



THE NATURE OF EXPLANATION

PETER ACHINSTEIN

New York Oxford
OXFORD UNIVERSITY PRESS
1983

The Nature of Explanation

THE NATURE OF EXPLANATION

PETER ACHINSTEIN

New York Oxford
OXFORD UNIVERSITY PRESS
1983

Copyright © 1983 by Oxford University Press, Inc.

Library of Congress Cataloging in Publication Data
Achinstein, Peter.

The nature of explanation.

Includes index.

1. Explanation (Philosophy) Addresses, essays,
lectures. I. Title.

BD237.A25 1983 121 82-22571

ISBN 0-19-509215-2

For my son Jonathan

Printing (last digit): 9 8 7 6 5 4 3 2 1

Printed in the United States of America

Preface

For more than a third of a century philosophical discussions of explanation have been dominated by various formal “models.” Included among these are the deductive and statistical models of Hempel and variants proposed by others, and the statistical model of Salmon. Yet there are well-known counterexamples to these models, and good reasons for supposing that they fail to provide even necessary conditions. The typical response is to add more conditions to the model or to modify others. But each new variation on the old theme is simply an invitation to philosophical sharpshooters to hit the mark with fresh counterexamples.

What this book presents is not another variation but a different theme. Unlike the standard theories, it focuses, to begin with, on the explaining act itself—the act in which by uttering or writing words someone explains something. From that act a “product” emerges: an explanation. To characterize what kind of entity that product is, as well as how it can be evaluated, essential reference must be made to the concept of an explaining act. Otherwise, I argue, we will be unable to distinguish explanations from products of non-explaining acts; and we will be unable to say why various explanations, particularly in the sciences, deserve praise or blame.

A theory of the explaining act, of the product and its ontological status, and of the evaluation of explanations, is presented in the second, third, and fourth chapters. Following this the theory is brought to bear on a number of issues: why have the standard models of explanation been unsuccessful (are they seeking the impossible)? What is a causal explanation, and must explanations in the sciences be causal? What sort of explanation is one that appeals to the function of something (is it, e.g., causal)? Are some things too fundamental to be explained, and what does that mean? What is the relationship between

explanation and evidence (can the latter concept be defined by reference to the former)? These issues, and others, will be approached by making essential use of the theory of explanation developed in the earlier chapters.

One trend in recent philosophy of science is to reject the formalism and precise schemas of logical positivism, from which the early models of explanation emerged. Although I too will not accept the tenets of positivism and its attendant picture of explanation, I will try to develop the theory of explanation in a precise way. This will necessitate some "formal" definitions. (Those who want to take their philosophy at a gallop are forewarned.) However, numerous examples are presented in order to make these definitions clear. Where appropriate, the examples are simple, everyday ones. In some cases—particularly in the discussion of the evaluation of scientific explanations, of causal explanations, and of the limits of explanation in science—I have chosen examples from physics.

The theory of explanation presented here had a forerunner in Chapter 4 of my book *Law and Explanation*, published in 1971. That account I no longer regard as adequate. However, in some of the chapters that follow I have made use of revised and expanded material from recent articles published in the *American Philosophical Quarterly* (vol. 14), *Midwest Studies in Philosophy* (vol. 4), *Mind* (vol. 76), and *Philosophy of Science* (vol. 44).

I am indebted to the National Science Foundation for research grants during the tenure of which this book was written. In the spring of 1976 when I served as Lady Davis Visiting Professor at the Hebrew University of Jerusalem I offered a graduate seminar on explanation, and began to think particularly about the problem of functional explanation discussed in Chapter 8. This and many other issues were also aired in an N.E.H. seminar for college teachers I directed in the summer of 1978 and in graduate seminars at Johns Hopkins. Participants in these seminars as well as persons from various universities in the U.S. and Israel where material from several chapters was read bestowed vigorous criticisms on my proposals. They are to be credited for my numerous attempts in what follows to counter objections.

Stimulating discussions with Jaegwon Kim, who read the entire manuscript, sharpened my thoughts, as did perceptive comments by George Wilson, Dale Gottlieb, and Barbara von Eckardt. These persons made important suggestions for improving the quality of the work and I am deeply indebted to them. Luckily for me some of their suggestions conflicted; otherwise I would still be revising.

Alan Berger raised questions about Chapter 2 that I had not previously considered. Robert Causey and Bas van Fraassen helped me better understand views of theirs which I treat. David Zaret, with whom

I taught a graduate seminar on evidence in the fall of 1981, offered criticisms which improved the last chapter. Robert Cummins and Daniel Garber, both lively philosophical conversationalists, gave me the opportunity to explore ideas over lunch at the Hopkins faculty club. And David Sachs furnished wise counsel when it was needed.

I want to express my thanks to Nancy Thompson, the Philosophy Department's chief secretary, for her generous help on many matters, to Cecelia Hrdlick and Molly Mitchell for their excellent typing, and to Vera Herst and Fred Kronz for assistance with the proofs.

To my wife Merle Ann, who shared with me the exhilarating as well as the difficult moments in the enterprise of explaining explaining, and who provided moral and physical support, I am forever grateful.

Baltimore
November 1982

P.A.

Contents

1. Introduction, 3
2. Explaining, 15
3. What Is an Explanation?, 74
4. The Evaluation of Explanations, 103
5. Can There Be a Model
of Scientific Explanation?, 157
6. The Causal Relation, 193
7. Causal Explanation, 218
8. Functional Explanation, 263
9. The Limits of Explanation, 291
10. Evidence and Explanation, 322
11. Evidence: Additional Topics, 351
- Index, 383

CHAPTER 1

Introduction

1. THE PROJECT

Suppose that a speaker *S* explains something *q* by uttering (or writing) a sentence *u*. For example, *S* explains why that metal expanded, by uttering

That metal expanded because it was heated.

An *act* of explaining has occurred, which took some time, however short. We might also speak of the *product* of this act—viz. the explanation given by *S*—which did not take any time, but was produced in or by the act of explaining. (“The explanation given by *S*” can be used to refer either to the act or to the product; in what follows it will be used for the product.) If *S* repeats the sentence above on different occasions when he explains this phenomenon, he has engaged in several explaining acts (he has explained several times), even though the product—his explanation—is the same on each occasion.

Various questions can be raised about explaining acts and products. The ones of concern to me in this chapter and the three that follow are quite general:

a. What is an explaining act? More particularly, can necessary and sufficient conditions be supplied for sentences of the following form?

(1) *S* explains *q* by uttering *u*.

b. What is the product of an explaining act? (What is an explanation?) Can necessary and sufficient conditions be supplied for sentences of the form

(2) *E* is an explanation of *q*.

c. *How should explanations (i.e., products) be evaluated?* Can necessary and sufficient conditions be supplied for sentences of the form

(3) E is a good explanation of q.

(Perhaps we will want to consider other terms as well for positive or negative evaluations.)

My aim is to develop a theory of explanation that will answer these questions, and in particular, propose informative conditions for sentences of forms (1), (2), and (3).

The terms "explain" and "explanation" can be used broadly to refer to explaining acts and products that may or may not be good (adequate, successful, "scientific"). They can also be used more narrowly to refer only to acts and products that are (regarded as) good. In accordance with the broader, but not the narrower, use an atheist could admit that his religious friends are explaining the origin of man when they assert that man was created by God. And he could refer to the product of such acts as an explanation. In what follows, the terms "explain" and "explanation" in (1) and (2) will be used in this broader way. The narrower, evaluative use will be considered when we turn to sentences of form (3).

The theory of explanation I propose begins with explaining acts, and in particular with sentences of form (1). These will be my concern in Chapter 2. Questions (b) and (c) will be addressed in Chapters 3 and 4, respectively. My thesis is that the concept of an explaining act is fundamental, and that the concept of an explanation (as product), and that of a good explanation, must be understood, in important ways, by reference to the former. Thus, the answers I will propose to questions (b) and (c) will depend on a prior consideration of question (a).

This is very different from the usual strategy of explanation theorists, which is to concentrate on question (c); to offer some views, or at least clues, about (b) but not dwell on this (since the concern is not to characterize explanations generally, but only good ones); and largely to ignore (a) by focusing on what its proponents call the "logic" of explanation rather than its "pragmatics." In the remainder of this chapter I shall describe some leading theories of this sort. My aim is not to try capturing all the details of these theories, but only their main outlines. In later chapters I will show that by developing views about the product of explanation, and about the evaluation of such products, that are independent of the concept of an explaining act these theories, and others like them, provide inadequate accounts.

2. ARISTOTLE'S DOCTRINE OF THE FOUR CAUSES

There are, Aristotle believed, four causes or determining factors which correspond to the meanings of the question "why": the material, formal, efficient, and final causes.¹ Aristotle is concerned with why-questions of the form

(1) Why does X have (property) P?

In *Metaphysics* Z he writes:

The "why" is always sought in this form—"why does one thing attach to some other?" . . . We are inquiring, then, why something is predicable of something. (1041a)

And in the *Physics*, where the doctrine is most fully developed, his examples of why-questions include "Why is he walking about?" and "Why did they go to war?"

Typical of the form of the answer to the why-question—i.e., the form of the explanation—Aristotle considers is "because r," or more fully,

(2) X has P because r.

("They went to war because they were raided.") It is Aristotle's view that the r-position will be filled by reference to various causes: the matter or constituents of which X is composed in virtue of which it has P (material cause), the form or structure of X (formal cause), an external source of motion or change by which X comes to have P (efficient cause), and "that for the sake of which" X has P (final cause).

Aristotle makes little attempt to define his four causes. He seems to treat them as primitives, that can be illustrated but not further analyzed. A complete explanation in physics, and perhaps in science generally, invokes causes of all four types:

Now, the causes being four, it is the business of the physicist to know about them all, and if he refers his problems back to all of them, he will assign the "why" in the way proper to his science—the matter, the form, the mover, "that for the sake of which." [*Physics*, 198a]

I follow Moravcsik in construing Aristotle's doctrine of the four causes as a doctrine about explanation.² (Perhaps, then, "explanatory factor" is a better term to use than "cause" to express Aristotle's thought.) However, I do not agree with Moravcsik's view that Aristotle selects the four basic causes he does solely because of his onto-

1. This is Aristotle's view in the *Physics* and *Metaphysics*. A more elaborate view in the *Posterior Analytics* will be noted in Section 5.

2. Julius M. E. Moravcsik, "Aristotle on Adequate Explanations," *Synthese* 28 (1974), pp. 3–17.

logical doctrine of substance (which, according to Moravcsik, commits Aristotle to there being these four kinds). Rather, I take Aristotle's view to be, in part, a semantical one concerning the meaning of the word "cause" (or "explanatory factor" or "explanation"). In the *Physics*, after introducing the doctrine, he writes:

This then perhaps exhausts the number of ways in which the term "cause" is used. (195a)

And in *Metaphysics* D (1013a):

Cause means (1) that from which, as immanent material, a thing comes into being. . . . [Aristotle continues with all the causes.]

Aristotle's view is that by an explanation (i.e., by something which gives a cause) we mean something which gives one or more of the causes he mentions. However, he does go on to make the ontological claim that the four types of causes exist in nature. His doctrine, then, seems to combine both semantical and ontological features, as follows. Each substance has both matter and form (ontological doctrine). But these are two senses of the word "cause" (semantical doctrine). So each substance has both a material and a formal cause. Furthermore, substances exhibit behavior for the sake of an end (ontological fact). But one sense of "cause" is teleological, i.e., for the sake of an end. Therefore, there exists behavior that has a teleological cause. Finally, substances do exhibit behavior produced by external sources. But one sense of "cause" is external, i.e., efficient, cause. So efficient causes exist. In each case, the conclusion that there is in nature a cause of the sort Aristotle cites is reached by means of two premises, one ontological and one semantical.

Turning now to a simple example, consider the question

Why did that metal expand?

Assume that someone offers the explanation

(3) That metal expanded because it was heated.

The explanation here (i.e., the product), Aristotle seems to be saying, is a proposition, viz. the proposition expressed by (3). This is an explanation because it is a proposition that purports to describe a cause—in this case, an efficient cause—of the metal's expansion.

More generally, if we consider questions of the form

(1) Why does X have (property) P?,

Aristotle's answer to question (b) of the previous section (What is the product of an explaining act, i.e., what is an explanation?) can perhaps be formulated as follows. Where q is an indirect question whose direct form is given by (1),

(4) E is an explanation of q if and only if E is a proposition of the form "X has P because r," in which r purports to give one or more of Aristotle's four causes of X's having P.

Aristotle's answer to question (c) of the previous section (How should explanations be evaluated?) might then be given as follows. Where q is an indirect question whose direct form is (1),

(5) E is a good explanation of q if and only if E is a proposition of the form "X has P because r," in which r correctly gives (and does not merely purport to give) one or more of Aristotle's four causes of X's having P. (Or perhaps Aristotle's view is the stronger one that E must give causes of all four types, if they exist.)

Aristotle's doctrine of the four causes does not concern itself with explaining acts. It does not address question (a) (What is an explaining act?). However, taking a cue from what has been said above, Aristotle might have wished to espouse the view that explaining acts consist in the uttering of sentences that express propositions that are explanations. That is, where q is an indirect question whose direct form is (1),

(6) S explains q by uttering u if and only if S utters u, and u expresses a proposition of the form "X has P because r" in which r purports to give one or more of Aristotle's four causes of X's having P.

For example, S explains why that metal expanded by uttering "that metal expanded because it was heated," since S utters a sentence that expresses a proposition of the form "X has P because r" in which r is replaced by a sentence purporting to give an efficient cause of the metal's expansion.

However, it should be emphasized, I am not here attributing (6) to Aristotle himself. The main point I want to stress is that Aristotle seems to be defending a view which divorces an account of explanations from that of explaining acts. On this view, one can understand the concept of an explanation independently of that of an explaining act. Neither (4) nor (5) invokes the latter concept. Indeed, if (6) were Aristotle's position on explaining acts, then his theory of such acts would depend on an independent account of explanations.

3. HEMPEL'S DEDUCTIVE-NOMOLOGICAL MODEL

The view that an explanation of a phenomenon brings it under a law is implicit in many writings, but its fullest and most influential expres-

sion is due to Hempel.³ According to this view, an explanation is an argument whose premises include laws and whose conclusion is a description of the phenomenon to be explained. Hempel distinguishes explanations of two types: deductive-nomological (D-N), and inductive-statistical. In the former, the premises entail the conclusion; in the latter, the premises make the conclusion probable without entailing it. Hempel is concerned with what he calls explanation-seeking why-questions.⁴ These have the form

Why is it the case that p?

in which "p" is replaced by a sentence. Let Q(p) be a question of this form which presupposes that p is true, and let q(p) be the indirect form of this question. Hempel's view can now be put as follows:

- (1) E is a potential D-N explanation of q(p) if and only if E is a valid deductive argument whose premises contain lawlike sentences that are actually used in the deduction and whose conclusion is p.⁵
- (2) E is a good (correct, scientific) D-N explanation of q(p) if and only if E is a potential D-N explanation of q(p) all of whose premises are true.

For example, let Q(p) be

Why is it the case that this metal expanded?

Consider the argument

- (3) This metal was heated
All metals expand when heated
Therefore,
This metal expanded.

By (1), (3) is a potential D-N explanation of why this metal expanded, since (3) is a valid deductive argument whose premises contain a lawlike sentence (the second premise) that is actually used in the deduction. If the premises of (3) are both true, then, by (2), (3) is also a good (correct, scientific) explanation of why this metal expanded.

Let us now turn to questions (a), (b), and (c) in Section 1 (What is an explaining act? What is an explanation? How should explanations be evaluated?). I take (1) to be an answer to (b), and (2) to be an answer to (c). An explanation, Hempel is urging, is an argument of a sort described in (1). (Or else it is a certain type of inductive argument

3. Carl G. Hempel, *Aspects of Scientific Explanation* (New York, 1965).

4. *Ibid.*, p. 334.

5. This is a simplification of Hempel's model; some additional conditions will be discussed in Chapter 5.

which also contains lawlike sentences; but I shall ignore the inductive cases in what follows.) A *good* explanation is an argument of the sort described in (1) in which all of the premises are true.

By contrast, Hempel fails to devote much attention to acts of explaining. He does recognize what he calls a "pragmatic" dimension of explanation. The use of the term "explanation" and its cognates "requires reference to the *persons involved in the process of explaining*."⁶ And in this pragmatic sense, what explains something for someone might not explain it for someone else, because of differences in beliefs, puzzlements, or intelligence. Although he does not explicitly say so, perhaps Hempel would suggest a condition such as the following for explaining acts:

- (4) S explains q by uttering u if and only if S utters u, and u expresses a deductive argument of the sort described in (1) (or a comparable inductive argument of the sort prescribed by the inductive-statistical model).

In accordance with this condition, if S utters (3) then S is explaining why this metal expanded, since (3) expresses a deductive argument of the sort described in (1). However, it should be stressed, Hempel does not devote himself to the project of providing conditions for explaining acts. Rather, he identifies his task as one of

constructing a nonpragmatic concept of scientific explanation—a concept which is abstracted, as it were, from the pragmatic one, and which does not require relativization with respect to questioning individuals any more than does the concept of mathematical proof. It is this nonpragmatic conception of explanation which the covering-law models are meant to explicate.⁷

Hempel believes that there is a concept of explanation which can be understood without reference to the idea of an explaining act. (1) and (2) above invoke no such idea. Indeed, if (4)—or something like it—were Hempel's account of explaining acts, then an understanding of explaining acts would depend on a prior understanding of explanations.

4. SALMON'S STATISTICAL RELEVANCE (S-R) MODEL

Like Hempel, Salmon stresses the need for laws in explanations.⁸ Unlike Hempel, he construes an explanation not as an argument but simply as a set of sentences. These sentences provide a basis for an

6. *Ibid.*, p. 425, my emphasis.

7. *Ibid.*, p. 426.

8. Wesley C. Salmon, *Statistical Explanation and Statistical Relevance* (Pittsburgh, 1971).

inference concerning the event to be explained; but Salmon, by contrast with Hempel, does not require that this inference be to the conclusion that the event definitely, or even probably, occurred. The explanation need provide only a basis for inferring with what probability (however small) it was to be expected to occur. Salmon writes:

An explanation does not show that the event was to be expected; it shows what sorts of expectations would have been reasonable and under what circumstances it was to be expected.⁹

Salmon's statistical relevance model embodies this idea. It is concerned with explanations which answer questions of the form

Why is X, which is a member of class A, a member of class B?

An explanation consists of a set of empirical probability laws relating classes A and B, together with a class inclusion sentence for X, as follows:

$$(1) \begin{aligned} p(B, A \& C_1) &= p_1 \\ p(B, A \& C_2) &= p_2 \end{aligned}$$

...

$$\begin{aligned} p(B, A \& C_n) &= p_n \\ X \in C_k \quad (1 \leq k \leq n) \end{aligned}$$

Salmon imposes two conditions on the explanation. One is that the probability values p_1, \dots, p_n all be different. The other is

The homogeneity condition: $A \& C_1, A \& C_2, \dots, A \& C_n$ is a partition of A, and each $A \& C_i$ is homogeneous with respect to B.

$A \& C_1, \dots, A \& C_n$ is a partition of A if and only if these sets comprise a set of mutually exclusive and exhaustive subsets of A. Set A is homogeneous with respect to B if and only if there is no way, even in principle, to effect a partition of A that is statistically relevant to B without already knowing which members of A are also members of B. (C is statistically relevant to B within A if and only if $p(B, A \& C) \neq p(B, A)$.) Intuitively, if A is homogeneous with respect to B, then A is a random class with respect to B.

Consider a simple example in which the explanatory question is

- (2) Why is this substance, which is a member of the class of metals, a member of the class of things that melt at 1083°C? (That is, why does this piece of metal melt at 1083°C?)

9. *Ibid.*, p. 79.

Letting

A = the class of metals

B = the class of things that melt at 1083°C

C_1 = the class of things that are copper

C_2 = the class of things that are not copper,

we can construct the following explanation for (2):

$$(3) \begin{aligned} p(B, A \& C_1) &= 1 \\ p(B, A \& C_2) &= 0 \\ X \in C_1. \end{aligned}$$

What this explanation tells us is that this substance, which is a metal, melts at 1083°C, because

the probability that something melts at 1083°C, given that it is copper, is 1; and

the probability that something melts at 1083°C, given that it is not copper, is 0; and

this substance is copper.

Salmon's two conditions are satisfied: the probability values in (3) are different; $A \& C_1$ and $A \& C_2$ form a partition of A and both these subclasses are homogeneous with respect to B.

Salmon's statistical relevance model is, I suggest, best construed as an attempt to answer question (c) of Section 1. It is an attempt to provide conditions for being a *good* (adequate, scientific) explanation. Perhaps Salmon would say that his model gives at least this much of an answer to question (b): An explanation is a set of sentences of form (1) containing probability laws and a class-inclusion sentence. However, the only necessary and sufficient conditions he proposes are ones for being a good explanation. Nor does Salmon address the question of what conditions are necessary and sufficient for an explaining act. Clearly, the conditions for explanations that he does formulate are independent of the concept of such an act.

5. BRODY'S TWO MODELS

A theory of explanation developed by Baruch Brody combines certain features of Hempel's D-N model with some ideas from Aristotle. Brody formulates two models, in both of which an explanation is a deductive argument satisfying the conditions of Hempel's D-N model. According to the first, the causal model, Hempel's D-N conditions must be supplemented by the

Causal condition: The premises of the argument must contain essentially a description of the event which is the cause of the event described in the conclusion.¹⁰

So, e.g.,

- (1) This metal was heated
All metals expand when heated
Therefore,
This metal expanded

is an explanation of the expansion, on Brody's causal model, since it is a deductive argument satisfying Hempel's D-N conditions plus the causal condition above. Its premises contain the sentence "this metal was heated," which is a description of the event that caused the event described in the conclusion.

Brody's second model, the essential property model, requires that Hempel's D-N conditions be supplemented by the

Essential property condition: The premises of the argument contain "a statement attributing to a certain class of objects a property had essentially by that class of objects (even if the statement does not say that they have it essentially) and . . . at least one object involved in the event described in the [conclusion] is a member of that class of objects."¹¹

In a later work Brody attempts to explicate the concept of an essential property by developing the idea that a has property P essentially if and only if there is no possible future in which a continues to exist but does not have P.¹² His examples include those in which (he claims) atomic numbers of substances are essential properties of those substances. Thus consider

- (2) This substance is copper
Copper has the atomic number 29
Whatever has the atomic number 29 conducts electricity
Therefore,
This substance conducts electricity.

This, Brody would say, satisfies not only Hempel's D-N model but the essential property condition as well. Its premises contain a statement—the second one—that attributes to copper an essential property; and the object described in the conclusion (this substance) is, indeed, copper.

10. B. A. Brody, "Towards an Aristotelean Theory of Scientific Explanation," *Philosophy of Science* 39 (1972), pp. 20–31; see p. 23.

11. *Ibid.*, p. 26.

12. *Identity and Essence* (Princeton, 1980), pp. 115ff.

Brody's models are, I suggest, proposed as answers to question (c) of Section 1. They are meant to provide necessary and sufficient conditions for being a good explanation. Brody, unlike Hempel, does not introduce the concept of a potential explanation; and it is not clear to me whether what he says can be used to formulate necessary and sufficient conditions for being an explanation (where goodness is not implied). Nor does he characterize explaining acts. But it is evident that the conditions for explanations that he does supply do not invoke the concept of an explaining act.

Brody claims to derive his view from Aristotle's doctrine of scientific knowledge in the *Posterior Analytics*. In this work Aristotle does not develop separate models, as Brody has done, but only one. According to it, an explanation (or what Aristotle calls a "demonstration") is "a syllogism productive of scientific knowledge" (bk. 1, ch. 2). The premises in such a syllogism must show not simply that the phenomenon to be explained did occur, but that it had to occur. To do this, according to Aristotle, they must ascribe to things properties that are essential to them, the having of which caused the phenomenon to be explained. On Aristotle's view, then, an explanation is a deductive argument whose premises cite both an essential property of a substance involved in the phenomenon to be explained and the cause of that phenomenon, which together permit the inference that the phenomenon had to occur.

6. CONCLUSIONS

There are other models of explanation in the literature, a number of which will be discussed in Chapter 5. But, I suggest, the ones briefly outlined here are typical in the following respects. They are concerned primarily with providing 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.

Conditions C impose requirements on the kinds of propositions which comprise the explanation and on their relationships to one another. But they make no mention of persons or types of persons who are explaining, or of particular or general types of explaining acts, or of audiences for whom explanations are given or intended. Nor do the modelists in question even attempt to characterize explaining acts. One might speculate that if they were to do so, their view would be that an explaining act—one in which a speaker explains by uttering something—can be understood quite simply as an act in which the sentence

or sentences uttered express a proposition or argument satisfying their conditions for being a (good) explanation. Views of this sort would hold that a sentence of the form "S explains q by uttering u" is true if and only if S utters u, and u expresses an explanation. Thus, the concept of an explanation is to be understood independently of that of an explaining act, while that of an explaining act is to be understood by reference to the concept of an explanation.

I turn now to the development of a very different type of theory. It begins with explaining acts, which are the subject of the chapter that follows. In Chapters 3 and 4, respectively, the theory will characterize the concept of an explanation, and that of a good explanation, by reference to such acts. In the course of the latter discussion I will argue that serious problems in the characterization of these concepts will arise if explaining acts are ignored or de-emphasized.

CHAPTER 2

Explaining

1. CONDITIONS FOR AN ACT OF EXPLAINING

The verb "to explain" is, to borrow a classification from Zeno Vendler, an accomplishment term.¹ It has a continuous present, "is explaining," that indicates that an act is occurring that occupies some stretch of time. But unlike some other verbs which also have a continuous present, such as "to run" and "to push" (which Vendler calls activity terms), it has a past tense which indicates not simply a stop to the act but a conclusion or completion. If John was running, then no matter for how long he was running, he ran. But if the doctor was explaining Bill's stomach ache, then it is not necessarily true that he explained it, since his act may have been interrupted before completion.

Sylvain Bromberger suggests that although the accomplishment use of "explain" is the most fundamental there is also a non-accomplishment use, illustrated by saying that Newton explained the tides. Here

one need not mean that some explaining episode took place in which Newton was the tutor. One may mean that Newton solved the problem, found the answer to the question.²

Using Vendler's terminology, Bromberger classifies this as an "achievement" use of "explains." Achievement terms (e.g., "winning a race"), unlike accomplishment terms, describe something that occurs at a single moment rather than over a stretch of time.

I believe that Bromberger is mistaken. If Newton has simply solved the problem or found the answer, although he may be in a position

1. Zeno Vendler, *Linguistics in Philosophy* (Ithaca, 1967), p. 102.

2. Sylvain Bromberger, "An Approach to Explanation," in R. J. Butler, ed., *Analytical Philosophy 2* (Oxford, 1965), pp. 72-105.

to explain the tides (we say that he *would* or *can* explain them as follows), he has not yet done so until he has said, or written, or at least communicated, something. One does not explain simply by believing, or even by solving a problem or finding an answer, unless that belief, solution, or answer is expressed in some act of uttering or writing. (We sometimes explain rather simple things by non-verbal acts such as gesturing, but such cases will not be of concern to me here.) This does not mean that we must construe "Newton explained the tides" as describing a particular explaining episode. Following Davidson,³ this can be treated as an existentially general sentence, i.e., as saying that there was at least one act which was an explaining of the tides by Newton.

Explaining is what Austin calls an illocutionary act.⁴ Like warning and promising, it is typically performed by uttering words in certain contexts with appropriate intentions. It is to be distinguished from what Austin calls perlocutionary acts, such as enlightening someone, or getting someone to understand, or removing someone's puzzlement, which are the effects one's act of explaining can have upon the thoughts and beliefs of others.

The illocutionary character of explaining can be exposed by formulating a set of conditions for performing such an act. To do so I shall consider sentences of the form "S explains q by uttering u," in which S denotes some person, q expresses an indirect question, and u is a sentence. (I will assume that any sentence of this form in which q is not an indirect question is transformable into one that is.)⁵

The first condition 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. If by uttering u I am asking you, or agreeing with you about, why the tides occur, by contrast to explaining it, I will not be doing so with the intention that my utterance render why the tides occur understandable. (I shall return to the concept of understanding in Section 3 after formulating the remaining conditions.)

3. Donald Davidson, "The Logical Form of Action Sentences," in N. Rescher, ed., *The Logic of Decision and Action* (Pittsburgh, 1967).

4. J. L. Austin, *How To Do Things with Words* (Oxford, 1962). Austin includes "explain" on his list of "expositives," pp. 160-61.

5. In Section 9, the field will be broadened to include cases in which u is not a complete sentence.

To explain q is not to utter just anything with the intention that the utterance render q understandable. Suppose I believe that the words "truth is beauty" are so causally efficacious with you that the mere uttering of them will cause you to understand anything, including why the tides occur. By uttering these words I have not thereby explained why the tides occur, even if I have satisfied (1). The reason is that I do not believe that "truth is beauty" expresses a correct answer to the question "Why do the tides occur?" More generally, assuming that answers to questions are propositions (see Section 3), we may say that S explains q by uttering u only if

- (2) S believes that u expresses a proposition that is a correct answer to Q. (Q is the direct form of the question whose indirect form is q.)

Often people will present hints, clues, or instructions which do not themselves answer the question but enable an answer to be found by others. To the question "Why do the tides occur?" I might respond: "Look it up in Chapter 10 of your physics text," or "Newton's *Principia* has the answer," or "Think of gravity." Some hints, no doubt, border on being answers to the question. But in those cases where they do not, it is not completely appropriate to speak of explaining. By uttering "Look it up in Chapter 10 of your physics text" I am not explaining why the tides occur, though I am uttering something which, I believe, will put you in a position to explain this.

These conditions are not yet sufficient. Suppose that S intends that his utterance of u render q understandable not by producing the knowledge that u expresses a correct answer to Q but by causing people to come to think of some non-equivalent sentence u' which, like u, S believes expresses a proposition that is a correct answer to Q. In such a case, although S utters something which he believes will cause others to be able to explain q, S does not himself explain q by uttering u. For example, to an audience that I believe already knows that the tides occur because of gravitational attraction, I say

- u: The tides occur because of gravitational attraction of the sort described by Newton.

Although I believe that u does express a correct answer to Q (Why do the tides occur?), suppose that I utter u with the following intention: that this utterance will render q understandable not by producing the knowledge of the proposition expressed by u that it is a correct answer to Q, but by causing my audience to look up the more detailed and precise answer actually supplied by Newton, which I don't present. This is like the situation in which I give the audience a hint that in this case is a correct answer, but is not the answer in virtue of which I intend q to be understandable to that audience.

To preclude such cases we can say that S explains q by uttering u only if

- (3) S utters u with the intention that his utterance of u render q understandable by producing the knowledge, of the proposition expressed by u, that it is a correct answer to Q.

In the case of the tides mentioned above, I do not intend that my utterance of u render q understandable by producing the knowledge of the proposition expressed by u that it is a correct answer to Q, but by producing such knowledge with respect to another proposition. So, according to condition (3), in such a case by uttering u I am not explaining why the tides occur.

Suppose, by contrast, I know that my audience is familiar with the answer supplied by Newton, but its members have no idea whether this answer is correct. Since the audience knows what sort of gravitational attraction Newton describes, I might explain why the tides occur, simply by uttering u. In this case I intend to render q understandable by producing the knowledge, of the proposition expressed by u, that it is a correct answer to Q. It is possible for me to have this intention with respect to u since I know that the audience is aware of the sort of gravitational attraction described by Newton.

Let us change the example once more. Suppose I believe that the audience does not know that the tides are due to gravitational attraction. I now proceed to utter u above with the intention that my utterance of u will render q understandable by the following combination of means (which I regard as jointly but not separately sufficient for rendering q understandable): (i) producing the knowledge, of the proposition expressed by u, that it is a correct answer to Q; and (ii) causing others to look up some different, more detailed, proposition (supplied by Newton) which is also a correct answer to Q. By uttering u am I explaining why the tides occur?

One might be inclined to say that I am *both* explaining q by uttering u *and* giving a clue about where to find another answer to Q. If this is correct, then (3) should be understood in a way that allows S to intend to render q understandable by a combination of means that includes producing the knowledge, of the proposition expressed by u, that it is a correct answer to Q. On the other hand, in the case just envisaged one might be tempted to say that I am doing something that falls between explaining and giving clues but is not exactly either. If this is correct, then (3) should be understood in a way that requires S to intend to render q understandable *solely* by producing the knowledge, of the proposition expressed by u, that it is a correct answer to Q. I am inclined to regard the latter interpretation of (3) as preferable, but I will not press the point. (This, of course, does not preclude

S from explaining q by formulating a number of different propositions whose conjunction constitutes an answer to Q, or from engaging in several acts in which different, though not necessarily competing, answers to Q are provided.)

In Section 9 some further conditions (involving restrictions on q and u) will be suggested whose formulation requires concepts to be introduced later. For the present I shall treat these three conditions as not only necessary but jointly sufficient. If so, then the same honor can be accorded to (3) by itself, since (3) entails both (1) and (2).

Although "explain" may be used in describing an act governed by these conditions, it can also be employed in a more restricted way to cover only correct explainings. We can say that Galileo explained why the tides occur, even though he did so incorrectly, or that he failed to explain this, even though he tried. When one has correctly explained q by uttering u one has performed the illocutionary act of explaining q and in doing so one has provided a correct answer to Q. In what follows, however, when reference is made to acts of explaining I shall mean acts for which this is not a requirement. (The concept of a correct explanation, and, more generally, the question of the evaluation of explanations, will be taken up in Chapter 4.)

2. SOME PRELIMINARY OBJECTIONS

It may be objected at this point that, to explain, what is required is not simply that S utters u with the intention to produce a certain effect, but that S intends to do so *by means of the recognition of this intention*.⁶ On this proposal, to (3) of Section 1 we must add that S also intends that q be rendered understandable by means of the audience's coming to recognize that S has the intention to render q understandable.

This addition, I suggest, should not be made. For one thing, S's audience may be tired of explanations; at the moment its members may not want anything rendered understandable to them. In such a case S could explain q to this audience with the intention that its members not recognize that S has the intention to render q understandable to them. It is possible for S to explain q to an audience while concealing from it his explanatory intentions. More importantly, the means by which S intends to render q understandable is already included in (3)—viz. producing the knowledge, of the proposition expressed by u, that it is a correct answer to Q. S does not intend that q be rendered understandable by means of the audience's recognition

6. See H. P. Grice, "Meaning," *Philosophical Review* 66 (1957), pp. 377-88.

of his intention. The latter is neither necessary nor sufficient for S to produce a state of understanding in the audience.⁷

A different objection has been raised by Robert J. Matthews.⁸ In a criticism of (3) he points out that in an explaining episode S may not only be supplying a ("direct") answer to Q; he may also be providing needed background information, correcting certain assumptions of the audience, showing the audience how the answer is compatible with its beliefs, and so forth. As a result, what S utters (a conjunction $u_1 \& u_2 \& \dots \& u_n$, let us suppose) can be construed as an "answer" to Q only in a very broad sense (one not intended by (3)). Indeed, Matthews urges, S can explain q even if S does not believe he is providing any ("direct") answer to Q.

His own account, which reflects these ideas, is this:

S explains q to (audience) A by citing (proposition) E if and only if S cites E with the intention of (i) presenting to A such information as A must learn in order to bring A to understand q, and (ii) having A come to understand q as a result of having recognized that E provides this information. (p. 75)

The proposition E may contain information in addition to an answer to Q. Indeed, according to Matthews, it need not even contain (what S believes to be) an answer to Q.

In response, let me acknowledge that during the course of uttering $u_1 \& u_2 \& \dots \& u_n$, S can be explaining q while performing other illocutionary acts as well, such as providing background information and correcting assumptions. These other acts may help set the stage for, clarify, defend, or otherwise buttress, the act of explaining q. This is true for illocutionary acts generally. Suppose I say

Three weeks ago I bought a watch at Cartiers. It has four diamonds, and it cost \$10,000. No, it is not the one I am now wearing. I promise to give the watch to you in three months. I always keep my promises to you: remember when I promised you a car.

During the course of uttering these sentences I have performed the illocutionary act of promising to give you the watch in three months. I have also buttressed my promise by describing the watch, correcting a mistake you may be making about its identity, and defending my promising record. However, it is by uttering the fourth sentence ("I promise to give the watch to you in three months") that I have performed the act of promising. Without uttering this sentence there is

7. For general arguments against this Gricean condition, see Stephen R. Schiffer, *Meaning* (Oxford, 1972).

8. Robert J. Matthews, "Explaining and Explanation," *American Philosophical Quarterly* 18 (1981), pp. 71-77.

no such act; with it there is, even if the other sentences are not uttered.

Similarly, suppose I say

Three weeks ago I bought a watch at Cartiers. It has four diamonds, and it cost \$10,000. No, it is not the one I am now wearing. The reason that I keep the watch in my safe deposit box is that I don't want to pay the high insurance premiums. I know the premiums are high because I checked with my insurance broker.

During the course of uttering these sentences I have explained why I keep the watch in my safe deposit box. I have also set the stage for this act by describing the watch, I have corrected a mistake you may be making about its identity, and I have defended a claim I make in my explanation. However, it is by uttering the fourth sentence that I have performed the act of explaining why I keep the watch in my safe deposit box. Without uttering this sentence there is no such act; with it there is, even if the other sentences are not uttered. This is not to impugn the importance of these other sentences for my explaining act. Uttering such sentences may be necessary to make the explaining act effective in this case (e.g., by allowing you to correctly identify the watch in question).

Turning to Matthews's own definition, note that it does not require that S provide what S believes to be a ("direct") answer to Q. Because of the absence of this requirement it is subject to a difficulty mentioned earlier. Suppose S believes that citing the proposition "truth is beauty" is so causally efficacious with his audience that the mere citing of it will cause the audience to understand anything, including why the tides occur. Suppose further that S believes that his audience must learn that truth is beauty in order to be brought to understand (to be caused to understand) why the tides occur. Suppose, finally, that S cites the proposition "truth is beauty" with the intention of (i) presenting to A such information as (he believes) A must learn in order to bring A to understand why the tides occur, and (ii) having A come to understand why the tides occur as a result of having recognized that the proposition "truth is beauty" provides this information. Matthews's conditions are now satisfied. But if S does not believe that the proposition that truth is beauty is a correct answer to the question Q (Why do the tides occur?), then he is not explaining q.

In the remainder of this section I shall briefly note some other uses of "explain" and some concepts that remain to be discussed.

The verb "to explain" is used not only in cases in which a particular explaining act is being described. As already noted, we may say that Newton explained the tides even if we are not describing a particular explaining episode. The truth of our claim, however, depends on the