experiments in which *some* beams of light actually did transmit marks. This requires some movement from tokens to types and back to tokens. In the experiment, some token light beams manifest their ability to transmit a mark. More specifically, a light beam, *when* it transmits a mark, demonstrates its ability to transmit a mark, and a light beam that has transmitted a mark demonstrated its ability *at that time* to transmit a mark. Presumably, we are supposed to move from those tokens to types: all light beams have the ability to transmit marks. We can then apply this conclusion regarding a type of phenomenon back to particular token instances of that

type: this beam of light that was not in the experiment has the ability to transmit a mark; this beam of light that was in the experiment continues to have the ability to transmit a mark.

This is just the sort of thing empiricists don't like. Through this experimental procedure, we have not *proven, discovered, or experienced* the ability to transmit a mark in token light beams that have not actually transmitted a mark. We have merely accumulated some evidence that inductively supports the hypothesis that all relevantly similar light beams transmit marks when marks are introduced. In other words, we have confirmed, to a certain degree, the hypothesis that a certain law-like statement expresses a law. But is this anything more than the claim, based on a sample of observed correlations, that there is a constant conjunction between being a beam of light (i.e., being a certain group of phenomenal qualities) and transmitting a mark if a mark is introduced (i.e., including another phenomenal quality)? All of Hume's (and others') comments on the problem of induction apply here. It is hard to say why the observed correlation justifies predictions that it will continue and assertions that it would hold in counterfactual situations.

We need a subjunctive: If a mark were introduced into this light beam, it would be transmitted to all subsequent parts of the light beam. One typically supports a subjunctive or counterfactual statement with a law. An empiricist would see this as a simple case of logical entailment, however. If *all* light beams transmit marks when marks are introduced into them, then *this* light beam would transmit a mark if a mark

were introduced into it. Salmon would insist, I believe, that it is a matter of *physical necessity* that any light beam would transmit a mark (see pp. 91-92).

Salmon said that ability to transmit a mark was a particularly important *species* of constant conjunction. But he says nothing about how to isolate the particular species of constant conjunction he has in mind. He describes certain experiments, but does not say how we are to determine whether a constant conjunction that shows up in an experimental setting belongs to the species that constitutes ability to transmit a mark. Perhaps he means to say that that species of constant conjunction he has in mind is the kind that holds up under certain conditions of experimentation. He says some things suggestive of the appropriate *method* of experimentation (see pp. 149-50). He says nothing, however, about the *extent* of experimental pressure a constant conjunction must withstand in order to count as belonging to his “particularly important species.”

In short, Salmon has really done little to inform us how to identify an ability to transmit a mark, something that remains a “mysterious power.” It seems that he would need to rely on something along the lines of essential intuition, an intuition telling us that the ability manifested in some particular tokens is *essential* to the type these tokens represent. It should be clear that controversy over causation cannot be cleared up with appeals to *experimentation*, for this is bound to beg the question for anyone who disputes the distinction between causation and *mere* constant conjunction, i.e., anyone who disputes that there is any particularly important species of constant conjunction (objectively, apart from any specific *purpose*). In experimentation, one is confronted with constant conjunctions, and the only thing that makes them special in any way is,



perhaps, their *extent*, their *durability*. In experimentation, we try to do whatever we can to disrupt correlations. Either we succeed or we fail, and if we fail, we merely have a very hardy correlation. The fact that a correlation is hardy does not warrant, according to standard empiricists, the conclusion that the correlation is due to an underlying metaphysical necessity (or a physical necessity). Mill's methods give us further techniques for testing the hardiness of a correlation, but unless they are taken as *definitive* of what it is to be a causal connection, they do not help.

I think Salmon is wrong to claim that he has an account of causation that passes empiricist muster. I think he (along with many others) is right that an adequate account of scientific explanation requires a causal account, *if* we go along with his assumptions about what counts as a scientific explanation. In order to account for his cases of scientific explanation, we have to build on metaphysical *postulates*. The crucial examples are those that distinguish uniformities in experience that are explanatory from those that are not, where the latter constitute the very sorts of things that *require* explanation (see pp. 14 and 121). He uses his intuitions to distinguish between the two types of regularity, intuitions about what is explanatory and what is not. What reasons, if any, do we have to go along with Salmon's intuitions about explanation? Can a philosopher such as Hempel, for example, have *different* intuitions about what is explanatory and what is not, where these intuitions pick out cases of scientific explanation allowing for a strictly empiricist (non-metaphysical) account? The question of whether any common notion of explanation allows for an analysis that is not committed to contingent existence

claims forces us to confront the question of the nature and source of intuitions about what is explanatory and what is not.

### Objective Homogeneity

The examples of scientific explanation best supporting Salmon's theory are statistical explanations. Salmon claims that modern science can explain low probability events, thereby bringing into sharp relief the distinction between Salmon's ontic conception of explanation and other conceptions (1984, pp. 20, 47 and 84). Salmon says that, *if* radioactive decay is genuinely indeterministic, we can explain the radioactive decay in a given time period of an atom in a sample of  $U^{238}$  statistically, even if the prior probability of decay for that atom was very small (p. 82).

Salmon's ontic conception of scientific explanation, along with his defense of the legitimacy of statistical explanation as *independent* from deductive explanation in an indeterministic world, requires him to eliminate the epistemic relativity that Hempel claimed was *essential* to I-S explanation. Hempel's account was essentially epistemically relativized because it made essential reference to an *epistemically* homogeneous reference class for the statistical law(s) involved. Apparently Hempel rejected the idea of an *objectively homogeneous* reference class for the statistical laws. An account of objective homogeneity is exactly what Salmon needs, then, to remove epistemic relativity from statistical explanation. Salmon says that he must deal with the "difficult concept of physical randomness—that is, objective homogeneity of reference classes—which is

*required* to implement an ontic treatment of statistical explanation adequate to the possibly indeterministic context of contemporary science” (1984, 23, emphasis added).<sup>3</sup>

Salmon clarifies the notion of objective homogeneity: “Let us say that a reference class is *objectively homogeneous* with respect to a given attribute if there is in fact no way of effecting a relevant partition” (Salmon 1984, 49). Universal laws have *trivially* objectively homogeneous reference classes. Assuming it is true that all copper conducts electricity, the probability that a thing conducts electricity, given that it is copper, is unity. There is no partition of the class of all (samples of) copper (into two or more nonempty subclasses such that every member of the original class is a member of one and only one subclass) such that the probability of conducting electricity within any one subclass is other than unity.

For a reference class  $R$  of a statistical law to be objectively homogeneous with respect to a feature  $F$  means that  $0 < pr(F/R) < 1$  and there is no relevant partition of  $R$  including a subclass  $R_i$  such that  $pr(F/R_i) \neq pr(F/R)$ , where  $R_i$  cannot be the same as  $F$ . For example, if the probability of recovery upon treatment, given that one has strep and takes penicillin, is .95, and there is no relevant partition of the class of strep patients who take penicillin such that the probability of recovery in a subclass is other than .95, then the class of strep patients treated with penicillin is objectively homogeneous with

---

<sup>3</sup> See also pp. 54-55: “To formulate the thesis of indeterminism ... we seem to need the concept of [an objectively] homogeneous reference class.... We need the concept of objective homogeneity to capture the idea of an irreducibly statistical law.... [I]f our philosophical theory of scientific explanation is to do full justice to twentieth-century science, it must furnish an autonomous pattern of statistical explanation. As I argued previously, the epistemic relativization of I-S explanations seems to make that mode of scientific explanation parasitic upon D-N explanation.”

respect to recovery upon treatment. If, however, there is a relevant partition, say into patients with normal strep and patients with penicillin-resistant strep, then the class is not objectively homogeneous. This class might, nonetheless, be *epistemically homogeneous* for anyone who does not *know* that some of the cases of strep are cases of a penicillin-resistant strain and others are not. The reference class will remain epistemically homogeneous for a person as long as that person has no knowledge allowing for a relevant partition of the reference class (i.e., that person does not know how to make a relevant partition, regardless of whether there is such a partition).

Salmon's account of objective homogeneity in a reference class is quite complicated (1984, ch. 3). He mentions that the thesis of indeterminism implies that at least some non-universal reference classes will be objectively homogeneous with respect to at least one feature: "If, however, indeterminism is true—on any reasonable construal of that doctrine with which I am acquainted—some reference classes will be actually, genuinely, objectively homogeneous in cases where no universal generalization is possible" (1984, 53). The final qualification is necessary because any reference class is trivially objectively homogeneous with respect to a feature had by *all* its members, i.e., in which it is universally true that a member of that class has that feature. Indeterminism implies that, in addition to these trivially objectively homogeneous reference classes, there will be a class  $R$  in which the probability for a feature  $F$  is strictly between zero and one ( $0 < pr(F/R) < 1$ ), where  $R$  is objectively homogeneous with respect to  $F$ . This means that, although not all members of  $R$  are  $F$ , there is no way of making a relevant

partition of the class  $R$  into  $R \& S$  and  $R \& \sim S$  such that  $pr(F/R) \neq pr(F/R \& S)$ , where  $pr(F/R \& S)$  is not a theorem of mathematical probability theory.

We need to place certain restrictions on this definition, however. One needed restriction is immediately apparent. We are not interested in partitions that are made in terms of the very attribute in respect of which we are assessing homogeneity (see Salmon 1984, 56). Such relevant partitions are always possible, but if we allow such partitions to disrupt homogeneity, then there will be no cases of homogeneity other than the trivial cases; the concept of a non-trivially objectively homogeneous class will be vacuous. For example, we could partition the class of all strep patients treated with penicillin into the subclass of those who recover upon treatment and the subclass of those who do not recover upon treatment. This would be a relevant partition, but it is useless. The probability that you will recover ( $R$ ), given that you have strep ( $S$ ), are treated with penicillin ( $P$ ), and are a person who recovers upon treatment ( $T$ ), is unity [ $pr(R/S \& P \& T)=1$ ]. Correlatively,  $pr(R/S \& P \& \sim T)=0$ . It seems the partition must be made in terms of attributes that are logically independent of the attribute with respect to which we are assessing homogeneity; they must be definable without any reference to that attribute. This turns out to be a difficult task for Salmon, however, since he is determined not to make homogeneity epistemically *or linguistically* relativized in any way, and the mention of logical independence and conditions of definability relativizes this restriction to a language.

The claim that there could be non-trivial objectively homogeneous reference classes seems extremely counter-intuitive, because it seems extremely plausible to think

that, as long as some  $R$ 's are  $F$  and the rest are  $\sim F$ , there will *always* be a way of making a relevant partition, i.e., identifying an  $S$  such that  $pr(F/R) \neq pr(F/R \& S)$ . Salmon himself suggests one candidate<sup>4</sup>: the binary expansion of a real number between zero and one such that, on some pre-defined ordering of the members of  $R$ , a member of  $R$  is  $F$  if and only if it is paired with a 1 (and is  $\sim F$  iff it is paired with a 0). It is extremely tempting to think that there will exist for any class a mathematical technique for separating its members into two subclasses that matches exactly the subclasses distinguished by a given attribute defined independently of the mathematical technique. For this reason, Salmon begins with a discussion of the theory of mathematical randomness (pp. 55-60).

The payoff of an acceptable account of mathematical randomness would be that there are certain sequences that cannot be captured in terms of "effectively calculable" functions, i.e., general recursive, or  $\lambda$ -definable, functions (Salmon 1984, 58). This means at most that there would be some class such that there exists at least one ordering of its members such that there is no general recursive function that distinguishes just those elements having a given feature. This is not sufficient for Salmon's purposes, however. He needs at least one class  $R$  such that for *all* possible orderings of its members there exists no general recursive function that distinguishes just those elements having a given feature  $F$ . So even if we have an adequate account of mathematical

---

<sup>4</sup> See p. 57, where Salmon discusses the possibility of a *collective*, in the sense that Richard von Mises defines it, and uses the binary representation of a real number between zero and one to create a *place selection* for which the limiting frequency of an attribute does not match that in the main sequence. In example 4 on pp. 68-69, he deals further with the application of this technique to empirical cases.

randomness,<sup>5</sup> it will not suffice for Salmon's purposes of giving an account of objective homogeneity. Further, it seems intuitively virtually impossible that there be a class allowing for *no* ordering of its members for which there exists a mathematical function that isolates just those members having a given feature. For example, there is always the ordering that goes through all the elements having the feature  $F$  and then all the elements that are  $\sim F$ , and there is for this ordering a mathematical description identifying the point at which the  $F$ 's stop, hence dividing the class into  $F$ 's and  $\sim F$ 's.

Salmon defines homogeneity in an actual class (that is not trivially homogeneous):

*"A reference class A is objectively homogeneous with respect to an attribute B iff the probability of B within A is invariant under all selections by associated sequences"* (p. 70). I will not go into his account in detail here, but I will point out that the most he shows in this section is that, assuming indeterminism, a sample of radioactive material, for example, possesses no *physical* attributes that allow for a relevant partition with respect to radioactive decay within a given time period. He claims that this is sufficient for objective homogeneity. The restrictions he placed on relevant partitions effectively guarantee that only the physical properties actually had by the members of the reference class, and detectable prior to the possible assessment of whether they have the attribute with respect to which we are assessing homogeneity, are admissible for purposes of partitioning. As he says, many textbooks in quantum mechanics include as a basic *postulate* that the quantum mechanical description of a system is (physically) *complete*

---

<sup>5</sup> Salmon leaves this question open, "If that goal has not been fully achieved, we can at least take comfort in the fact that the job is in very good hands" (p. 60).

(p. 72). Again, he does not insist that the world is this way; he leaves it an open question whether the world is actually indeterministic. Even if the world is indeterministic and quantum mechanical descriptions are complete, however, it may still be that a sample of radioactive material possesses *non-physical* attributes (e.g., mathematically defined attributes) that allow of a relevant partition. It seems for this reason that the restrictions he places on objective homogeneity are artificial and *ad hoc*, being specifically designed to block the objective heterogeneity of indeterministic systems.

As far as I can tell, Salmon's restrictions on admissible methods of partitioning succeed in making the concept of a non-trivial homogeneous reference class non-vacuous. This was his objective. His restrictions are complicated and hard-won (and appear *ad hoc*), however, and I think there is a simpler way to achieve the same objective.

#### The Statistical and the Propensity Interpretation of Probability Statements

Salmon, like Hempel before him, works with a statistical, or relative-frequency, interpretation of the probability statements involved in probabilistic explanation; indeed, he refers to this type of explanation most frequently as *statistical* explanation, both reflecting part and rejecting part of Hempel's term, 'inductive-statistical explanation'. He rejects the 'inductive' because he rejects Hempel's claim that all explanations are arguments, but he keeps the 'statistical' because he claims there is an important connection between the explanations he is trying to capture and the explanations Hempel was trying to capture. The common characteristic is that these explanations do not



provide explanatory facts that necessitate the explanandum-event, rather they *make probable* (to some degree strictly between zero and one) the explanandum-event; they necessarily involve statistical laws, and statistical laws are always expressed in probability statements. The statistical interpretation of those statements is *not*, however, the only interpretation or even an obligatory interpretation.

Many interpretations of probability statements compete for prominence.<sup>6</sup> These interpretations reflect everyday discourse and scientific discourse. The claim that a standard die has a probability of 1/6 of coming up six and the claim that an American male over the age of fifty has a 2% chance of dying of a heart attack are both familiar sounding claims but already represent two different interpretations of probability statements. Some interpretations of probability statements are defective for scientific purposes, despite their intuitive appeal, because they allow for mathematically inconsistent systems of probability assessments. Two interpretations, however, that do not suffer from this defect are favored among scientists. The statistical interpretation is very common among scientists, for obvious reasons. The propensity interpretation is also common, however, and one can try to give an account of probabilistic explanation using a propensity interpretation of probability statements.

Salmon construes probabilities as relative frequencies, but denies that his account is tied essentially to that interpretation. "Those who prefer propensities," he says, "can easily make the appropriate terminological adjustments, by speaking of chance setups and

---

<sup>6</sup> Ian Hacking has called the notion of probability Janus-faced (1975, 12), but it is probably better described as Hydra-headed.

outcomes of trials where I refer to reference classes and attributes" (1984, 36). The adjustments are not that simple, however, nor are they inconsequential. Propensity theorists will also refer to attributes, but they give a very different interpretation to some statements about attributes. More importantly, a propensity interpretation makes it much easier to give an account of objective homogeneity than does the statistical, or relative-frequency, interpretation.<sup>7</sup>

An objectively homogeneous class is a class for which there is no relevant partition. Uniform classes are easy but trivial examples. What would a non-trivial case look like? Let's assume indeterminism and ask what it would be for a sample of carbon 14 to be objectively homogeneous with respect to the attribute of undergoing spontaneous radioactive decay within the next 5,730 years. The probability that an atom from this sample will decay within the specified time period is  $1/2$  (5,730 years is the half-life of carbon 14). Giving a statistical interpretation to this probability statement makes it equivalent to the claim that, within any (sufficiently large) sample of carbon 14, half the atoms will decay within 5,730 years (half will not). This interpretation leads naturally to the view that half the atoms in this sample have the attribute, in the present, of being an atom that will decay within the specified time. This view of attributes naturally leads to the view that the sample is not objectively homogeneous with respect to the attribute of being an atom that will decay within the specified time; half the atoms have the attribute, and half lack it. Any partition that manages to separate the atoms into those

---

<sup>7</sup> J. Alberto Coffa (1977) and Peter Railton (1978) treat probability as a single-case objective propensity, and consequently deal more simply with the problems Salmon tries to solve with a frequency interpretation.

having the attribute and those lacking it is relevant, for the (statistical) probability of decay in the former subclass is 1 and in the latter subclass is 0. None of this violates the assumption of indeterminism, because indeterminism merely tells us that there is no physical means of determining in the present which atoms have the relevant attribute and which lack it; the atoms are in fact physically identical at the present. It is consistent to say that all the atoms are physically identical at present but that they differ with respect to the attribute of being an atom that will decay within the specified time, for this means nothing more than that at the end of the 5,730 years half of these atoms will have decayed and half will remain atoms of carbon 14. Similarly, if I will be dead in 2050, then I have now the attribute of being a person who is dead in 2050.

Salmon blocks this move by saying, in effect, that he will consider a class to be objectively homogeneous if there is no relevant partition *in terms of an attribute that can (in principle) be determined to be relevant in the present*. He would say that the sample of carbon 14 is objectively homogeneous with respect to the attribute of undergoing spontaneous radioactive decay within the next 5,730 years *because* there is presently no way *in principle* to determine, for any attribute detectable in the present, whether it allows for an “associated sequence,” i.e., whether it relevantly partitions the class of carbon 14 atoms. This restriction succeeds in making the concept of non-trivial objective homogeneity non-vacuous, but it is much more involved than I have indicated here and, as I’ve said, appears *ad hoc*. It seems objectionable for reasons similar to those that led Salmon to exclude both epistemic and linguistic relativization; his restrictions are ultimately tied to an idealized, in-principle-possible knowledge situation.

Working with a propensity interpretation of probability statements alters much more than Salmon indicates. An adherent of the propensity interpretation would interpret the statement that the probability of spontaneous radioactive decay within 5,730 years for an atom in a sample of carbon 14 is  $1/2$  to mean that each atom in the sample presently has a propensity of  $1/2$  of decaying within 5,730 years. We need no artificial and complicated restrictions on the concept of objective homogeneity to say that this sample of carbon 14 is objectively homogeneous with respect to decay within the next 5,730 years. Objective homogeneity remains nothing more than the characteristic of allowing for no relevant partitions; the sample of carbon 14 is objectively homogeneous (assuming indeterminism) with respect to decay *in exactly the same sense* that the class of all pieces of copper is (presumably) objectively homogeneous with respect to being a good electrical conductor. *Any* partition of the sample of carbon 14 will give us subclasses all having a probability for decay of exactly  $1/2$ ; likewise, *any* partition of the class of samples of copper will give us subclasses all having a probability for electrical conductivity of exactly 1.

An adherent of the propensity interpretation will block the objection based on considerations of the possession, in the present, of the attribute of being an atom that will decay within the specified time more simply than Salmon does. Instead of saying rather artificially, along with Salmon, that certain attributes (that turn out to be very difficult to specify in a principled way) just do not figure in *relevant* partitions for the purposes of determining objective homogeneity, a propensity advocate can block the objection by rejecting the claim that the class of carbon 14 atoms *is* in fact heterogeneous with respect

to the above attribute. A propensity theorist can claim that each atom possesses that attribute to exactly the same degree, namely  $1/2$ , based solely on the fact that each atom presently has a propensity of  $1/2$  to decay within the specified time. A propensity theorist can (in fact, should) reject the claim that every entity either has completely or lacks completely any attribute. Spelling out the details of this would resemble to some degree Salmon's method of restricting admissible attributes for relevant partitioning, but the conceptual basis of the propensity approach is simpler and the restrictions seem much more natural, less *ad hoc*.

Salmon's claim that "if ... indeterminism is true—on any reasonable construal of that doctrine with which I am acquainted—some reference classes will be actually, genuinely, objectively homogeneous in cases where no universal generalization is possible" is not obviously true when one is using a relative-frequency interpretation of probability statements. Salmon needs to put many restrictions on the concept of a relevant partition in order to get the implication. They include the restriction that the partition could not be effected with the very attribute in terms of which one is assessing homogeneity, and they continued through many more restrictions. If one uses a propensity interpretation of probability statements, however, and the accompanying *metaphysical* view of attribute possession, then Salmon's statement is obviously true. No restrictions are needed on Salmon's original (and simple) formulation of the concept of objective homogeneity, and it therefore remains clearly and closely related to the concept of epistemic homogeneity. Using a propensity interpretation, and assuming indeterminism, the class, say, of carbon 14 atoms is objectively homogeneous with

respect to spontaneous radioactive decay within the next 5,730 years, even if we partition using the attribute of being an atom that will decay within that time. So we don't even need Salmon's first restriction, which seemed so obviously needed when we were using a relative-frequency interpretation of probability statements. These considerations make it seem useful, if we assume the metaphysical position of indeterminism, to employ a propensity interpretation of probability in sciences dealing with indeterministic systems.

For my purposes in this dissertation, the significance of the foregoing considerations is as follows: the question of how to interpret the probability statements in statistical laws is relevant to one's *intuitions* about explanation. Because Hempel explicitly employed a statistical concept of the probability statements that appear in statistical laws, it was very difficult, perhaps even impossible, for him to conceive of what it would be for a reference class of a probabilistic law to be objectively homogeneous; the notion of homogeneity seemed necessarily connected exclusively with reference classes of universal laws. This had a profound impact on his conception of probabilistic explanation (in fact, it led directly to the most controversial, most discussed aspect of his models); it led him to a conception of probabilistic explanation that is completely dependent on and subordinate to the notion of deductive explanation (see Coffa 1974 and Salmon 1984, 52-53 for a defense of the claim that I-S explanation is for Hempel completely parasitic upon D-N explanation).

Salmon has shown, I believe, that one can make sense of this notion and that it can be instantiated if indeterminism is true, with a statistical interpretation. He has simultaneously shown, however, how difficult it is to do this. It is not something one

would simply fall into, but only something one might work hard for if, like Salmon, one were thoroughly committed, first, to the conceptual possibility of indeterminism and, second, to the view that indeterminism entails that there are objectively homogeneous reference classes for probabilistic laws. The possibility of indeterminism seems a good thing to commit oneself to. The entailment is not obvious, however; I have claimed that it is consistent to accept indeterminism while holding that *necessarily* all homogeneous classes are the trivial ones. Indeterminism clearly entails that there are irreducibly (for any possible conceiver) probabilistic laws in physics, but not that there are objectively homogeneous reference classes for those laws. One will take it to entail that there are homogeneous reference classes for those laws if one is committed to the position that a physical description of a low-probability event in an indeterministic system in terms of probabilistic laws and antecedent conditions *can* constitute an *adequate explanation* of that event. If by 'adequate explanation' one means an explanation that is not epistemically relativized, one will have to have objectively homogeneous reference classes for probabilistic laws in order to get adequate probabilistic explanations. Salmon is, I believe, thoroughly committed to that position, but now we can ask *why* he insists that some quantum mechanical descriptions are (assuming indeterminism) fully adequate explanations.

#### Intuitions Contrary to Salmon's

Salmon aims to give an account of scientific explanation according to which current physics, especially quantum mechanics, is *explanatory*. He seems to think that a theory

of explanation that claims that quantum mechanics does not offer explanations somehow *diminishes* or *fails adequately to appreciate* an achievement of contemporary science that is in fact *great*. As I've said, it is possible adequately to appreciate the greatness of quantum mechanics without considering it to be explanatory. Why does Salmon insist that it is explanatory? What *standard of explanation* does he use to judge it to be explanatory? Whatever standard it is, I find it questionable; it is not, in any case, a criterion that everyone will readily accept as self-evident. According to his standards of what counts as an explanation, "when an alpha particle in the nucleus of a  $U^{238}$  atom approaches the potential barrier that constitutes the wall of the nucleus, it has a probability of about  $10^{-38}$  of tunneling through, but when such a spontaneous radioactive decay occurs, the explanation is that a low-probability event has occurred" (pp. 85-86). I honestly have no *decisive* intuitions about this alleged example of explanation, and I doubt that there would be a wide consensus among scientists and philosophers of science that this is a sketch of an *explanation* rather than of just a *description*. I feel that, in some sense, this event has *not* been explained, even when the sketchy account has been filled out with all the appropriate initial conditions and statistical laws.

### Contrastive Why-Questions

This example might seem similar to the example of a lottery win. It might seem that one can adequately explain why Janie Jones won the lottery merely by saying that she bought a ticket and thus had a chance, however small, of winning. Our sense that this is an adequate explanation is aided by the claim that, given the setup of the lottery, this is all



the explanation anyone can hope to get, just as our intuition that quantum mechanics explains the radioactive decay is aided by the assertion that that's all the explanation there is. However, our sense that this is an adequate explanation is hindered by the observation that many other people also bought tickets but did not win. When we ask the *contrastive* why-question 'Why did Janie Jones rather than any one of the other players win?', we get the definite feeling that we don't have an explanation when we say merely that she had a chance of winning. (We *would* feel that we had an answer to that question if we had an account of how the game was fixed so that Janie had to win.) So we might retreat to the claim that we can explain why *someone* won because, given the setup of the game, it was either likely or, perhaps, certain that someone would win, but then we are no longer explaining why *Janie Jones* won. We would then be explaining why someone rather than no one won. Similarly, we might explain why *a* radioactive decay occurred, since the occurrence of a decay is probable, but this is not to explain why *this* atom decayed; it is to explain why a decay occurred rather than no decay.

Apparently, our intuitions can be affected by where we stand on the controversy over contrastive why-questions. One might feel, for example, that we can explain why Janie Jones won the lottery although we cannot explain why Janie Jones rather than Juan Silva won the lottery. Likewise, one might feel that we can explain why this alpha particle tunneled through although we cannot explain why this alpha particle tunneled through at this time rather than earlier or later. Our intuitions about what is and what is not an explanation are tied to how the why-question is formulated, whether as a

contrastive why-question or not. The issue is whether every explanation-seeking why-question is a *contrastive* why-question.

Salmon apparently realizes this, because he tries to keep us from understanding the question why the alpha particle tunneled through as a certain type of contrastive why-question. He is cautious about contrastive why-questions: “If we assume that an explanation of why one outcome occurs must *ipso facto* be an explanation of why one rather than another occurs, we run a serious chance of finding ourselves involved in the notion that only those events that are strictly determined can be explained” (1984, 110). He apparently recommends that we *not* assume that an explanation of why one outcome occurs must *ipso facto* be an explanation of why one rather than another occurs. This is possibly correct with respect to *some* contrasts. Salmon calls it a “fundamental error” to suppose that an explanation of why  $P_k$  occurred is *automatically* to explain why  $P_k$  rather than  $P_j$  ( $j \neq k$ ) occurred (*ibid.*). An answer to an explanation-seeking why-question tells us *why*  $P_k$ , but “we have not necessarily been told *why*, but only *that*,  $P_j$  is false” (*ibid.*).

I will grant Salmon’s claim that an explanation of *why*  $P_k$  is not *necessarily* an explanation of *why not*  $P_j$ , where  $j \neq k$  and  $P_j$  is a member of the appropriate contrast class (although I am having difficulty thinking of an example where it isn’t). This does not entail, however, that there are *any* explanations of *why*  $P_k$  that are not automatically also explanations of *why not*  $P_j$ , for some  $P_j$  ( $j \neq k$ ). For example, I take it that an explanation of why the flame in the Bunsen burner turned *green* is *ipso facto* an explanation of why it did not turn blue, and of why it did not turn magenta, and of why

it did not turn yellow, etc. etc. This is because the flame's turning green physically necessitates its not turning blue, etc. Also, to explain why the alpha particle tunneled through is *ipso facto* to explain why it was not reflected, since its tunneling through physically necessitates its not being reflected.

We don't need to assume, for we can demonstrate, that an explanation of why one outcome occurs must *ipso facto* be an explanation of why one rather than another occurs, for an explanation of *why P* is *ipso facto* an explanation of *why not ~P*; an explanation of why one outcome occurs is *ipso facto* an explanation of why it rather than its complement-outcome occurs, or why it did not fail to occur. This is because the occurrence of one outcome always guarantees the non-occurrence of another outcome, its complement-outcome; the truth of *P* *logically* necessitates the falsity of  $\sim P$ . So the explanation of why the alpha particle tunnels through must *ipso facto* explain why the alpha particle tunnels through rather than fails to tunnel through, and it is a logically contingent but physically necessary fact that failing to tunnel through implies being reflected. Since an explanation of why one outcome occurs must *ipso facto* be an explanation of why one rather than another occurs, we apparently do have a good chance of finding ourselves involved in the notion that only those events that are strictly determined can be explained, and that is why I find Salmon's criterion for explanatoriness at least questionable.

Salmon claims that we can explain why an electron is reflected by a potential barrier even though it had a 0.1 chance of tunneling through (p. 110; I have no idea why

he chooses the explanation of a *high-probability* event for this example). But that explanation does *not* explain

why it was reflected rather than transmitted in this case; it just happened that way, as it does in 9/10 of such cases. In 1/10 of such cases, it just happens that the electron tunnels through and is not reflected. But, by hypothesis, there is no explanation for the fact that it does one of these on one occasion and not on another. (*ibid.*)

I am baffled by these comments. I understand how explaining why the Bunsen burner flame turned green does not necessarily explain why the candle flame did not turn green, though in some contexts it does *assert* that the candle flame did not turn green. It seems, however, that explaining why the electron was reflected would just *be* to explain why it did not tunnel through the potential barrier, since these are the only two physically possible options. Hence, to explain why it was reflected just is to explain why it was reflected rather than transmitted in this case.

In the last quotation from Salmon, he seems to use the phrase “it just happened that way” to indicate when an explanation is lacking. That phrase seems clearly appropriate in response to the question why the alpha particle tunneled through, or to a question why any indeterministic event happened, for that matter. We share, I believe, strong intuitions that we have an explanation for an event only if we have an account of what brought it about, but for an indeterministic event we cannot have an adequate account of what brought it about. In the case of indeterministic events, the antecedent conditions for one event will be identical with the antecedent conditions for a contrary event. It is counter-intuitive to say that the particle’s approaching a potential barrier brings it about (*if* the particle tunnels through) that it tunnels through the barrier, and that

exactly the same approach to the barrier brings it about (*if* the particle is reflected) that it does *not* tunnel through.

My intuitions seem pretty strong that we don't have an explanation of why the alpha particle tunneled through when we note that to explain that just is to explain why it tunneled through rather than being reflected, where it had a  $10^{-38}$  probability of tunneling through and a  $1-10^{-38}$  probability of being reflected. It seems all we can say is "it just happened that way, as it does  $10^{-38}$  of the time." If this is *all* we can say, it seems clear that we cannot *give an account* of why it happened.

A plausible (though unproven) approach to the contrastive nature of why-questions supports intuitions contrary to Salmon's. We have here a basis for rejecting some of Salmon's claims about what is and what is not explanatory. If we reject some of Salmon's intuitions about explanation, we might be able to construct an adequate theory of explanation that has no metaphysical commitments (such as commitments about causation or commitments about the nature of attribute-instantiation, examples we've looked at in this chapter). Examining different approaches to the contrastive nature of why-questions does not, however, give us a complete account of the nature and origin of intuitions about what is explanatory. One's approach to the contrastive nature of why-questions seems rather to stem from the same source as one's intuitions about what is and what is not explanatory. It may even follow directly from some strong intuitions about explanation. (Salmon's rejection of the claim that all why-questions are essentially contrastive is the direct result of his commitment to the explanatory status of certain

quantum mechanical explanations.) We thus need to look further for an adequate account of the nature and origin of our intuitions about what is and what is not an explanation.

### Conclusion

We began this chapter with the observation that Hempel's classic analyses of scientific explanation seem defensible only on the metaphysical assumption of determinism, and we asked whether it was possible to explicate any common notion of explanation without relying on metaphysical assumptions. We found first that recent critics of Hempel have in fact helped themselves to hefty metaphysical assumptions in their theorizing about explanation. Wesley Salmon is one prominent recent theorist of explanation, and I claimed that, despite his protestations to the contrary, he relies on an account of causation that a Humean empiricist would consider objectionable metaphysics.

We found that whether it is possible to explicate any common notion of explanation without employing metaphysical assumptions depends on how one identifies particular cases of explanation. The identification of explanations depends in turn on one's intuitions about what is explanatory and what is not. We thus had to turn to an examination of the nature and source of our intuitions about explanation. We examined in depth the special significance in the present philosophical climate of conflicting intuitions about whether quantum mechanics provides *explanations* (in addition to accurate descriptions) of low-probability subatomic events. Salmon feels that quantum-mechanical accounts are genuinely explanatory, and develops his doctrine of objective homogeneity as a necessary component of his theory of explanation. His doctrine of

objective homogeneity, we saw, would be greatly aided by yet another metaphysical postulate; a view of attribute-instantiation that goes with a propensity interpretation of probability statements. This is yet one more metaphysical commitment that will shape one's intuitions about explanation in such a way as to support a particular theory of explanation.

Is there any metaphysically neutral way to account for the nature and origin of our intuitions about what is and what is not explanatory? A doctrine concerning the contrastive nature of why-questions shows some initial promise of being a metaphysically neutral force shaping our intuitions about explanation. One's view about contrastives does not, however, shed much light on the nature and origin of one's intuitions about explanation, for it stems from the same source as those intuitions or even from some of those intuitions themselves.

In the next chapter, we will look more deeply into the nature and origin of our intuitions. We shall find that they are, in fact, embedded in a presupposed metaphysical system; they are not metaphysically neutral. We shall finally be in a position to address the problem of there being either (a) a circular relationship between theories of explanation presupposing metaphysical claims that are themselves supported essentially by explanatory arguments, or (b) a regress of convictions about explanation embedded in metaphysical assumptions, metaphysical claims supported essentially by explanatory arguments, more general claims about the nature of explanation embedded in more general metaphysical assumptions, and so forth. We shall find a satisfactory way to

avoid both a vicious circularity and an endless regress in the relationship between explanation and metaphysics.



## CHAPTER 7

### EXPLANATION, UNDERSTANDING, AND ONTOLOGY

#### Intuitions About Explanation, and Understanding

Combining his statistical-relevance basis with his account of causation, Salmon claims to have a complete account of scientific explanation. What exactly is it he's giving an account *of*? He claims to give much attention to the "clarification of the explicandum" (see Salmon 1984, x; see also Carnap 1962, ch. 1 for the classic exemplar of this method). There is one crucial point he neglects, however. He does not tell us how he selects the examples of scientific explanation that are supposed to clarify the general concept for which he wants to provide a philosophical explication. Salmon, like so many others, offers his line of preferred examples, pronouncing that *this* is what we aim to characterize, without telling us how one determines, independently of his theory of explanation, that they *are* explanations. He and many other philosophers of explanation work with the assumption that the aim is to construct a theory of scientific explanation that "fits" all the examples we can adduce. There will be disputes about examples, however. I shall look at one crucial area of current dispute about explanatoriness to help clarify the source of disputed intuitions.

Salmon provides compelling examples to encourage intuitions that will agree with his own. There are allegedly obvious problems with Hempel's I-S model. Hempel

allows for explanation of high probability events but rules out explanation of events that are not highly probable (he says the explanatory argument must confer a “more or less high degree of inductive support or of logical (inductive) probability” on the explanandum sentence; see his 1965, 385-6). Hempel intentionally leaves this vague. “Of course, an argument of this kind will count as explanatory only if [the probability] is fairly close to 1. But it seems impossible, without being arbitrary, to designate any particular number, say .8, as the minimum value of the probability  $r$  permissible in an explanation” (1965, 390).

The purported problem with his restriction is not its vagueness but the fact that it decisively rules out explanation of low probability events. It does not matter that we cannot give a precise cut-off point for the probability in question, at least in the present context. What matters most is that there are some probabilities Hempel would count as clearly sufficient for explanatoriness (.9, for example) and some that he would count as clearly insufficient for explanatoriness (say .6).

This leads to the problem of “conjunctivitis” (Kyburg 1970). Since the probability of the joint occurrence of two independent events is equal to the product of their individual probabilities, the probability for a conjunction of two independent events with probabilities less than 1 is always less than that of either conjunct. Hence, it is always possible to construct the conjunction of a number of independent events such that each event is explained with a sufficiently high probability but their conjunction has a low enough probability to guarantee that it is *not* explained. For example, if one has explanations for 5 independent events, each of which is given a probability of .9 by its

explanatory facts, the conjunction of all those events will have a probability below .6 given all those same explanatory facts.

Salmon calls conjunctivitis a “serious malady” for Hempel’s model of I-S explanation (1984, 87). Hempel merely says that I-S explanations are nonconjunctive, whereas deductive explanations are conjunctive (1965, 410-12). He is apparently biting the bullet and admitting that in some cases we cannot explain the conjunction of a number of events even when we can explain each event individually. Biting the bullet is a telling philosophical move. It amounts to holding on to a theory in spite of its allegedly counter-intuitive consequences, *rejecting the intuitions* that run counter to the theory. Hempel seems quite comfortable rejecting the notion that whenever we can explain a number of events individually we can automatically explain the conjunction of all those events *with the same facts we used to explain them individually*. And, come to think of it, that doesn’t seem obviously true after all.<sup>1</sup>

Salmon brings other intuitions to bear, however. He claims that Hempel’s theory leads to the counter-intuitive result that in a population of pea blossoms where 3/4 of the blossoms are red and 1/4 are white, we can explain why a particular red blossom has the color it has (because it is probable) but cannot explain why a particular white blossom has the color it has (because it is improbable) (1984, 86). The example comes from Hempel, and again we are presented with conflicting intuitions from Salmon and Hempel. Hempel’s example, however, is actually slightly *different*. He imagines a population of

---

<sup>1</sup> See Colin Howson 1988 for a defense of Hempel’s account of statistical explanation based on the rejection of the conjunctive closure principle for explanation. See also Ruben 1990, 30-1.

pea blossoms, all of which are offspring of a pure-bred red blossom and a pure-bred white blossom. He says that we can explain why the distribution of red blossoms to white blossoms in this population is approximately 3:1, because Mendelian genetics makes this result highly probable (1965, 391-2). He does not say that we can explain why a particular red blossom has the color it does. So the appropriate case to consider is a (sufficiently large) population of blossoms that diverges significantly from the 3:1 ratio. Mendelian genetics renders this result improbable. Hempel is apparently comfortable with the consequence that we could not explain (*on the basis of Mendelian genetics*, at least) a sufficiently large population from the same two pure-bred blossoms with, say, a 1:3 ratio of red to white blossoms.

Salmon, however, is trying to make Hempel's view seem as uncomfortable (i.e., counter-intuitive) as he can. He says that in a population of 3/4 red blossoms and 1/4 white blossoms, it is clear that we can understand just as well why a particular blossom is white as why another blossom is red. "I find it hard to believe that we do not understand the occurrence of white blossoms in this population just as adequately as we understand the occurrence of red" (1984, 86).

It is interesting and quite significant that here, and in many other places where Salmon tries to draw on supporting intuitions, he falls back on considerations of *understanding* as a criterion for when we have an explanation. We have seen some instances where Hempel's and Salmon's intuitions about what is and what is not an explanation apparently conflict. This leads us naturally to ask how they determine when we have a scientific explanation.

When one tries to assess Salmon's explication of scientific explanation in comparison with Hempel's, Salmon wants us to think that it is a straightforward matter of comparing how well each handles a large number of examples of scientific explanation. It seems, however, that the counterexamples Salmon (and others) bring against Hempel would not be taken as counterexamples by Hempel himself (or at least need not be admitted as such by someone defending Hempel's analysis), because he would not regard them as genuine cases of scientific explanation. He has *different intuitions* about such cases.

The key point of dispute between Hempel and Salmon concerns what science would be like if the universe is actually indeterministic. Salmon believes that there would then be genuine scientific explanations of low-probability events, such as the radioactive decay of an atom of carbon 14. Everything Hempel says about explanation suggests that he would say that such an event is simply inexplicable. It is nonetheless difficult to find in Hempel any clear principle used to identify scientific explanations. Other authors give us a much clearer picture of certain intuitions about explanation that are directly opposed to Salmon's. G. H. von Wright, for example, represents an approach to explanation that Salmon has named the "modal conception." The modal conception of explanation requires that every complete and adequate explanation render its explanandum necessary given the explanatory facts invoked. Von Wright *denies* the possibility of statistical explanation:

[I]t is part and parcel of an inductive-probabilistic explanation that it admits the possibility that *E* might have *failed* to occur. It therefore leaves room for an additional quest for explanation: why did *E*, on this occasion, actually occur and why did it not fail to occur? It would be the task of deductive-nomological

explanation to answer *this* question. Sometimes we can answer it.... [But] failing such additional information which gives us a deductive-nomological explanation of *E*, we have not explained why *E* occurred, but only why *E* was to be expected.... It seems to me better, however, not to say that the inductive probabilistic model *explains* [emphasis added] what happens, but to say only that it justifies certain expectations and predictions. (von Wright 1971, 13-14)

Salmon cites another defender of the modal conception (1989, 175): John Watkins expresses an attitude similar to von Wright's in defense of deductivism regarding explanation. "I have argued that it is a mistake to call upon microphysics to *explain* an individual micro-event, such as the disintegration of a radon atom or the reflection of a photon; for if it really was a matter of chance which way it went, then the fact that it chanced to go this way rather than that simply defies explanation. What we *can* explain with the help of an appropriate indeterministic microphysical theory is why there was a precise objective probability that it would go this way" (1984, 246).

Here we are presented with clearly opposed claims. Salmon would be perfectly content with an explanation of why an atom underwent radioactive decay within a given time period that says only that such an event had a non-zero probability (of, say,  $10^{-38}$ ) of occurring<sup>2</sup> (we must remember, of course, that Salmon always asserts such things only on the condition that the universe turns out in fact to be indeterministic). Von Wright and others would point out that, according to such an "explanation," the radioactive decay *might not have occurred* (in fact, had a high probability of not occurring), and consequently "leaves room for an additional quest for explanation." Salmon's "explanation" simply does not explain, according to von Wright. A consequence of

---

<sup>2</sup> He writes, "It is enough to be able to assign the actual probability" (1984, 120).

Salmon's view is that, assuming indeterminism, some scientific explanations will be irreducibly statistical. A consequence of von Wright's view is that, assuming indeterminism, some events will not be *explainable* by science, even though they might very well be predictable to some extent (e.g., we won't be able to predict that *this* atom will undergo radioactive decay in the given time period, but we will be able to predict that approximately  $x$  many atoms from a sample will decay in the time period).<sup>3</sup> Which intuitions about explanation do we share? Which view of explanation should we accept?

Salmon tells us we should reject the modal conception of scientific explanation. "Since I find the consequences unacceptable, I think we must reject the modal conception" (1984, 111). Salmon finds it unacceptable to say that, if indeterminism is true, there are events that occur but are unexplainable by science. Why is this unacceptable? He tells us:

Quite possibly the greatest success story of modern science, with respect to its capacity for explaining a wide range of phenomena, is that of quantum mechanics. Under the most widely accepted interpretation, quantum mechanics is a statistical theory, and the explanations it provides are statistical explanations. The modal conception cannot survive the transition from the deterministic context to the indeterministic context. (*ibid.*)

It should be immediately clear how this response begs the question the advocate of the modal conception raises. In effect, Salmon is saying that since the modal conception rules out the possibility of statistical explanations, and since there possibly are statistical explanations (in quantum mechanics), the modal conception is wrong. Salmon asserts in this passage that quantum mechanics, under the most widely accepted interpretation,

---

<sup>3</sup> It is important to notice that von Wright's position does *not* rule out Hempel's D-S explanation or Railton's (1978) D-N-P explanation.

offers statistical explanations. The modal theorist, however, rejects that interpretation, rejects the claim that quantum mechanics offers statistical explanations of particular events.

Is Salmon trying to resolve a philosophical dispute with a show of hands? Is the “most widely accepted interpretation” automatically to be judged the correct interpretation? It’s not even clear that Salmon’s interpretation *is* the most widely accepted interpretation. There are many theorists, even major figures in the development of quantum mechanics, who seem to say that quantum physics, despite giving us very powerful tools for prediction and control, presents us with an utter mystery; it does not *explain* how the world can be the way it is, although it is very successful in telling us what the world is like.<sup>4</sup> Salmon himself cites an advocate of such a view (1984, 248; see also the quote from McCalla that Salmon gives on p. 8). Bernard d’Espagnat writes that “it is not entirely clear what physical interpretation should be given to the wave function or what picture of the electron one should keep in mind. Because of ambiguities such as this many physicists find it most sensible to regard quantum mechanics as merely a set of rules that prescribe the outcome of experiments” (1979, 158). Even if Salmon’s interpretation is the most widely accepted, Salmon presents no evidence of this, and it

---

<sup>4</sup> Niels Bohr has said, “Anyone who is not shocked by quantum theory has not understood it,” suggesting at least that it is not a paradigm of explanatoriness. Richard Feynman claims, “I think I can safely say that nobody understands quantum mechanics” (1965, 129). “[W]hile I am describing to you *how* Nature works [in the realm of quantum electrodynamics], you won’t understand *why* Nature works that way. But you see, nobody understands that. I can’t explain why Nature behaves in this peculiar way.... [T]here are no good theories to explain that” (1985, 10 and 12).



seems to me to be irrelevant to the question of whether this interpretation is in any sense *correct*.

I will agree that, according to the most widely accepted interpretation of quantum mechanics, the evidence suggests that the world is indeterministic.<sup>5</sup> This does not by itself support Salmon's claim that in an indeterministic world there will be adequate but irreducibly statistical *explanations* of particular events. The possibility of indeterminism merely raises the question, Does science in an indeterministic world offer irreducibly statistical explanations, or does science in such a world fail to explain any particular (indeterministic) event, explaining only assigned probabilities for outcomes and statistical laws?

In Salmon's most extended treatment of the modal conception of scientific explanation (pp. 111-120), he clearly tries to make the advocates of the modal conception feel that they do not adequately appreciate "the greatest success story of modern science." He virtually claims that their view of explanation is "a relic of a bygone age," borrowing a colorful turn of phrase from Bertrand Russell (see Salmon 1984, 120). Salmon apparently thinks that advocates of the modal conception have missed the boat, that they are stuck in some thoroughly outdated conception of science, happily pondering the elliptical orbits of the various planets, oblivious to the wave function, electron

---

<sup>5</sup> I remain much more doubtful than Salmon that the question of determinism vs. indeterminism can be answered decisively by empirical means. Remember Einstein and "God does not play dice with the universe." It is *always* possible, it seems to me, to maintain a "hidden-variables" story, to maintain that the appearance of indeterminism is merely the product of our current position of partial ignorance. Such a story might nonetheless become quite uncomfortable, even embarrassing, for relative, context-dependent reasons.

scattering, the two-slit experiment, and so on. Such reference to a bygone age as contrasted to the latest trends of thought is completely out of place, however, for it completely misses the point of dispute.

The modal theorist can praise just as heartily as Salmon the achievements of quantum mechanics. These achievements are achievements of prediction and control, however, and they do *not* suffice to answer the question of whether quantum mechanics is adequately *explaining* quantum phenomena. Salmon apparently thinks they do (he uses them that way, anyway). He seems to take it as obvious that, assuming indeterminism, modern science produces adequate statistical explanations. His theory of explanation can account for these, and the modal conception cannot. Therefore, he says, the modal conception is wrong and his theory is confirmed to an extent.

We must ask ourselves what could prevent someone from maintaining that quantum mechanics is a very powerful tool but explains nothing. Nothing could prevent someone from saying that, unless they take it as axiomatic that what science says always *is* explanatory. What could *recommend* the view that quantum mechanics is explanatory? A desire to account for the sense of *understanding* one gets from learning quantum mechanics? This seems to be Salmon's reason for taking quantum mechanics as explanatory; it appears that Salmon's criterion for what counts as an explanation is the sense of *understanding*.

Salmon is definitely not alone in assuming some kind of connection between explanation and *understanding*. Peter Achinstein has argued that we should think of explanation primarily as a speech act in which one person explains something to someone

else by uttering certain sentences, and a necessary condition of the occurrence of an act of explanation is that the person giving the explanation intend to impart a sense of understanding. He writes,

The first condition [for performing an explanatory speech act] expresses what I take to be a fundamental relationship between explaining and understanding. It is that *S* explains *q* by uttering *u* only if

(1) *S* utters *u* with the intention that his utterance of *u* render *q* understandable.

This expresses the central point of *S*'s act. It is the most important feature which distinguishes explaining from other illocutionary acts, even ones that can have indirect questions as objects. (Achinstein 1983, 16).

While developing his theory of singular causal explanation, James Woodward says that such a theory "ought to identify the structural features of such explanation which function so as to produce understanding in the ordinary user" (1984, 249). David Lewis claims that in an act of explaining, one who possesses explanatory information tries to convey it to someone else, and a condition of that person's coming to possess the explanatory information, or simply the explanation, is that he understand what he is told (1986, 185). Peter Railton connects explanation and understanding, as I believe Salmon would, with reference to quantum theory: "The deepening and widening of scientific knowledge under quantum theory gives the lie to those who prophesied that explanation by chance would be the enemy of enlarged inquiry and understanding" (1981, 181). Philip Kitcher claims, "A theory of explanation should show us *how* scientific explanation advances our understanding" (1981, 168). Michael Friedman declares that "the central problem of the theory of scientific explanation comes down to this: what is the relation between phenomena in virtue of which one phenomenon can constitute an explanation of another,

and what is it about this relation that gives understanding of the explained phenomenon?" (1974, 189). Paul Moser defends the position that "a proposition, *P*, is an explanation for one of one's subjective nonconceptual contents, *C*, if and only if *P* explains *C* and one understands *P*" (1989, 95). Finally, Robert Prevost explicates and defends a type of explanation that is actually defined in terms of its generating understanding. He describes a "concept [of explanation] more broadly conceived as making a state of affairs intelligible," and calls it "integrative explanation" (1990, 4-5).

Carl Hempel also has commented on the connection between explanation and understanding: "[T]he [D-N] argument shows that, given the particular circumstances and the laws in question, the occurrence of the phenomenon *was to be expected*; and it is in this sense that the explanation enables us to *understand why* the phenomenon occurred" (1965, 337, author's emphasis). At the same time, however, he tries to drive a wedge between considerations of understanding and the task of analyzing the logical structure of explanation. He explicitly claims to be concerned only with the "logical aspect[s] of explanation," adding that "such expressions as 'realm of understanding' and 'comprehensible' do not belong to the vocabulary of logic, for they refer to the psychological or pragmatic aspects of explanation" (1965, 413). Even in "Studies," he takes care to warn his readers against giving too much weight to the psychological, subjective sense of understanding:

Another aspect that lends appeal to teleological considerations is their anthropomorphic character. A teleological explanation tends to make us feel that we really "understand" the phenomenon in question, because it is accounted for in terms of purposes, with which we are familiar from our own experience of purposive behavior. But it is important to distinguish here understanding in the psychological sense of a feeling of empathic familiarity from understanding in the

theoretical, or cognitive, sense of exhibiting the phenomenon to be explained as a special case of some general regularity. The frequent insistence that explanation means the reduction of something unfamiliar to ideas or experiences already familiar to us is indeed misleading. For while some scientific explanations do have this psychological effect, it is by no means universal.... (1948, 256-257)

We can grant explanations that do not accomplish a reduction to something familiar, but is it possible that there be a scientific explanation that produces no psychological sense of understanding whatsoever in any of the (let's say) very many psychologically healthy and intelligent people who think about it? If there is a wedge to be driven between explanation and understanding, it should not be driven that far. But how far does Hempel want to drive it? How far should we drive it? What exactly is the relationship between explanation and understanding?

Hempel's comments are not fully clear, but he seems to suggest a position I find untenable. When Hempel says that the theoretical, or cognitive, sense of 'understanding' consists of exhibiting the phenomenon as a special case of some general regularity, he appears to be using his theory of what scientific explanation is to dictate where there is scientific understanding (i.e., understanding in the theoretical, or cognitive, sense). But then it is left as a mystery how Hempel determined in the first place what are and are not instances of scientific explanation, the very things he set out to "model." We need to ask what distinguishes an *explanatory* scientific account of something from a *merely descriptive* scientific account of something.

For example, in section 2 of "Studies in the Logic of Explanation" Hempel and Oppenheim provide "some illustrations," one involving the behavior of a mercury thermometer, one involving the refraction of light, and one involving Galileo's law for

the free fall of bodies. There is, in this context of inquiry, no dispute that these are scientific accounts. But how, one is led to ask, do Hempel and Oppenheim determine that these are explanatory rather than merely descriptive scientific accounts? The question takes on special strength when we remember that some Logical Positivists had not long before Hempel and Oppenheim's article been saying that there are *no explanatory* scientific accounts. Shall we let a theory dictate for us what things are (scientifically) understood, disregarding the subjective, psychological sense of understanding? Or shall we use a communal sense of understanding to bolster *counterexamples* to theories of explanation, identifying instances of explanation that do not fit the model or cases that fit the model but are not explanations?

Hempel's comments here, as in many other places, have the ring of stipulations. Certain Positivists had been saying that the function of science does not include giving explanations but consists of "describing phenomena in terms of laws," to use Carnap's phrase. Now Hempel comes along and pronounces that a certain type of description in terms of laws just is *scientific explanation* in terms of laws. He does not give an account of why we should think the arguments he describes are explanations rather than mere descriptions in terms of laws (or merely potential predictions, or potential justifications for belief, etc.). One gets the distinct impression that Hempel is declaring, "Let us henceforth mean *this* by 'scientific explanation'." He seems to be stipulating a meaning for 'explanation' in the restricted context of science. In doing so, he seems also to be stipulating a meaning for 'scientific understanding', asserting that where there is scientific

explanation so also shall there be scientific understanding, understanding “in the theoretical, or cognitive, sense.”

This is indeed objectionable. The problem is simple: one *may not* find sole support for a theory in particular examples while finding sole justification for the identification (involving *categorization*) of those particular examples in the same theory. Such a method exhibits a type of circularity that eliminates the possibility of cogency for that particular method. If the theory is not supported on the basis of particular examples but rather determines particular examples, the theory must have some other support, but Hempel provides *none*. I *suspect* that Hempel is motivated primarily by an overwhelming desire to bolster the claim that *science is explanatory*. I also suspect that a similar desire pushes Wesley Salmon to insist that *quantum mechanics is explanatory*. And I think they are both committed strongly to those claims because of their personal, subjective sense of getting increased *understanding* of the world from science. Despite the fact that Hempel’s comments are intended to isolate explanation from considerations of a subjective sense of understanding, both the first quote from him given above and his choice of examples, or “illustrations,” indicate his reliance on a subjective sense of understanding in identifying scientific accounts that are explanatory.

The accounts of the behavior of the mercury thermometer, the appearance of objects partially submerged in water, and Galileo’s laws do indeed give one the sense of greater understanding. Not only do these examples appear to be chosen for their efficacy in giving one a sense of understanding, they also appear to be chosen for dealing with phenomena that appear particularly puzzling, resistant to attempts to understand them,

prior to the scientific account. A thermometer submerged in hot water initially goes *down* slightly, when we would expect, of course, that it go *up*. Objects partially submerged in water appear bent, when we know that they are not bent and hence initially expect them to appear straight. Galileo's law of free falling bodies seems merely descriptive, and it goes directly against the initial expectation, widespread before Galileo, that heavier bodies fall faster than lighter bodies. Hempel seems, all things considered, to identify cases of explanation in practice initially just as the rest of us do, *via* the subjective sense of understanding.

### The Subjective Sense of Understanding

Before I examine the relationship between explanation and understanding, I should clarify what I mean by "understanding." We must first distinguish understanding (or, for emphasis, genuine understanding) from the mere feeling of understanding ("sham understanding"). It is important to notice, however, that one can make this distinction only on the basis of some prior distinction between what is genuinely explanatory and what is not. Someone might claim to understand why a particular political assassination occurred, for example, on the basis of some astrological or numerological account that would have predicted this event or something of the kind (perhaps "some grave misfortune for a prominent person"). This person would report having the distinctive psychological sense of gaining an increased understanding of why the event occurred as the result of hearing the astrological or numerological account. We insist, of course, that



there is no genuine understanding of the event in this case. It is merely sham understanding because it is not based on a genuine explanation.

There are two possible bases for rejecting the proposed explanation. The first is the claim that it is not true, or not wholly true. That is, it contains at least one proposition essential for the explanation that is false. The second is the claim that, even if the account contains only true statements, it is not explanatory. That is, it does not present *explanatory* facts, even if it does present *facts*. If we reject the explanation for the first reason, we make only the very familiar claim that someone is mistaken about the facts. One can bring this sort of criticism against any sort of proposed explanation, even potential explanations meeting the highest standards of scientific explanation. It is not significantly different from any other charge of falsehood, at least for our purposes here. The second reason for rejecting the proposed explanation has special bearing on the question of what it is to be explanatory. It depends on prior criteria for being either a potential or a true explanation. In this sort of case, the charge that there is no genuine understanding clearly depends on a prior distinction between the explanatory (including the potentially explanatory) and the non-explanatory. It thus depends on some sort of theory of explanation.

What I shall argue in the next section, however, is that understanding is *prior* to explanation in the sense that we must use a subjective psychological sense of gaining increased understanding as an initial guide to what is at least potentially explanatory. It is not the only determinant of what counts as explanatory, for a feeling of understanding can sometimes be dismissed as mere sham understanding on the basis of an established

theory of explanation. I shall argue that it is, however, indispensable as a *prima facie* indicator of what is potentially explanatory. Before we have a theory of explanation, we rely on our sense of understanding to point out individual *cases* of explanation. After we have a theory of explanation, we still sometimes use a sense of understanding (or lack of understanding) to buttress *counterexamples* to a theory of explanation. In order to use the case of the flagpole and its shadow to demonstrate weaknesses in Hempel's D-N account of explanation, for example, we must have some independent way of arguing that the length of the shadow *does not* explain the height of the pole (and that the height of the pole *does* explain the length of the shadow). This independent argument is based on our sense of understanding the length of the shadow on the basis of the height of the pole and our lack of understanding the height of the pole on the basis of the length of the shadow.

What is this subjective psychological sense of gaining increased understanding? In lieu of a complete psychological account of this subjective experience, it is best to define it by ostension. We can point out paradigm cases that increase understanding and paradigm cases that generate no increase in understanding. One of my favorite explanations is the explanation of why clocks run clockwise, that is, why the hands go to the right rather than to the left from an initial position of straight up.<sup>6</sup> If someone

---

<sup>6</sup> Asking why clocks run clockwise is ambiguous, but it has a dramatic impact. One might want to explain that "clockwise" is defined as the direction the hands move on a clock, so that it is an analytic truth that clocks run clockwise. What I am asking here is better expressed, "Why do the hands on clocks move in the direction they do rather than in the opposite direction?" This example comes from *Why Do Clocks Run Clockwise?*, by David Feldman (New York: Harper & Row, 1987). This is one book in David Feldman's *Imponderables* series of books. Incidentally, not all clocks run

offered as an explanation merely the claim that the standard clock is divided into twelve equal intervals, we would not feel any sense of gaining an increased understanding of why clocks run clockwise. The claim is relevant in its topic. It tells us something about clocks, and what it tells us is true. However, since we feel no increase in understanding (and, even more, see no way that this information when coupled with some other background information would produce increased understanding), we deny initially at least that this information is either explanatory or potentially explanatory. It is true and relevant to our topic, but it is not relevant in an *explanatory* way.

By contrast, when someone points out that clocks were first developed in the Northern Hemisphere, that the precursor to the mechanical clock as a timekeeping device was the sundial, and that the shadow on a sundial in the Northern Hemisphere sweeps out in the same direction as the hands on clocks now do, we feel (I presume) that this information is not only topically relevant but also explanatorily relevant. We designate this account as being at least potentially explanatory because it makes us feel that we understand why clocks run clockwise. There is a feeling that this account at least *could* be the true explanation of our explanandum, a feeling that was missing in connection with the previous information about the twelve divisions on a clock.

It is important to notice that we do not have to be convinced that this is *the* explanation of why clocks run clockwise in order to identify it as potentially explanatory

---

clockwise. Some clocks are designed for use in barbershops, and they run backwards so that customers in the chairs, viewing them in a mirror, will see them moving in a normal clockwise motion (Feldman, *Why Do Dogs Have Wet Noses?* [New York: HarperCollins, 1990], p. 225).

(explanatory-if-true). It might be that the first person who designed a clock lived in the Southern Hemisphere or near the equator. In that case, we would still say that the account could have been the explanation if it had been true. It might also be that the account is fully true (clocks were invented in the Northern Hemisphere), although it does not present the true *reasons* for clocks running clockwise. The real explanation might be that the first designer of a clock was right-handed and that right-handed people (whether in the Northern Hemisphere or the Southern Hemisphere) find “clockwise” motion more natural when they are given a choice of how to rotate something. This designer might not have been influenced by sundials at all. This is a *competing* explanatory account. If we lack evidence to eliminate either account as being false, we will continue to hold both as potential explanations of why clocks run clockwise. We hold both as potential explanations because each one exhibits explanatory relevance capable of generating the feeling of having increased understanding of the explanandum. This feeling was lacking in the first account.

The subjective psychological sense of gaining increased understanding is, I claim, present in the second and third accounts given above and lacking from the first. If some readers consider these examples and still insist that they do not see the relevant difference I am trying to point out, I do not know how else to identify the subjective sense I need to identify. I would then be in a situation analogous to the situation of a person trying to describe the subjective experience of green to a red-green color-blind person. If I point to a red square and to a green square and claim that they are obviously different, and if my companion claims not to notice any difference, I am at a loss to explain the

difference as it appears to me. I can explain some of the physics of color and some of the physiology of color vision. I will not, however, be able to explain the distinctive, subjective, qualitative experience of the color green. People with color vision learn how to identify green by ostension, and that depends on a distinctive experience of the right sort. Likewise, I do not know how to identify the distinctive subjective experience of getting a sense of increased understanding of something except by pointing out cases that produce it and cases that do not produce it and claiming that there is a marked difference between such cases. So I shall hope that I have identified the relevant subjective psychological experience well enough to bring most of my readers (i.e., at least two of the three) along.

I shall argue in the next section that this subjective sense of understanding is indispensable in the identification of explanations. Before one can begin to examine cases of explanation in search of the salient features that make them *explanatory*, one must establish that one in fact has a case of explanation (or potential explanation). The theorist of explanation who studies the first account I gave (citing the twelve divisions on the clock) in search of the salient features of explanation deserves our pity. He will not find them. He might come to believe he has found them, but he will be mistaken. He cannot find the salient features of explanation, because he is not looking in the right place. How do I know he's not looking in the right place? After all, he might construct a theory of explanation that qualifies our first account as explanatory. I know he's looking in the wrong place because I have no idea of how that first account as it stands could ever produce the appropriate sense of understanding for me or for anyone else.

On the other hand, any theory of explanation that *disqualifies* the second and third accounts as potential explanations will most likely be rejected just for that reason. Those accounts should qualify as potential explanations because they produce the relevant sense of understanding in abundance.

Everything I have said here leaves room for the subsequent distinction between genuine understanding and sham understanding. If the sundial account is false, then it does not produce genuine understanding of why clocks run clockwise. It might still produce the sense of understanding in many people, but it will be sham understanding because it is based on falsehoods. The rejection of this sense of understanding is similar to a rejection of a claim to knowledge if the proposition believed turns out to be false. One might have a subjective experience strongly indicative of knowledge (e.g., a feeling of certainty) while nonetheless failing to meet the truth-condition of knowledge. At least one of the conditions of genuine understanding is independent of any subjective feeling; it is the condition of truth in the explanation producing the feeling of understanding.

If the “twelve divisions” account produces in someone the subjective sense of understanding, this too is sham understanding. This, however, presents a much more complicated case. An individual’s sense of understanding in this case will be corrected socially, through critical comparison with the subjective experiences of other people in the community. Social corrections of the “explanations” offered by astrology or numerology, for example, sometimes proceed not by attacking any false claims in the explanations themselves but rather by indicating that most people do not acknowledge any mechanism by which the positions of the stars and planets or numerical relationships

could bring about or influence the events they are evoked to explain. That is to say that the explanations are sometimes attacked not for containing explicit false statements but rather for employing a socially unacceptable explanatory strategy, or method. They are attacked as not qualifying as potential explanations by any generally accepted criteria for explanation. The conditions of such social correction of explanations are not fully clear, although it is clear that such corrections do occur.

### The Connection Between Explanation and Understanding

How then are explanation and understanding connected? Is providing understanding either necessary or sufficient for an explanation?

First, is understanding necessary *and* sufficient for explanation? If it were, we could say that explanation just is whatever affords understanding to someone and that careful consideration of an explanation always produces a sense of understanding. A consequence of this position would be the relativization of explanations to persons. Knowledge of certain facts conceivably could produce a sense of understanding of a certain phenomenon for someone while knowledge of the same facts produces no sense of understanding of the relevant phenomenon for another person. Those facts would then constitute an explanation of the phenomenon for the first person but not for the second. *If* we find such thorough relativization of explanation to persons anathema, we should reject the view that understanding is necessary and sufficient for explanation.

Salmon's statement, "We secure scientific understanding by providing scientific explanations," suggests a connection between understanding and explanation: we secure

understanding *only* by providing explanations. This means that producing a sense of understanding (either actually or counterfactually, we could add) is a sufficient condition for explanation. It may be that a set of statements is an explanation of something *for a person* if that person's believing those statements brings that person some sense of understanding. It seems, however, that one can get a sense of understanding from something other than an explanation. Understanding is not sufficient for explanation, because one can get a sense of understanding, as thorough as is humanly possible, from *false* beliefs. If the beliefs that bring understanding are false, the statements believed cannot constitute an explanation.

But do they necessarily constitute a *potential* explanation? It seems not. I think it's best to keep in mind the variety among human beings and their cognitive lives. It would be very difficult to say what can and cannot produce a sense of understanding in a person. An account might seem completely unrelated to an event-to-be-explained to one person while providing a sense of understanding and an apparent explanation to another person. What brings about a sense of understanding is too variable from person to person or from culture to culture to be a sure indicator of the presence of an explanation. It seems best to leave open the possibility that a person get a sense of understanding from an account that, if true, still would *not* be an adequate explanation.

But *why* does that seem best, *why* should we leave open that possibility? The justification for doing so is not objective in the sense that there is some objective feature of the world that *compels* us to admit that possibility. We could define 'explanation' in such a way that the sense of an increase in understanding for a person entails the



presence of an explanation, so that an explanation just is whatever promotes a sense of understanding in someone. I think, however, that most of us would not want to do that, because we have experienced people reporting that they understand something on the basis of an account that seems not to be an adequate explanation. First, there are cases where a person reports understanding a phenomenon on the basis of statements that we know to be false. They cannot have an explanation if some of their relevant beliefs are false, but they still might have a potential explanation. Second, there are cases where a person reports understanding a phenomenon on the basis of statements that are not even a potential explanation.

Although someone can claim to understand the occurrence of a political assassination on the basis of certain astronomical signs, our view of the world assures us that even if there were a correlation over a large period of time between political assassinations and certain astronomical signs this correlation would be *accidental* and thus could not be *lawful* or *explanatory*.<sup>7</sup> We cannot conceive of how astronomical events could figure in an explanation of a political assassination; our ontology excludes this possibility, for it tells us there just is no way the astronomical events and the assassination could be related in appropriate ways. Hence, the astronomical facts cannot figure even in a potential explanation of the assassination, although it is conceivable that someone with a significantly different view of the world and its ways, a significantly

---

<sup>7</sup> Even if there were a lawful connection based, say, on there being a common cause, the connection would still not be potentially explanatory.

different metaphysics, would feel a sense of understanding the assassination on the basis of the astronomical facts.<sup>8</sup>

It is not advisable, then, to consider providing a sense of understanding as a sufficient condition either for explanation or for potential explanation. Is it *necessary* that an explanation provide (either actually or subjunctively) a sense of understanding to everyone or anyone? Another possible reading of Salmon's statement suggests as much: we secure understanding *always* in providing explanations. In other words, if something is an explanation of *X*, then it produces a sense of understanding *X* in everyone who considers it. Again, we have considerable freedom in answering these questions. We might want to define 'explanation' in such a way that, if an account does not produce a sense of understanding for a person, then that account cannot be an explanation *for that person*. Once again, however, this does not seem advisable. Capacity to understand is too variable from person to person for a useful notion of explanation to require understanding. Some people, we recall, seem to have trouble understanding just about anything. So we draw analogies. Just as an answer to a problem in mathematics can be the correct answer even though some people cannot calculate it or understand it, so also a scientific account can explain an event even if most people cannot understand it.

---

<sup>8</sup> We might wish to say that their account would be explanatory if both it and all of their background metaphysical beliefs were true. Is this maybe a *potential* potential explanation? We should ask why we pay homage to one account as a Potential Explanation just because its explanans if true would explain, on the assumption of *our* (scientific) ontology, and downgrade some other account as thoroughly non-explanatory even though its explanans if true would explain, on the assumption of some other ontology or metaphysics.

For some reason, most of us typically require a degree of *objectivity* for explanatoriness, such that a few rare geniuses can *discover* an explanation for something, and they at least have a chance of being *correct* in thinking that they have learned the explanation although the vast majority of us dullards have no understanding of it. Folklore has it that soon after Einstein published his special theory of relativity, very few people could understand it. It seems clear, however, in retrospect, that Einstein's theory does in fact explain, say, the results of the Michelson-Morley experiment, regardless of who does and does not understand the explanation. We almost want to say that there can be explanations no one at present understands. We at least want to say that if *A* explains *B*, then *A* explains *B for everyone*, even for those who do not understand it. *A* just is *the explanation of B*.

So it seems advisable not to say that producing a sense of understanding is necessary for an explanation. This "advice" not to make producing a sense of understanding either a necessary or sufficient condition for explanation is hard to justify. Why can't we say that Einstein's theory just fails to explain *for those who do not understand it*, while it explains for those who do understand it? The people who don't understand it just haven't yet got an explanation; if they come to understand, then the theory *will become* an explanation for them. There is room for saying that what is objective in this situation is accurate description. Einstein's theory is an accurate description of some aspects of reality, regardless of who understands it. In fact, it was an accurate description long before Einstein ever thought of it. It is an explanation, however, only if and when someone comes to understand it and comes, by way of it, to

understand the facts it explains. It is an explanation only to those people, to others it is an accurate description (objectively, regardless of what they might think about it) but not an explanation.

Our usage of 'explanation' is quite fuzzy, it turns out. The conflicting statements in the previous paragraphs all have considerable intuitive appeal. We might feel comfortable enough in the conclusion that producing a sense of understanding is neither necessary for an explanation nor sufficient for a potential explanation, and want to leave it at that. We ought to notice, however, that although producing (actually or counterfactually) understanding in everyone is not a necessary condition for an account's being an explanation, it is difficult to think of any reason to call something an explanation that never has and never will produce any sense of understanding in any member of a very large and diverse group of intelligent, psychologically healthy people who have actually thought carefully about it. This holds true even if all hands agree that they are dealing with an accurate description (or a fact). It at least becomes a very pressing question for one who claims it is an explanation what exactly makes it *explanatory*. And although producing a sense of understanding for someone does not seem sufficient for a potential explanation, it is difficult to think of any reason for saying that an account that produces a significant sense of understanding for every member of a very large and diverse group of intelligent, psychologically healthy people who have thought carefully about it is nonetheless not even a potential explanation. It would at least take of lot of explaining in that case why that account is not even a potential explanation.

Consequently, it is clearly wrong to say that there is *no* connection between understanding and explanation. Another reason, which we've already mentioned, for thinking there is a connection between understanding and explanation is that one cannot defend one's theory of explanation with appeals to its adequacy to particular cases if one identifies particular cases only with the theory. We need some way of knowing how to identify particular explanations independently of a theory of explanation, and using a sense of understanding somehow as a guide seems the only plausible candidate to fill that need, and it does in fact seem to be at work in that capacity for all theorists of explanation. This is not merely a point about ordinary language or about the majority opinion among a certain group of theorists. It is rather a point about the presuppositions of the universally employed *method* in all theorizing about the nature of explanation. If someone called something an explanation despite the fact that very many intelligent (etc.) people have thought carefully about it and get absolutely no sense of understanding from it, we must ask what reason they have for calling it an explanation. If they reply that their theory of explanation entails that it is an explanation, we must ask why we should think it is an explanation just because their theory says so *rather* than think it is a *counterexample* to their theory. Just think of how one decides whether (a) the tower and its shadow is an adequate explanation *because Hempel's theory says it is*, or (b) the tower and its shadow is a *counterexample to Hempel's theory*. We take it as a given that deducing the height of the tower from the length of its shadow, the position of the sun, and the laws of the rectilinear propagation of light does not explain the tower's height,

and we take this as given apparently because we get no subjective sense that this deductive argument gives us *understanding* of why the tower is the height it is.

We could turn to the common distinction between acts of explaining (*P* explains *q* by uttering *u*) and the products of such acts (the account referred to by *u* in that act) for help here (see Achinstein 1983 and Bromberger 1965). Perhaps the apparently conflicting intuitions we have about explanation and understanding fall naturally into two camps, one appropriate for acts of explaining and the other appropriate for the product of explaining. Let me refer to the product of explaining simply as an *explanation* and the act of communicating it with the intention of making someone understand something *explaining*.

It seems that an account's *being an explanation* should be largely independent of whether anyone has thought about it or gains understanding from it; at least, this is the way many people, including philosophers and scientists, handle explanations. An account of the moon's (and sun's) gravitational effect on the oceans of the earth is an explanation of the tides, regardless of whether anyone understands the tides in terms of the gravitational effect of the moon. And an account of some astronomical facts is not an explanation of a political assassination, regardless of how many people feel that they understand the assassination by knowing those astronomical facts.

On the other hand, it seems that the process of *explaining* does depend in some important ways on understanding. As Achinstein says, it seems that a necessary condition of offering an explanation is the intention to impart a sense of understanding. We might even go further than Achinstein and say that, in some sense, failure to impart

understanding implies failure to explain; not only the *intention* to impart understanding but also *success* at imparting understanding is necessary for one to have explained something to someone. Perhaps one can succeed in conveying all the information that is contained in the explanation of *X*, but fail to explain *X* (to someone) because the recipient does not gain a sense of understanding from knowledge of the explanatory account.

We might feel, then, that an account can be an *explanation* of something although no one ever has or ever will (or even would) gain a sense of understanding by coming to know it. We might feel in addition that an account can fail to be even a potential *explanation* although every one of a great number of people who have thought about it gets an immense sense of increased understanding from it. It might also seem that creating a sense of understanding in someone is a necessary condition for *explaining* something to them, and that creating that sense of understanding by conveying an explanation is sufficient for explaining.

This is not satisfactory, however. There is nothing wrong with saying that I have explained something to you even though you have failed to understand it. It is quite normal to mean by 'explaining *q* to *S*' nothing more nor less than 'presenting the explanation of *q* to *S*.' This reduces questions of what an act of explaining is to questions of what an explanation is, and we run into the same difficulties as we did above in determining the connection between an account's being an *explanation* and its producing a sense of understanding in someone.

Peter Achinstein (1983, esp. ch. 2 and 3) has argued that the act of explaining is more basic than the explanation provided; the concept of an explanation can be analyzed only by using the concept of explaining, but the concept of explaining can be analyzed without using the concept of an explanation. This is wrong, however. He seems to base his conclusion on the correct observation that one can present the same information (the same account) either in an act of explaining or in some other act, such as criticizing or merely describing. He concludes from this that the explanatory act makes something an explanation, since the thing's being an explanation cannot make the act of presenting it an explanatory act. This conclusion does not follow. It can be that a certain account is an explanation of *X* even though explaining is only one among many things one is able to do in presenting that account to someone. There is no difficulty in saying that one may criticize *by* presenting an account that is, in fact, explanatory. At most, we can say that an explanation is an account with which one *would* be able under appropriate conditions to explain the relevant explanandum. This does not mean, however, that the notion of explanation is dependent on the notion of an act of explaining, only that they are related.

The notion of an explanation is very rich. We should not artificially restrict it, as so many have tried to do. Even when we deal only with scientific explanation, the notion is complex. Our urge to formulate necessary and sufficient conditions relating it to the subjective sense of understanding must go unsatisfied. It is uncomfortable for philosophers to accept this, but we must be satisfied merely with the claim that producing



the subjective sense of understanding is *relevant* to whether something is an explanation while being neither necessary nor sufficient.

In the particular context at hand, the main thing to realize is that the subjective sense of understanding is *absolutely indispensable* as a guide to which accounts qualify as explanatory and which count as merely descriptive. This is evident from reflection on the very distinction between description and explanation. Both a descriptive account of an event and an explanatory account consist of descriptive propositions. They present information about the event. No philosopher, as far as I know, has ever suggested that explanations are distinguished as a class by doing something other than stating how the world is. Explanations are not distinguished as a class in the way that normative claims, for example, are distinguished. Normative claims do something *other than* stating how the world is; they state how the world *should be* (in some sense). This characteristic marks them off from descriptive propositions. Explanations, on the other hand, cannot be distinguished from (non-explanatory) descriptions simply in terms of their intended function. Explanations just are descriptive accounts *of a certain sort*. This accounts for the possibility of someone listening to a purported explanation and responding, "Everything you've told me is quite correct, and I agree with it all, but you still haven't *explained* what needs explaining."

Explanations are descriptive accounts *of a certain sort*. The central task for any philosopher of explanation is to spell out what that sort is. Hempel tried to do it strictly in terms of logical structure, in terms of syntax and semantics. Most philosophers now think this attempt was a failure. Wesley Salmon, David Lewis, James Woodward, and

many others have claimed that the distinguishing mark of explanations is that they make causal claims having a particular sort of objective relevance for the explanandum. Philip Kitcher and Michael Friedman have said that it is the unification of disparate facts about the world that marks explanations, where unification is characterized objectively in terms of information. None of these philosophers, I claim, has given enough attention to the relevance of the *subjective* sense of understanding. They all strive for some objective (logical, linguistic, or metaphysical) criterion for explanation.

In support of my claim that the subjective sense of understanding is indispensable as a guide to what explanation is, I can only fall back on my firm conviction that any account, even if it fulfills the objective standards of any of the aforementioned philosophers, fails to be an explanation if it *could not* produce in anyone under any circumstances a sense of increased understanding. An account that met someone's objective standards but failed to show promise of bringing a sense of increased understanding to anyone would, *prima facie*, be a *counterexample* to the proposed theory of explanation. Again, the central method of counterexamplifying within theorizing about explanation depends on the function for the subjective sense of understanding that I have identified.

### Understanding and an Ontological Background

Wesley Salmon recognizes that the question of what distinguishes explanation and description is a fundamental question (1984, 124-133). The question is, "what sort of knowledge, over and above predictive and descriptive knowledge, is involved in

explanation?” (p. 124). Salmon reiterates the fundamental importance of understanding: “A detailed knowledge of the mechanisms may not be required for successful prediction; it is indispensable for the attainment of genuine scientific understanding” (p. 133).

The capacity for producing a sense of understanding distinguishes explanatory from merely predictive or descriptive knowledge. Salmon argues that knowledge of underlying mechanisms is essential for the appropriate sort of knowledge. It is too restrictive, however, to insist on knowledge of an underlying mechanism. We can and do gain understanding of a phenomenon sometimes without a knowledge of an underlying mechanism. It is enough in some cases to know *what brought the phenomenon about*, where knowing what brought it about can be expressed strictly in terms of observable events. There will have to be postulated *connections* between events, but there need not be any postulated unobservable entities that are connected. We can understand, and hence explain, the melting of ice resulting from an increase in temperature because of a postulated connection between temperature and the formation and melting of ice, even though we have no idea of the underlying molecular mechanisms of melting. In fact, it seems that we are always in such a situation, taking some connections as explanatorily basic even when we do give some description of underlying mechanisms. For whenever we provide an explanation in terms of underlying mechanisms, there must remain connections between things within those mechanisms that are crucial for bringing about phenomena but are taken (at present at least) as explanatorily basic, in the sense that they are not explained in terms of further underlying mechanisms. That further explanation always stands out ahead of us as the domain of future research. So it is not surprising

that in some explanations we do not even progress to the first level of underlying mechanism, but leave that in the domain of future research.

What is essential for understanding a phenomenon is an ontological or metaphysical view of the world in which phenomena occur, a view that includes connections between phenomena and an apparatus for bringing phenomena about. In this sense, explanation requires a metaphysical or ontological background. As Peter Railton has put it, “Theories broadly conceived, complete with fundamental notions about how nature works—corpuscularianism, action-at-a-distance theory, ether theory, atomic theory, elementary particle theory, the hoped-for unified field theory, etc.—not laws alone, are the touchstone in explanation” (1981, 169). It is consequently not surprising that there were latent metaphysical presuppositions in the classic empiricist analyses of explanation and that the discussions have lately moved into explicitly metaphysical topics. David-Hillel Ruben (1990) has said that explanation requires a metaphysical “backing” (p. 2) or metaphysical presuppositions (p. 35). Railton has said that “one’s background picture of the world is involved in one’s *conception* of explanation” (1989, 224).

Neither Railton nor Ruben has explained in these works why this should be so, i.e., why a notion of explanation typically is not metaphysically neutral. I believe that the key to the dependence of a notion of explanation on a metaphysical view is the necessity of a metaphysical view for *understanding*, and the connection between understanding and explanation. The connection between a conception of explanation and a metaphysical view is mediated by the subjective sense of understanding.

It is necessary to have some background metaphysical view of the world in order to understand or comprehend some event. Without a background view of the world, there is no possibility of stating what *brought about* some experience or some experienced event. There will be no meaningful connections between experiences. Understanding consists of setting an event within some meaningful pattern and exhibiting its connections to other events. One can do this only by employing some general inventory of what (kinds of things) the world contains and some general guidelines regarding how it behaves.

Differences in a general conception of the world and its ways account for differences in what people will admit as a potential explanation. The person who entertains astrological explanations probably has a metaphysical view of the world that includes the sorts of entities and connections between events appropriate to such explanations. The person who entertains full-blown teleological explanations of natural events (events not involving humans) most likely includes persons other than humans (e.g., God) in a general view of the world and its ways. The eliminative materialist bases a rejection of folk-psychological explanations of actions in part on a view of the world that excludes a distinct mental realm of existence. In general, any person will admit and exclude potential explanations on the basis of what can possibly provide a sense of understanding within the given framework of their own metaphysical view of the world. Accounts that rely on elements alien to one's metaphysical framework will generally not produce any sense of understanding or comprehension and hence will not count as explanations. This connection between explanation, understanding, and

metaphysics explains why recent debates about explanation have been forced back into a reliance on metaphysical assumptions about such things as theoretical entities, causation, and laws.

### The Threat of Circularity Again

In chapter 4 I argued that ontological reasoning, or argumentation for existence claims, relies essentially on assumptions about what explanation is and what makes one explanation better than another. I have argued in this chapter that a theory of explanation typically depends on a background ontology or metaphysics. We saw in the case of Leibniz's argument from contingency for the existence of God how this two-way path of influence can generate a justificatory circle that is philosophically vicious. We did not, however, wind up in a vicious circle in our explication of the two-way influence between explanation and ontology in an aristotelian style of philosophy. There might be a vicious circle, however, if we push the investigation further. Let's forego additional analysis of metaphysics on an aristotelian model and ask in general whether there is any way to justify both an ontology/metaphysics and an analysis of explanation that does not generate vicious justificatory circularity, either by resting ultimately on a virtuous circle or by escaping circularity altogether.

First, it is possible that a conception of explanation and a general view of the world and its ways exhibit a positive, progressive back-and-forth influence. As we refine our knowledge of the general make-up of the universe, we refine our conception of what it takes to explain events in the world, because we refine our knowledge of what

“bringing about” actually amounts to in the world. Further, as we develop a more precise conception of what it takes to provide explanations, we substitute ontological schemes that have more explanatory power for older, weaker ontological schemes.

Much of the development of science might be characterized according to this pattern. Our scientific knowledge of the world has a deep influence on our conception of scientific (and non-scientific) explanation. And our conception of explanation guides scientific research and theorizing. There is doubtless an intricate story of how, for example, many subatomic particles were postulated to explain the outcomes of various experiments, but investigation of these entities led scientists and philosophers to change their conception of explanation radically to accommodate indeterminism and allow probabilistic notions of causality. Explanatory, theoretical concerns led scientists to change the ontological picture of what is inside the atom, abandoning the “solar system” picture and the ordinary conception of a particle as being distinct from a wave. Then the new ontological picture of the atom led to the inclusion of new patterns of explanation of events inside, and outside, of the atom. We thus see a constructive reciprocal relationship between ontological claims and conceptions of explanation in the sciences.

The reciprocal pattern of influence presents no problems as long as there are deeper levels of ontological theorizing into which the discussion can progress for the purpose of justification. When we revise our conception of explanation in light of ontological claims, we need to have some justification for those ontological claims that does not depend essentially on the specific nature of that conception of explanation; it

must be defensible on the basis of a broader, more general conception of explanation that has itself some justification independent of the particular ontological claims at issue.

The reciprocal influence between ontology and explanation does not produce a virtuous circle, but a descending spiral. It spirals down into more fundamental ontological claims and more general conceptions of explanation. When we trace the reciprocal line of influence we do not return exactly to where we start. If we try to support statistical explanations of low-probability events with an appeal to probabilistic causality, for example, we should not support the notion of probabilistic causality merely by pointing out that statistical explanations rely on it. Rather we move to more general considerations of what explanation *is* to argue that allowing for probabilistic causality is the best way to explain a large number of experimental outcomes. A defense of the general conception of explanation at work here should then appeal to more fundamental, more general ontological assumptions. And the pattern repeats itself.

We have, then, a spiralling regress that presents the problem of accounting for the termination of the regress in an ultimate source of justification. So we should find some way to terminate the regress in something that derives its justification from outside of either specific conceptions of explanation or claims about the nature of the world. We can take a cue from the position, based on Hintikka's research, that Aristotle had certain substantive assumptions about what calls for explanation and what counts as adequate explanation that were supported by epistemological and logical methodology. In the case of Aristotle, there were also metaphysical assumptions of a Heraclitean nature at work in producing the distinctive explanatory need that Aristotle felt and tried to meet with his



metaphysics of substance. But we can ask whether there is a way to defend either a metaphysics or a conception of explanation without appeal to yet another metaphysics or conception of explanation.

The way that logical and epistemological positions seemed to function in generating the felt need for an explanation of the possibility of scientific knowledge (*epistēmē*) in the ancient Greeks suggests a way out. If we take a background of logical and epistemological categories as given, we can understand some fundamental notion of explanation as bringing the world of experience into *order* in terms of those categories, so as to eliminate, or at least minimize, conceptual conflicts such as derivable contradictions (e.g., Zeno's Paradoxes, or the conflicts raised by the Michelson-Morley experiment). We can call this *conceptual explanation*.<sup>9</sup> In this context, we can usefully describe it as a type of *proto-explanation*. It is the ultimate source of justification for the entire enterprise of theorizing about explanation and developing ontological theories.

A conceptual explanation, together with the conceptual scheme on which it is based, is not obviously in need of the same sort of justification as ontological or metaphysical claims or theories of explanation. It can be justified instrumentally, on the

---

<sup>9</sup> This notion of explanation bears some resemblance to what John W. Yolton calls "systemic explanation," describing it as "a more primitive, generic form of explanation of which scientific explanation is only a species" (1959, 194). See also Yolton 1958. This notion of conceptual explanation, or proto-explanation, also resembles a type of explanation Paul Churchland discusses in 1989, chapter 10. He claims that explanatory understanding and perceptual recognition are closely related cognitive processes, both involving "prototype activation." One can agree with his assimilation of explanatory understanding and perceptual recognition and agree with his characterization of this fundamental type of explanation without necessarily basing it on Churchland's connectionist theory of mind.

basis of its usefulness as a tool for coping with the world as it comes to us through experience. There is probably some evolutionary story of how a propensity to impose order on experience by seeking out regularities is selected because of its basic usefulness for coping successfully with an environment. This sort of explanation, conceptual explanation, does not depend on prior notions of causation or on a prior ontology, rather it produces an ontology of objects of experience from an imposition of order upon experience.

There is no virtuous circular relationship between concepts of explanation and ontology, because there is, when things work properly, no genuine circle at all. When circles occur, as they did with Leibniz's cosmological argument for the existence of God, they are always vicious in that they fail to provide justification for the relevant existence claims or for the particular standards of explanation involved. There is instead a regress of justification basing analyses of explanation on ontological grounds, basing those ontological claims on more fundamental notions of explanation, and basing those notions of explanation on more fundamental ontologies or metaphysical systems. The most basic type of support holding between a set of existence claims and a notion of explanation is the support for our simplest level of interpretation of experience in terms, for most of us, of a physical object ontology. That is, the most basic support is the support coming from the most basic form of explanation, which consists of ordering experience for the most part as experience *of* a stable external world.

Why can't we say that some very simple and general metaphysical view of the world is the proper starting point both for defending a conception of explanation and for

defending more specific, detailed ontological claims? Say, for example, we may start with a concept of self as a persistent, thinking something existing within something other than itself and of which it has experiences. This view of self and world needs some defense because there are alternative metaphysical systems consistent with experience. Philosophers have in fact elaborated such systems: some including a self (mind) without any mind-independent world, others including a world and experiences of it without any enduring self (mind). Metaphysical systems always have alternatives. Metaphysical claims must be defended *vis-à-vis* coherent alternatives. Experience itself is the only thing that cannot be coherently disputed.

Why can't we take as basic a conception of explanation more substantive than the mere ordering of experience recommended on the basis of its utility? Why can't we start, for example, with a claim about what explanation essentially is independent of our handling of experience? Such a claim would simply be another metaphysical claim in need of defense. It would be a claim about how the world is with respect to the relationship of explanation (i.e., what it is for one fact about the world to explain another fact about the world).

Proto-explanation, in the form of some sort of conceptual ordering of experience is not optional. Without it, there can be no life of the mind whatsoever; no knowledge, no language, no discussion of the problems of ontology or explanation. Without it, there is only chaos. Proto-explanation is the first necessary step on the way to the conceptual modelling of a world that can be understood to some extent through the construction of explanations.

### Conclusion

We have found that explanation has a fundamental role to play in our thinking about the world. It is not merely a pleasant by-product of scientific discovery; it is a basic part of scientific theorizing and progress. It is not something we pursue only after we have arrived at what we take to be a satisfactory descriptive account of the world around us; it shapes virtually every part of our thinking about the nature of the world we live in, *including* our attempts at accurate description of parts of the world.

It is of the utmost importance to recognize not only that our standards of explanation profoundly affect the inferences we draw about what the world is like, but also that our ontological view of the world and its ways exerts a powerful influence over our thinking about what in our experience needs to be explained and about how properly to explain it. There is a genuine danger of failing to justify existence claims if they are based solely on explanatory arguments that employ criteria of acceptable explanation that are themselves defensible only on the assumption of the truth of the existence claims standing in need of justification.

The ultimate source of justification for any ontological position lies in the power of some basic conceptual ordering of experience to make it meaningful, to make it intelligible, and to allow for successful dealings within the world of our ontological theorizing.

## REFERENCES

- Achinstein, Peter. 1983. *The nature of explanation*. New York and Oxford: Oxford University Press.
- \_\_\_\_\_. 1992. Inference to the best explanation: Or, who won the Mill-Whewell debate? *Studies in History and Philosophy of Science* 23: 349-364.
- Aquinas. 1948. *Introduction to Saint Thomas Aquinas*. Edited by Anton Pegis. New York: Random House.
- Aristotle. 1941. *The basic works of Aristotle*. Edited by Richard McKeon. New York: Random House.
- \_\_\_\_\_. 1952. *Metaphysics*. Translated by Richard Hope. New York: Columbia University Press.
- Arrington, Robert L. 1989. *Rationalism, realism, and relativism*. Ithaca: Cornell University Press.
- Ashbaugh, A. Freire. 1988. *Plato's theory of explanation: A study of the cosmological account in the Timaeus*. Albany, N.Y.: State University of New York Press.
- Ayer, A. J. 1973. *The central problems of philosophy*. London: Weidenfeld and Nicolson.
- Ayer, A. J., and F. C. Copleston. 1949. Logical positivism—a debate. In *Contemporary approaches to philosophy*, ed. Paul Moser and Dwayne Mulder, 141-170. New York: Macmillan, 1994.
- Boyd, Richard. 1984. The current status of scientific realism. In *Scientific realism*, ed. Jarrett Leplin, 41-82. Berkeley and Los Angeles: University of California Press.
- \_\_\_\_\_. 1991. Confirmation, semantics, and the interpretation of scientific theories: Introductory essay. In *The philosophy of science*, ed. Richard Boyd, Philip Gasper, and J. D. Trout, 3-35. Cambridge, Mass.: The MIT Press.

- Braithwaite, Richard Bevin. 1953. *Scientific explanation*. Cambridge: Cambridge University Press.
- Bridgman, P. W. 1927. *The logic of modern physics*. New York: Macmillan.
- Bromberger, Sylvain. 1965. An approach to explanation. In *On what we know we don't know*, 18-51. Chicago: University of Chicago Press, 1992.
- \_\_\_\_\_. 1966. Why-questions. In *On what we know we don't know*, 75-100. Chicago: University of Chicago Press, 1992.
- \_\_\_\_\_. 1992. *On what we know we don't know*. Chicago: University of Chicago Press.
- Brown, James Robert. 1988. Platonic explanation: Or, what abstract entities can do for you. *International Studies in the Philosophy of Science* 3: 51-67.
- Burt, E. A. 1932. *The metaphysical foundations of modern physical science*. Rev. ed. Atlantic Highlands, N.J.: Humanities Press.
- Butterfield, H. 1957. *The origins of modern science 1300-1800*. Rev. ed. New York: Macmillan.
- Carnap, Rudolf. 1928. Pseudoproblems in philosophy. In *The logical structure of the world*, 301-343. Translated by Rolf A. George. Berkeley and Los Angeles: University of California Press, 1967.
- \_\_\_\_\_. 1932. The elimination of metaphysics through logical analysis. In *Logical positivism*, ed. A. J. Ayer, 60-81. Translated by Arthur Pap. Glencoe, Illinois: The Free Press, 1959.
- \_\_\_\_\_. 1950. Empiricism, semantics, and ontology. *Revue Internationale de Philosophie* 4: 20-40.
- \_\_\_\_\_. 1962. *Logical foundations of probability*. 2nd ed. Chicago: University of Chicago Press.
- \_\_\_\_\_. 1966. *Philosophical foundations of physics*. Edited by Martin Gardner. New York: Basic Books.
- Carruthers, Peter. 1992. *Human knowledge and human nature*. New York and Oxford: Oxford University Press.

- Churchland, Paul M. 1989. *A neurocomputational perspective*. Cambridge, Mass.: The MIT Press.
- Coffa, J. Alberto. 1974. Hempel's ambiguity. *Synthese* 28: 141-163. Reprinted in *Explanation*, ed. David-Hillel Ruben, 56-77. New York and Oxford: Oxford University Press, 1993. Page references are to this reprint.
- \_\_\_\_\_. 1977. Probabilities: Reasonable or true? *Philosophy of Science* 44: 186-198.
- \_\_\_\_\_. 1991. *The semantic tradition from Kant to Carnap: To the Vienna station*. Cambridge: Cambridge University Press.
- Copleston, F. C. 1950. The possibility of metaphysics. *Proceedings of the Aristotelian Society* 50 (1949-1950): 65-82.
- Dawkins, Richard. 1995. *River out of Eden*. New York: BasicBooks.
- Dennett, Daniel C. 1978. Where am I? Chap. 17 in *Brainstorms*. Montgomery, VT: Bradford Books.
- d'Espagnat, Bernard. 1979. The quantum theory and reality. *Scientific American* 241, 5.
- Dray, William. 1961. *Laws and explanation in history*. New York and Oxford: Oxford University Press.
- Einstein, Albert. 1961. *Relativity: The special and the general theory*. New York: Bonanza Books.
- Feynman, Richard P. 1965. *The character of physical law*. Cambridge, Mass.: The MIT Press.
- \_\_\_\_\_. 1985. *QED: The strange theory of light and matter*. Princeton, New Jersey: Princeton University Press.
- Fodor, Jerry. 1975. *The language of thought*. New York: Cromwell.
- Frankfort, Henri, H. A. Frankfort, John A. Wilson, Thorkild Jacobsen, and William A. Irwin. 1946. *The intellectual adventure of ancient man*. Chicago: University of Chicago Press.
- Friedman, Michael. 1974. Explanation and scientific understanding. *Journal of Philosophy* 71: 5-19. Reprinted in *Theories of explanation*, ed. Joseph C. Pitt,

188-198. New York and Oxford: Oxford University Press, 1988. Page references are to this reprint.

\_\_\_\_\_. 1981. Theoretical explanation. In *Reduction, time and reality*, ed. Richard Healy, 1-16. Cambridge: Cambridge University Press.

Fumerton, R. A. 1980. Induction and reasoning to the best explanation. *Philosophy of Science* 47: 589-600.

Garfinkel, Alan. 1981. *Forms of explanation*. New Haven, Conn.: Yale University Press.

Glymour, Clark. 1984. Explanation and realism. In *Scientific realism*, ed. Jarrett Leplin, 173-192. Berkeley and Los Angeles: University of California Press.

Goldman, Alan H. 1988. *Empirical knowledge*. Berkeley and Los Angeles: University of California Press.

Goodman, Nelson. 1955. *Fact, fiction, and forecast*. Cambridge, Mass.: Harvard University Press. (2d ed., 1965; 3d ed., 1973; 4th ed., 1983)

Hacking, Ian. 1975. *The emergence of probability*. Cambridge: Cambridge University Press.

Hampshire, Stuart. 1951. *Spinoza*. Harmondsworth: Penguin.

Hare, R. M. 1963. *Freedom and reason*. Oxford: Clarendon Press.

Harman, Gilbert. 1965. Inference to the best explanation. *Philosophical Review* 74: 88-95.

\_\_\_\_\_. 1967. Quine on meaning and existence, I and II. *Review of Metaphysics* 21: 124-151, 343-367.

\_\_\_\_\_. 1968a. Enumerative induction as inference to the best explanation. *Journal of Philosophy* 65: 529-532.

\_\_\_\_\_. 1968b. Knowledge, inference, and explanation. *American Philosophical Quarterly* 5: 164-173.

\_\_\_\_\_. 1977. Ethics and observation. Chapter 1 of *The nature of morality*. New York and Oxford: Oxford University Press. Reprinted in *Essays on moral realism*, ed. Geoffrey Sayre-McCord, 119-124. Ithaca: Cornell University Press, 1988. Page references are to this reprint.



- Hempel, Carl G. 1962. Deductive-nomological vs. statistical explanation. In *Minnesota studies in the philosophy of science*, vol. III, ed. Herbert Feigl and Grover Maxwell, 98-169. Minneapolis: University of Minnesota Press.
- \_\_\_\_\_. 1965. Aspects of scientific explanation. In *Aspects of scientific explanation and other essays*, 331-496. New York: The Free Press.
- \_\_\_\_\_. 1968. Maximal specificity and lawlikeness in probabilistic explanation. *Philosophy of Science* 35: 116-133.
- Hempel, Carl G., and Paul Oppenheim. 1948. Studies in the logic of explanation. *Philosophy of Science* 15: 135-175. Reprinted with a Postscript in *Aspects of scientific explanation and other essays*, 245-295. New York: The Free Press, 1965. Page references are to this edition.
- Hintikka, Jaakko. 1962. *Cogito, ergo sum: Inference or performance*. *Philosophical Review* 71: 3-32. (Reprinted in Hintikka, *Knowledge and the known*. Dordrecht and Boston: D. Reidel, 1974.)
- \_\_\_\_\_. 1974. Time, truth, and knowledge in Aristotle and other greek philosophers. Chap. 3 in *Knowledge and the known*. Dordrecht and Boston: D. Reidel.
- Holton, Gerald, and Stephen G. Brush. 1973. *Introduction to concepts and theories in physical science*. 2d ed. Reading, Mass.: Addison-Wesley.
- Howson, Colin. 1988. On a recent argument for the impossibility of a statistical explanation of single events, and a defence of a modified form of Hempel's theory of statistical explanation. *Erkenntnis* 29: 113-24.
- Hughes, R. I. G. 1989. *The structure and interpretation of quantum mechanics*. Cambridge, Mass.: Harvard University Press.
- Humphreys, Paul. 1984. Explanation in philosophy and in science. In *Principles of philosophical reasoning*, ed. James H. Fetzer, 172-189. Totowa, N. J.: Rowman & Allanheld.
- Hussey, Edward. 1972. *The Presocratics*. London: Duckworth.
- Irwin, Terence. 1988. *Aristotle's first principles*. Oxford: Clarendon Press.
- Kirk, G. S., J. E. Raven, and M. Schofield. 1983. *The Presocratic philosophers*. 2d ed. Cambridge: Cambridge University Press.

- Kitcher, Philip. 1976. Explanation, conjunction, and unification. *Journal of Philosophy* 73: 207-212.
- \_\_\_\_\_. 1981. Explanatory unification. *Philosophy of Science* 48: 507-531. Reprinted in *Theories of explanation*, ed. Joseph C. Pitt, 167-187. New York and Oxford: Oxford University Press, 1988. Page references are to this reprint.
- \_\_\_\_\_. 1989. Explanatory unification and the causal structure of the world. In *Scientific explanation*, ed. Philip Kitcher and Wesley C. Salmon, 410-505. Minneapolis: University of Minnesota Press.
- Kuhn, Thomas. 1970. *The structure of scientific revolutions*. 2d ed. Chicago: University of Chicago Press.
- Kyburg, Henry E., Jr. 1970. Conjunctivitis. In *Induction, acceptance, and rational belief*, ed. Marshall Swain, 55-82. Dordrecht and Boston: D. Reidel.
- Leibniz. 1973. *Philosophical writings*. Edited by G. H. R. Parkinson. London: Dent.
- Leplin, Jarrett. 1984. Introduction. In *Scientific realism*, ed. Jarrett Leplin, 1-7. Berkeley and Los Angeles: University of California Press.
- Lewis, David. 1986. Causal explanation. In *Philosophical papers*, II, 214-240. New York and Oxford: Oxford University Press. Reprinted in *Explanation*, ed. D.-H. Ruben, 182-206. New York and Oxford: Oxford University Press, 1993. Page references are to this reprint.
- Lipton, Peter. 1991. *Inference to the best explanation*. London and New York: Routledge.
- Locke, John. 1975. *An essay concerning human understanding*. Edited by Peter H. Nidditch. Oxford: Clarendon Press.
- Lugg, Andrew. 1991. Review of Wesley C. Salmon, *Four decades of scientific explanation*. *Canadian Philosophical Reviews*, 68-69.
- Lycan, William. 1988. *Judgement and justification*. Cambridge: Cambridge University Press.
- Mackie, J. L. 1976. *Problems from Locke*. Oxford: Clarendon Press.
- \_\_\_\_\_. 1977. *Ethics: Inventing right and wrong*. London: Penguin.

- Mill, J. S. 1858. *A system of logic*. New York: Harper & Brothers.
- Miller, David L. 1946. The meaning of explanation. *Psychological Review* 53: 241-246.
- \_\_\_\_\_. 1947. Explanation versus description. *Philosophical Review* 56: 306-312.
- Moravcsik, Julius M. E. 1974. Aristotle on adequate explanations. *Synthese* 28: 3-17.
- Moser, Paul K. 1985. *Empirical justification*. Dordrecht and Boston: D. Reidel.
- \_\_\_\_\_. 1989. *Knowledge and evidence*. Cambridge: Cambridge University Press.
- \_\_\_\_\_. 1993. *Philosophy after objectivity*. New York and Oxford: Oxford University Press.
- Mulder, Dwayne Hudson. 1996. The existential assumptions of traditional logic. *History and Philosophy of Logic*.
- Nagel, Ernest. 1961. *The structure of science: Problems in the logic of scientific explanation*. New York: Harcourt, Brace & World.
- Nagel, Thomas. 1974. What is it like to be a bat? *Philosophical Review* 83: 435-450. Reprinted in *Mortal questions*, 165-180. Cambridge: Cambridge University Press, 1979. Page references are to this reprint.
- Nozick, Robert. 1981. *Philosophical explanations*. Cambridge, Mass.: Harvard University Press.
- Parfit, Derek. 1992. The puzzle of reality: Why does the universe exist? *Times Literary Supplement*, July 3, 1992.
- Passmore, John. 1961. *Philosophical reasoning*. New York: Basic Books, 1969.
- Pearson, Karl. 1957. *The grammar of science*. 3d ed. New York: Meridian Books.
- Plantinga, Alvin. 1993. *Warrant and proper function*. New York and Oxford: Oxford University Press.

- Plato. 1941. *The Republic of Plato*. Translated by Francis MacDonald Cornford. New York and Oxford: Oxford University Press.
- \_\_\_\_\_. 1961. *The collected dialogues of Plato*. Edited by Edith Hamilton and Huntington Cairns. Princeton: Princeton University Press.
- Popper, Karl. 1935. *Logik der Forschung*. Vienna: Springer.
- \_\_\_\_\_. 1945. *The open society and its enemies*. London: Hutchinson.
- \_\_\_\_\_. 1959. *The logic of scientific discovery*. New York: Basic Books. (English translation of Popper 1935.)
- \_\_\_\_\_. 1972. *Objective knowledge*. Rev. ed., 1979. Oxford: Clarendon Press.
- Prevost, Robert. 1990. *Probability and theistic explanation*. Oxford: Clarendon Press.
- Putnam, Hilary. 1978. What is realism? Lecture II of *Meaning and the moral sciences*, 18-33. London: Routledge. (Reprinted in *Scientific realism*, ed. Jarrett Leplin, 140-153. Berkeley and Los Angeles: University of California Press, 1984.)
- Quine, Willard Van Orman. 1960. *Word and object*. Cambridge, Mass.: The MIT Press.
- \_\_\_\_\_. 1966. *The ways of paradox and other essays*. New York: Random House.
- Railton, Peter. 1978. A deductive-nomological model of probabilistic explanation. *Philosophy of Science* 45: 206-226. (Reprinted in *Theories of explanation*, ed. Joseph Pitt, 119-135. New York and Oxford: Oxford University Press, 1988.)
- \_\_\_\_\_. 1981. Probability, explanation, and information. *Synthese* 48: 233-256. Reprinted in *Explanation*, ed. David Hillel-Ruben, 160-181. New York and Oxford: Oxford University Press, 1993. Page references are to this reprint.
- \_\_\_\_\_. 1986. Moral realism. *Philosophical Review* 95: 163-207.
- \_\_\_\_\_. 1989. Explanation and metaphysical controversy. In *Scientific explanation*, ed. Philip Kitcher and Wesley C. Salmon, 220-252. Minneapolis: University of Minnesota Press.

- Rawls, John. 1971. *A theory of justice*. Cambridge, Mass.: Harvard University Press.
- Redhead, Michael. 1990. Explanation. In *Explanation and its limits*, ed. Dudley Knowles, 135-154. Cambridge: Cambridge University Press.
- Ross, Sir David. 1949. *Aristotle*. 5th ed. London: Methuen.
- Ruben, David-Hillel. 1990. *Explaining explanation*. London and New York: Routledge.
- \_\_\_\_\_, ed. 1993. *Explanation*. New York and Oxford: Oxford University Press.
- Russell, Bertrand. 1912. *The problems of philosophy*. New York and Oxford: Oxford University Press.
- \_\_\_\_\_. 1918. The philosophy of logical atomism. In *Logic and knowledge: Essays 1901-1950*, ed. Robert Charles Marsh, 177-281. London: George Allen & Unwin, 1956.
- Russell, Bertrand, and F. C. Copleston. 1948. The existence of God—A debate. In *Reality in focus: Contemporary readings on metaphysics*, ed. Paul K. Moser, 405-417. Englewood Cliffs, NJ: Prentice-Hall, 1990.
- Salmon, Wesley C. 1970. Introduction. In *Zeno's paradoxes*, ed. W. C. Salmon. New York: Bobbs-Merrill.
- \_\_\_\_\_. 1980. *Space, time, and motion: A philosophical introduction*. Minneapolis: University of Minnesota Press.
- \_\_\_\_\_. 1984. *Scientific explanation and the causal structure of the world*. Princeton: Princeton University Press.
- \_\_\_\_\_. 1989. *Four decades of scientific explanation*. Minneapolis: University of Minnesota Press. Also included in *Scientific explanation*, ed. Philip Kitcher and Wesley Salmon, 3-219. Minneapolis: University of Minnesota Press, 1989.
- \_\_\_\_\_. 1993. The value of scientific understanding. *Philosophica* 51: 9-19.
- Sayre-McCord, Geoffrey. 1988a. The many moral realisms. In *Essays on moral realism*, ed. Geoffrey Sayre-McCord, pp. 1-23. Ithaca: Cornell University Press.

- \_\_\_\_\_. 1988b. Moral theory and explanatory impotence. In *Essays on moral realism*, ed. Geoffrey Sayre-McCord, pp. 256-281. Ithaca: Cornell University Press.
- Schlick, Moritz. 1932. Positivism and realism. In *Logical positivism*, ed. A. J. Ayer, 82-107. Translated by David Rynin. Glencoe, Illinois: The Free Press, 1959.
- Soames, Scott. 1989. Presupposition. In *Handbook of philosophical logic*. Vol. IV, *Topics in the philosophy of language*, ed. D. Gabbay and F. Guenther, 553-616. Dordrecht and Boston: D. Reidel.
- Stroud, Barry. 1968. Transcendental arguments. *Journal of Philosophy* 65: 241-56. Reprinted in *Kant on pure reason*, ed. Ralph C. S. Walker, 117-131. New York and Oxford: Oxford University Press, 1982. Page references are to this reprint.
- Sturgeon, Nicholas L. 1985. Moral explanations. In *Morality, reason and truth*, ed. David Copp and David Zimmerman, 49-78. Totowa, N. J.: Rowman and Allenheld. Reprinted in *Essays on moral realism*, ed. Geoffrey Sayre-McCord, 229-255. Ithaca: Cornell University Press, 1988. Page references are to this reprint.
- Swinburne, Richard. 1979. *The existence of God*. Oxford: Clarendon Press.
- Urmson, J. O. 1956. *Philosophical analysis*. Oxford: Clarendon Press.
- van Fraassen, Bas. 1980. *The scientific image*. Oxford: Clarendon Press.
- von Wright, Georg Henrik. 1971. *Explanation and understanding*. Ithaca: Cornell University Press.
- Watkins, John. 1984. *Science and scepticism*. Princeton: Princeton University Press.
- Weinberg, Steven. 1992. *Dreams of a final theory*. New York: Pantheon.
- Whewell, William. 1967. *The philosophy of the inductive sciences, founded upon their history*, vol. 2. New York: Johnson Reprint Corp.
- Wittgenstein, Ludwig. 1921. *Tractatus logico-philosophicus*. Translated by D. F. Pears and B. F. McGuinness. Introduction by Bertrand Russell. London: Routledge, 1961.

- \_\_\_\_\_. 1953. *Philosophical investigations*. Edited by G. E. M. Anscombe and R. Rhees. Translated by G. E. M. Anscombe. New York: Macmillan.
- \_\_\_\_\_. 1991. Philosophy. *Synthese* 87: 3-22. (Reprinted in *Contemporary approaches to philosophy*, ed. Paul Moser and Dwayne Mulder, 125-140. New York: Macmillan, 1994.)
- Woodward, James. 1984. A theory of singular causal explanation. *Erkenntnis* 21: 231-262. Reprinted in *Explanation*, ed. David-Hillel Ruben, 246-274. New York and Oxford: Oxford University Press, 1993. Page references are to this reprint.
- Yolton, John W. 1958. Philosophical and scientific explanation. *Journal of Philosophy* 55: 133-143.
- \_\_\_\_\_. 1959. Explanation. *British Journal for the Philosophy of Science* 10: 194-208.
- Zimmerman, David. 1985. Moral theory and explanatory necessity. In *Morality, Reason and Truth*, ed. David Copp and David Zimmerman, 79-103. Totowa, N.J.: Rowman & Littlefield, 1985.

## VITA

Dwayne Hudson Mulder was born in Grand Rapids, Michigan on July 4, 1965. He received a Bachelor of Arts degree in philosophy, with a cognate in mathematics and a German minor, from Calvin College in 1987. He began taking graduate courses at Loyola University of Chicago in 1988, and began full-time graduate studies under a graduate assistantship in the Fall of 1989. He was granted graduate assistantships for the following three years, and received the Master of Arts degree in philosophy in 1991. He was awarded a University Teaching Fellowship for 1992-93 and received a Schmitt Dissertation Fellowship for 1993-94.

In 1992 and 1993, Dwayne Mulder and Prof. Paul K. Moser co-edited an anthology, *Contemporary Approaches to Philosophy*, published in 1994 by Macmillan. In 1993, they published "Methodology and Ontology" in *Methodology and Science*, volume 26. In 1994, they published "Probability and Rational Decision-Making," in *Philosophical Papers*, volume 23. They published a review article on Alvin Plantinga's books, *Warrant: The Current Debate* and *Warrant and Proper Function*, in *Philosophical Books* in 1996. Mr. Mulder published "The Existential Assumptions of Traditional Logic" in *History and Philosophy of Logic* in 1996. In May, 1994, Mr. Mulder presented a colloquium paper, "Probability and Rational Decision-Making," at the annual meetings of the Central Division of the American Philosophical Association in Kansas



City, Missouri. In April, 1995, he presented a commentary on a colloquium paper at the annual meetings of the Pacific Division of the American Philosophical Association in San Francisco, California. He presented a paper entitled "Explanation and Understanding" to the Parkside Philosophical Society at the University of Wisconsin-Parkside in March, 1996. From 1991 to 1996, he has taught sections of Logic, Epistemology, Metaphysics, Philosophy of Science, Philosophy of Mind, Introductory Philosophy, and Business Ethics in the Philosophy Department of Loyola University Chicago, at Oakton Community College, and at the University of Wisconsin-Parkside.

## DISSERTATION APPROVAL SHEET

The dissertation submitted by Dwayne Hudson Mulder has been read and approved by the following committee:

Paul K. Moser, Ph.D., Director  
Professor of Philosophy  
Loyola University of Chicago

Suzanne Cunningham, Ph.D.  
Associate Professor of Philosophy  
Loyola University of Chicago

J. D. Trout, Ph.D.  
Associate Professor of Philosophy  
Loyola University of Chicago

The final copies have been examined by the director of the dissertation and the signature which appears below verifies the fact that any necessary changes have been incorporated and that the dissertation is now given final approval by the committee with reference to content and form.

The dissertation is, therefore, accepted in partial fulfillment of the requirements for the degree of Ph.D.

6-3-96

Date

Paul K. Moser

Director's Signature