1973

Explanation In Science

Brian William Cupples

EXPLANATION IN SCIENCE

by

Brian William Cupples

Department of Philosophy

Submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy

Faculty of Graduate Studies
The University of Western Ontario
London Canada

April 1973

Abstract

This thesis is an examination of one prominent analysis of explanation: the deductive-nomological model of Hempel and Oppenheim. The purpose of the thesis is to evaluate the present status of this model and to erase some misconceptions which have arisen in the literature.

Chapter one consists of a statement of the model, along with its rationale. The basic assumption of Hempel and Oppenheim appears to be quite natural and simple, viz., that explanation of singular sentences takes place by the deductive subsumption of such sentences under other sentences consisting of laws and initial conditions. This view, however, is not free from difficulty.

In Chapter two we survey all of the recent literature which has arisen since the discovery by Eberle, Kaplan, and Montague that the Hempel and Oppenheim model is trivializable. The authors show that, on the Hempel and Oppenheim analysis, a relation of explanation holds between almost any law and almost any singular sentence. Several authors have proposed revisions, most of which have only served to further confuse the issues. After dispensing with the inadequate proposals, we argue that a revision proposed by David Kaplan is essentially sound, and that the criticism leveled against his analysis by J. Kim is unsound. Kim's error will be shown to lie in his confusing epistemic conditions for explanation with logical conditions for explanation.

Briefly, Kaplan's analysis retains the rather idealized requirement of the original Hempelian model that the explanans be true. Kim's objection rests upon the fact that certain sentences may be true and yet not be known to be true. Now while we certainly do not want to dispute this distinction, we argue that it is simply irrelevant to Kaplan's analysis. This point is developed in more detail in Chapter six, where we do introduce a model which explicitly includes certain epistemic requirements. Yet when we discuss revisions of the original proposal, it is essential that we keep such models distinct.

In Chapter three, we make use of Kaplan's analysis to prove that a newly proposed model by Charles Morgan is also inadequate.

Chapters four and five deal with explanations of laws. It is pointed out that, contrary to popular belief, the deductive-nomological model is not defined relative to non-singular explananda. These two chapters are more programmatic than final. We believe that a model of explanation for generalized explananda requires a more thorough analysis of "theory" than has yet been provided. We therefore content ourselves in these chapters with a preliminary account of the sorts of examples that any such analysis of explanation ought not to permit.

Preface

In this thesis we critically examine the logical, semantical and epistemological conditions of scientific explanation. Our starting point is the deductive-nomological or covering-law model of explanation, as formulated by Karl Popper and subsequently developed by Carl Hempel and Paul Oppenheim.

Popper formulated the notion of explanation roughly as follows: to provide an explanation of an events means to deduce a statement which describes it, using as premises of the deduction one or more universal laws together with certain singular statements, the initial conditions. This brief sketch was elaborated upon by Hempel and Oppenheim, who attempted to state rigorous conditions for explanation with respect to a formalized language, the first-order predicate calculus without identity.

The Hempel and Oppenheim version has been subjected to sustained criticism in the literature. Such criticisms have been diverse, ranging from the question of the applicability of the model in explaining historical actions to the tenability of the logical conditions of the model itself. It is the latter problem with which we will be concerned. Although there is not at present any one version of the covering-law model which is free from logical difficulty, it is the aim of this thesis to provide such a model. In fact, we argue for the plausibility of two models, one requiring that the premises of a deductive explanation be _true_, and one which requires the satisfaction of a weaker condition, namely that the premises be _rationally acceptable_.

Chapter one provides the rationale and general framework of the deductive-nomological model. The problems with the Hempel and Oppenheim account of 'law' and of 'theory' are noted but not resolved in this work. Our objective is to lay bare the problems surrounding the logical structure of explanatory arguments. It is found, for example, to be incorrect to say simply that the statement of the event to be explained must be a logical consequence of the explanatory premises. This fact is established in Chapters two and three, where all the recent literature pertaining to the logical conditions of explanation is examined. Only the proposal by David Kaplan survives the critical survey.

In assessing the merits of the criticisms of Kaplan's proposal we are led to suggest in Chapter six a different model of explanation, which we call the "rationally-acceptable" model. This model is suggested in order that explanations may be appraised as to their adequacy when their explanans satisfy a criterion of acceptance rather than of truth. And although we believe that neither the model of S-explanation nor the model of rationally-acceptable explanation is entirely adequate (for roughly the reasons suggested by Sylvain Bromberger in his paper listed in the bibliography), we do feel that the present work provides a framework in which more profitable reasearch can be carried out.

Chapters four and five provide an analysis of explanation when what is to be explained is a law or a general regularity rather than the statement of a particular event. Robert Ackermann had proposed an extension of the Hempel and Oppenheim model of explanation which purported to be capable of providing such an analysis, but his proposal has since been shown to be inadequate. Our analysis in Chapter four is free from the difficulties embodied in Ackermann's approach and seems to exclude the intuitively undesirable counter-examples which could be brought against it.

I would like to akcnowledge here the rewarding experience that graduate work at the University of Western Ontario has been. The consideration and enthusiasm of the Graduate Faculty has always been a source of inspiration and I am grateful to them for their assistance and their example.

My fellow graduate students, Leon Ellsworth and Danny Steinberg in particular, have always served as ready critics and I should to express here my thanks to them for their helpful conversation and their friendship.

Professor James Leach has assisted me in immumerable ways and has earned my respect and gratitude. Of course, I am myself responsible for any and all of the shortcomings of this work.

CONTENTS

Chapter I

Deductive Explanation

Carl Hempel and Paul Oppenheim, in their now classic paper "Studies in the Logic of Explanation",[1] set out what they consider to be a set of necessary and sufficient[2] conditions definitive of scientific explanation. The adjective 'scientific' points to the fact that the authors were interested in explicating only one aspect of the concept of explanation, i.e., that aspect which amounts to providing an answer to the question "why?" asked of some natural phenomenon. Typically, answers to the question "why?" establish the fact to be explained, the explanandum, as an instance of a regularity in nature. Such an answer consists not simply in a statement of certain conditions which preceed the explanandum event, but also includes a law of nature which establishes a connection between some

---

1. Hempel, C., and Oppenheim, P., "Studies in the Logic of Explanation", *Philosophy of Science*, 15 (1948), pp. 135-175. Also reprinted in Hempel, *Aspects of Scientific Explanation*, 1965, pp. 245-290. All further references to Hempel and Oppenheim will be to the latter source.

2. It might be objected that only necessary conditions are offered; but while it is true that Hempel begins by considering only necessary conditions, his final formulation is quite clearly intended to be in terms of necessary and sufficient conditions, as definition (7.8), found later in this chapter, testifies.

of those antecedent conditions and the event demanding explanation. The important function of the law is to prohibit certain fortuitous circumstances from counting as explanatory.

In short, a scientific explanation clarifies why a given event occurred in terms of general regularities which we hold to be true of the world. These range from regularities between observable characteristics to regularities that hold between certain non-observational properties.

In the sense of Hempel and Oppenheim, therefore, a scientific explanation provides an answer to "why did x occur?" by subsuming x under certain regularities in nature. Other analyses of "explanation" can be given, however, in which one provides answers to questions which are not why-questions. Bromberger gives as examples "how-questions", "what-corresponds-at-the-microscopic-level-questions", "what-is-the-cause-of-questions", etc.[3] These questions call for answers peculiar to the type and different from those called for by why-questions.

In our discussion these other types of question and their answers will not concern us. It is answers to why-questions that we will consider as candidates for explanation. The problem will be to specify conditions which single out adequate answers to such questions. This, I think, is

---

3. Bromberger, S., "Why-Questions", in Mind and Cosmos, R. Colodny, ed., 1966, pp. 86-111.

the task which Hempel and Oppenheim set for themselves.

1.1. The definition of Hempel and Oppenheim

As mentioned above, Hempel and Oppenheim conceive of explanation as the provision of an answer to a why-question. This involves the derivation of the fact to be explained (or, more precisely, of a statement of the fact to be explained, the explanandum) from other statements which express antecedent conditions and general laws. To explain a particular fact, on their view, is to derive a sentence describing that fact from general laws, in conjunction with other singular sentences.

Some of the main characteristics of this conception of explanation are summarized in the following schema;[4]

$$\begin{array}{ll} T: & L_1 \ldots L_r \\ C: & \underline{C_1 \ldots C_k} \\ E: & \quad E \end{array}$$

where the C_i's represent statements of antecedent conditions, the L_i's statements of general laws, the horizontal line indicates logical derivability, and E represents the explanandum or statement to be explained. T and C together constitute what is called the explanans. (It is to be noted that Hempel recognizes, in addition to this conception of explanation, another type of explanation, which he calls "inductive-statistical" explanation, where the explanandum

4. Hempel, Aspects, p. 249.

follows not deductively from statements of law and antecedent conditions, but only with a certain degree of probability.[5] I mention this because Hempel has been accused of requiring all sound scientific explanations to be of deductive form, which, clearly, is not the case.)

Hempel and Oppenheim also propose certain criteria of adequacy that are to be imposed on the above schema, which are as follows: (1) the explanandum must be logically derivable from the explanans, (2) the explanans must contain general laws which are actually required for the derivation of the explanandum, (3) the explanans must have empirical content, and (4) the sentences of the explanans must be true.

For example, the following is taken to be a paradigm instance of Hempel and Oppenheim explanation:

$$\begin{array}{ll} T: & (x)(Cx \rightarrow Ex) \\ C: & \underline{\quad Ca \quad} \\ E: & \quad Ea \end{array} ,$$

which might be read "that object 'a' conducts electricity is explained by the law 'all copper conducts electricity' in conjunction with the specific information that 'a' is a specimen of copper".

1.1a. Explanations without laws?

By way of illustrating the rationale of some of these conditions of adequacy, in particular that the explanandum

---

5. See "Postscript", in Aspects, p. 291.

be logically derivable from the explanans and that the explanans contain at least one law which is essential for the derivation of the explanandum, I shall consider a recent objection to the Hempel and Oppenheim proposal requiring laws in the explanans. The objection is not a new one, although its defence is. (Scriven was first to voice disapproval in the article "Explanations, Predictions, and Laws",[6] to which Hempel replied in the long essay "Aspects of Scientific Explanation".[7]) Recently, an article appeared by J. Aronson[8] defending Scriven's argument against Hempel's objections. I turn now to consider this argument, not because I think it a good one (on the contrary, I think it just the opposite), but because it illustrates very nicely the point at issue.

Scriven's argument was that laws _may_ be relevant to an explanation, but only as a "role-justifying" ground for the explanation.[9] Explanations, said Scriven, might take the form 'q because p', where 'p' mentions particular facts but no laws.

> The citation of laws is appropriate, according to Scriven, not in response to the question 'Why q?', which 'q because p' serves

---

6. Scriven, M., "Explanations, Predictions, and Laws", in _Minnesota Studies in the Philosophy of Science_, Vol. III, pp. 170-230.
7. Hempel, C., "Aspects of Scientific Explanation", in _Aspects_, pp. 359-364.
8. Aronson, J., "Explanations without Laws", _The Journal of Philosophy_, vol. LXVI, no. 17, Sept. 4, 1969, pp. 541-557.
9. ibid., p. 542.

> to answer, but rather in response to the quite different question as to the grounds on which the facts mentioned in the 'p'-clause may be claimed to explain the facts referred to in the 'q'-clause. To include the relevant laws in the statement of the explanation itself would be, according to Scriven, to confound the statement of an explanation with a statement of its grounds.[10]

Thus, Scriven's argument is that laws play a justificatory role in explanations of the form 'q because p', but such laws themselves constitute no part of the explanans statement 'p'. For example, in "the ice cube melted because it was floating in water at room temperature" a law would justify attributing to "it was floating in water at room temperature" an explanatory role.

Hempel has pointed out that such an 'explanation' will often mention only some of a larger set of particular facts which jointly could explain the occurrence in question. Another factor which could be mentioned would be that the water and the surrounding air remained approximately at room temperature for an adequate time. Therefore, Hempel concludes that

> ... in order to justify attributing an explanatory role to the facts actually specified, one would have to cite here not only certain laws, but also the relevant particulars that had not been explicitly mentioned among the explanatory facts. Thus it is not clear why only laws should be singled out for the function of role-justification. And if statements of particular fact were equally allowed to serve as role-justifying grounds in explanations, then the distinction between explanatory facts and

10. ibid.

> role-justifying grounds would become obscure and arbitrary.[11]

Another way of stating the point is this: not just any fact 'p' is sufficient to explain 'q', for there are many facts which could be taken as conditions antecedent to the event demanding explanation. What makes some preceeding facts relevant to 'q' is the existence of a statement asserting a lawlike connection between 'p' and 'q'. This lawlike connection is all important. For example, to cite the fact that some specimen of water at sea level was heated to 212 F. to explain why the water boiled, we would need to know that these two occurrences were indeed nomically related. For, if it so happened that at other times, when water was heated to 212 F., the open container exploded, or ice formed a layer across the surface, we would be completely in the dark as to why on this particular occassion the water boiled.

Thus, on the Hempelian view, general laws not only serve to justify the connection of 'p' and 'q' statements, they also make the 'p' statements relevant to 'q'. Aronson asks us to note that both Hempel and Scriven presuppose that 'p' statements do not, in themselves, entail 'q' statements. This will become important in his subsequent analysis.

Now let us turn to Aronson's claim that 'q because p' does in fact function as an explanation even in the abscence

---

11. Aspects, p. 359.

of laws in the statement 'p'. His claim is that, due to the unique internal structure of some 'p' statements, a condition of explanatory relevance can be stated purely on the basis of an analysis of these 'p' statements.

Aronson says: "Presented below is a significant class of 'p'-statements that serve as complete explanations without containing laws in their expressions".[12] These statements are the following:

(a) John <u>knocked</u> over the ink bottle.
(b) The connecting rod <u>pushes</u> and <u>pulls</u> the piston back and forth.
(c) The bar <u>holds</u> A and B together.
(d) Pi mesons <u>bind</u> protons and neutrons in the nucleus.
(e) The gears <u>maintain</u> the angular velocities of the rods equal to one another.

Aronson asserts that (a)-(e) supply complete explanations of the form 'q because p' in the following way:

(a') The ink bottle is on its side because John knocked it over.
(b') The piston moves back and forth because the connecting rod pushes and pulls the piston back and forth.
(c') A and B stay together because the bar holds them together.
(d') Protons and neutrons remain in the nucleus because Pi mesons bind them in the nucleus.
(e') The rods move at equal angular velocity because the gears maintain their rates equal.[13]

What accounts (in part) for (a)-(e) being explanatory, accord-

---

12. Aronson, p. 544.
13. ibid.

ing to Aronson, is the occurrence of transitive verbs such as "knock", "pull", "hold", etc., in the 'p' statements. "The subject term of each of these expressions refers to that object which performs the action ... designated by the transitive verb; the object that receives the action, of course, is designated by the direct-object term".[14] Aronson maintains that (a')-(e') serve as complete explanations in that they enlighten us about things that happen as a result of one thing acting on another.

The claim is that the statements 'p' (i.e., (a)-(e)) entail their corresponding 'q' statements. Aronson asserts that by "entailment" he means what is traditionally meant; viz., the 'q' statement cannot be false if the 'p' statement is true.[15] But then, in a curious footnote to this passage, he remarks: "the entailment, here, may not always be an exact or strict entailment. For example, in (a'), one may wish to claim that 'John knocked over the ink bottle' does not entail 'the ink bottle is on its side' unless it is assumed that 'the ink bottle is on its side' is synonymous with 'the ink bottle is over'. Perhaps a synonymy assumption is required for such an entailment; ... but this is a far cry from a law relating stages of a physical process."[16]

---

14. ibid., p. 545.
15. ibid., p. 547.
16. ibid., pp. 547-548.

This notion of an "inexact entailment" is confused. First of all, the example given is not a case of entailment at all; from 'John knocked over the ink bottle' it by no means follows that 'the ink bottle is on its side'. For it is possible for the 'p' statement to be true and the 'q' statement false, contrary to the definition of "entails".

For example, an ink bottle of which all we know is that is has been knocked over does not thereby force us to conclude as a matter of logic that it is on its side anymore than we can conclude that it is resting upside down or that it immediately bounces back into an upright position. It is simply wrong to claim that the 'p' statement in (a') <u>entails</u> the 'q' statement thereof. And to claim that there is an entailment but not an 'exact' one is more confused still. Finally, an appeal to an assumption of synonymy is, clearly, of no help in clarifying the matter, if Quine's remarks in this regard are to be taken seriously.[17]

Aronson later gives us a rule for <u>generating</u> 'q' statements from 'p' statements, which is perhaps an improvment on calling the relationship between them one of entailment. In the generation rule provided, a lot turns upon the objective complement in a sentence with a transitive verb. For instance, in (a), the objective complement is "over", which serves to complete the action expressed by the verb

---

17. Quine, W.V.O., "Two Dogmas of Empiricism", in <u>From a Logical Point of View</u>.

and also to modify the direct object. The claim is that

> If a 'p' statement is meaningful and contains a transitive verb and objective complement, then that statement <u>in itself</u> can serve to explain the 'q' statement that can be generated from it.[18]

The manner in which a 'q' statement is generated from a 'p' statement is this: let the direct object of the 'p' statement perform the office of subject of the 'q' statement and the objective complement of the 'p' statement perform the office of predicate (and place the appropriate copula between them).[19]

In a footnote to the above rule, Aronson states that this condition "is not a sufficient condition for <u>the</u> explanation of 'q because p' or what 'q' refers to, but a sufficient condition for 'p' being relevant to explaining 'q'. For example, 'p' may be false, but still a relevant consideration...."[20] Presumably what he means is that this condition serves as a sufficient condition of explanation just in case 'p' is true; when 'p' is false, the condition specifies only that 'p' is relevant to 'q'.

I will return to this point shortly. Let us first look at the new model proposed by Aronson:

<u>Explanans</u>: ('p' statement). Noun phrase+transitive verb+direct object+objective complement.

---

18. Aronson, p. 549.
19. ibid., p. 548.
20. ibid., p. 549.

Explanandum: ('q' statement). Noun phrase (direct object of 'p' statement)+predicate (copula+objective complement of 'p' statement).[21]

Thus, in (a') above,

John knocked over the ink bottle

is said to generate, by the above rule,

The ink bottle is on its side.

Not only is the second generated from the first in the above specified manner, but the second is also said to be entailed by the first.

I have given reasons for denying that the first entails the second. Quite simply, the first could be true while the second was false. But let us hold our qualms with the "entailment" in abeyance for the moment. Let us consider whether the proposed model lives up to the claim that it determines the relevancy of 'p' to 'q'.

Consider the following 'p' statement:

The clock hands move the drive motor around.

From this statement we can generate the following 'q' statement:

The drive motor moves around.

But I would wish to deny that the motion of the hands of a clock in any way produces movement in the drive motor of

---

21. ibid.

that clock. In other words, I claim that the first is irrelevant to the second, as far as explanatory import goes. Remember, Aronson's criterion of relevance allowed for the 'p' statement to be false, but still relevant to the 'q' statement generable from it. In the case of the clock example, the 'p' statement is false, and I would say of no relevance at all in an explanation of the fact that the drive motor of the clock moves around.

By way of summarizing this section I would like to emphasize the need for laws in any explanatory account. Indeed, it is only in virtue of the fact that some law or other holds that we have good reason for believing one event to be even relevant to another. Without an expression in an explanans whose interpretation is that of a physical law, we would be left with a mere random association of events and rather spurious claims that one bears any relation at all to the other. To explain why a certain event occurs, say the boiling of a certain substance, we would be justified in saying that the substance was water and raised to a temperature of 212 F. and that this explains why this particular sample boils only if this holds in general. Indeed, if this was just an accidental pre-condition of the boiling of the water, it would not explain the boiling at all. It would be analogous to the natives' claiming that the beating of the drum at sunrise explains the sun's rising.

1.1b. <u>Further comments on the deductive-nomological schema</u>

The schema depicted in section 1.1 was intended to apply only to those sentences E which were free from occurrences of variables or quantifiers, i.e., to singular sentences. Hempel and Oppenheim envisaged an extension of their model to cover generalized E sentences, but they declined from characterizing such explanations because of the following consideration. A law such as Kepler's could be explained, on the basis of the conditions so far sketched, by deriving it from the conjunction 'K&B', where 'K' represents a statement of Kepler's law and 'B' a statement of Boyles' law. 'K' does indeed follow from this conjunction, but only trivially. That is to say, what happens in this case is that 'K' is used to explain itself, which is clearly no explanation at all.

For this reason, Hempel and Oppenheim propose an analysis of explanation for singular E sentences only, and they leave the solution of difficulties involved in explanations of non-singular E sentences an open problem. Thus, until otherwise indicated, it is always to be understood in what follows that E is a singular sentence.

We should also note that in their early paper Hempel and Oppenheim require that the explanans of an adequate explanation be <u>true</u>. This requirement of truth for laws "has the consequence that a given empirical statement S can never be definitely known to be a law".[22] Thus, the deductive-nomological model, as originally conceived, is not stated in terms of an explanans which is known to be true or is highly confirmed. It is an attempt to explicate the purely formal or logical aspects of scientific explanations in terms of an idealization of what is in fact the case.

However, in a more recent work Hempel has suggested a deductive-nomological model of explanation in which an explanans need be only highly confirmed. This model will be isolated for attention in Chapter VI. The distinction is mentioned here only to facilitate our evaluation of certain objections to be examined in Chapter II.

1.2. A definition of explanation

Hempel and Oppenheim propose their definition of "explanation" in a formal language, viz., the first-order predicate calculus without identity. Such a language is to be understood as presupposed in the following, and we shall refer to this language as L.

First, certain auxiliary notions are defined. A sentence S of L is said to be singular if it contains no variables, and a singular sentence which does not contain any sentential

---

22. Hempel, Aspects, p. 265.

connective is called atomic; a basic sentence is either an atomic sentence or the negation of an atomic sentence.

For future reference I introduce now the following definitional chain:

> S is said to be a generalized sentence if it consists of one or more quantifiers followed by an expression which contains no quantifiers.
> S is said to be of universal form if it is a generalized sentence and all the quantifiers occurring in it are universal.
> S is called purely generalized (purely universal) if S is a generalized sentence (is of universal form) and contains no individual constants.
> S is said to be essentially universal if it is of universal form and not equivalent to a singular sentence.
> S is called essentially generalized if it is generalized and not equivalent to a singular sentence.[23]

It remains only to define the notions of "law" and "theory", which Hempel and Oppenheim do as follows. A fundamental law is a true sentence consisting of one or more universal quantifiers followed by an expression without quantifiers or individual constants. A sentence S is called a derivative law if (1) S consists of one or more universal quantifiers followed by an expression without quantifiers, (2) S is not equivalent to any singular sentence, (3) at least one individual constant occurs in S, and (4) there is a class K of fundamental laws such that S is logically derivable from K. A law is a sentence which is either a fundamental law or a derivative law.

---

23. Eberle, Kaplan, and Montague, p. 419.

A _fundamental_ _theory_ is a true sentence consisting of one or more quantifiers followed by an expression without quantifiers or individual constants. A sentence S is called a _derivative_ _theory_ if (1) S consists of one or more quantifiers followed by an expression without quantifiers, (2) S is not logically equivalent to any singular sentence, (3) at least one individual constant occurs in S, and (4) there is a class K of fundamental theories such that S is logically derivable from K. A _theory_ is a sentence which is either a fundamental theory or a derivative theory.[24]

With the aid of these defined notions, Hempel and Oppenheim propose a precise characterization of the concept of explanation. They propose first a tentative statement of the necessary conditions for a _potential_ _explanans_:[25]

> (7.5) An ordered couple of sentences (T,C) is a potential explanans for the singular sentence E only if
> 1) T is essentially generalized and C is singular
> 2) E is derivable from T and C jointly but not from C alone.

I start with this tentative definition in order to expose the reasons for the final formulation. If (7.5) were an adequate definition of potential explanans, the notion of

---

24. These definitions can be found, in various forms, in Hempel, _Aspects_, p. 272, or in Eberle, Kaplan, and Montague, "Hempel and Oppenheim on Explanation", _Philosophy of Science_, 28 (1961), p. 419.

25. Hempel, _Aspects_, p. 273.

explanans could then be defined as follows:[26]

(7.6) An ordered couple of sentences (T,C) is an explanans for the singular sentence E if and only if
1) (T,C) is a potential explanans for E
2) T is a theory and C is true.

Let us now consider (7.5) in somewhat more detail. It should be noted that, by condition 2) of (7.5), trivial self-explanation is precluded; that is, any explanation of the form

$$\begin{array}{c} T \\ C \\ \hline C \end{array}$$

does not satisfy (7.5).

But now let us consider a form of explanation that has been called "partial self-explanation":

$$\text{(A)}\quad \begin{array}{ll} T: & (x)(Px \rightarrow Qx) \\ C: & Pa \& Rab \\ \hline E: & Qa \& Rab \end{array} \;.$$

Here, part of the explanandum is derivable from C alone. However, Hempel and Oppenheim decline from ruling out this form, for the following reason. If we were given the following paradigm form of explanation,

$$\text{(B)}\quad \begin{array}{ll} T: & (x)(Px \rightarrow Qx) \\ C: & Pa \\ \hline E: & Qa \end{array}$$

this could be written in the following logically equivalent form:

$$\text{(C)}\quad \begin{array}{ll} T: & (x)(\sim Px \vee Qx) \\ C: & (Pa \vee Qa) \& (Pa \vee \sim Qa) \\ \hline E: & (Pa \vee Qa) \& (\sim Pa \vee Qa) \end{array}$$

---

26. ibid.

This reformulation shows, according to Hempel and Oppenheim, that part of the content of the explanandum is always contained in the content of the singular sentence of the explanans, and thus is explained by itself, at least in part. For this reason they retain partial self-explanations in their system, while admitting that their intuitions are not altogether clear in this regard.

But let us be quite clear as to the problem involved here. Hempel and Oppenheim feel constrained to accept (C) on the basis of its being equivalent to (B), and thus they also feel constrained to accept (A). That is to say, they find no difference between (A) and (C); both allow for the self-explanation of one conjunct of the explanandum in terms of one conjunct of C. And since (C) must be admitted, therefore (A) must be.

However, I will argue in a later chapter that there are indeed conditions to be met which will exclude (A) but not (C). Briefly, the idea is this. In (A), 'Rab' will never be explainable in the strict sense of following from T and C jointly but not from C alone. This is clearly undesirable as it by-passes any use of laws. Yet there is a way of differentiating between this case and case (C). I will discuss this point in more detail in Chapter II, when David Kaplan's proposal is considered.

Now let us return to our discussion of (7.5). As was just mentioned, no change in (7.5) was made in spite of "partial self-explanation", which this definition allows. However, there is another difficulty with (7.5), a difficulty which does lead to a revision, as illustrated by the following example and discussion.

T: $(x)(Px)$
C: $\underline{Pa \rightarrow Rab}$
E: Rab

This form of explanation ought to be ruled out, for the following reason. T, being a law, must be true, so that the truth of C is made dependent upon the truth of E (notice that 'Pa→Rab' is logically equivalent to '∿PavRab'). That is to say, in the words of Hempel and Oppenheim, that the verification of C depends upon the verification of E. This, they claim, is a form of self-explanation, in a sense somewhat wider than that discussed above.

Furthermore, if such forms were accepted, then any law could be used to explain any event whatsoever. For we could always construct an explanation satisfying (7.5) in the following manner: E, the explanandum, is assumed to be true, so we can form the sentence C by choosing any antecedent we want to materially imply E. Choose the instantiation of any law whatsoever, and then take that law as the T of the definition. Then (7.5) is satisfied.

Hempel and Oppenheim remedy this defect by making the following stipulation: the assumption that T is true must not imply that the verification of C necessitates the verification of E. Now then let us consider the Hempel and Oppenheim account of verification. This will turn out to be quite essential to the following development (in Chapter II).

The following is an illustration of the verification of a molecular sentence M.[27] The sentence

M= ((Pa&Qa)vRa)

may be verified either by establishing the truth of 'Pa' and 'Qa', or by establishing the truth of 'Ra', both of which have M as a logical consequence. Verification of a molecular sentence S may be defined generally, in the Hempel and Oppenheim analysis, as the establishment of the truth of some class of basic sentences which has S as a consequence. Thus, their additional stipulation becomes: The assumption that T is true must not imply that every class of true basic sentences which has C as a consequence also has E as a consequence.

Before stating this condition in final form, the authors make the following comment:

> As brief reflection shows, this stipulation may be expressed in the following form, which avoids reference to truth: T must be compatible in L with at least one class of basic sentences which has C but not E as a consequence; or equivalently: There must exist at least one class of basic sentences

---

27. ibid., p. 277.

> which has C, but not E nor the negation of T as a consequence in L.[28]

However, notice that by removing the reference to truth, Hempel and Oppenheim are settling for what may be called "potential verification". That is, there must be a basic sentence, or a class of basic sentences, possibly false, which has C as a consequence, but not E; in other words, it must be possible to verify C independently of E.

For the moment, I wish to only point out this fact. More attention will be devoted to this consideration in the sequel.

Hempel and Oppenheim formulate their final definition of potential explanation as follows (note that the second condition of (7.5), requiring that E be derivable from T and C jointly but not from C alone, is redundant if the above verification clause is adopted):

> (7.8) The ordered couple (T,C) is a potential explanans for the singular sentence E if and only if
> 1) T is essentially generalized and C is singular
> 2) E is derivable from T and C jointly
> 3) T is compatible with at least one class of basic sentences which has C but not E as a consequence.[29]

The definition of explanation given in (7.6) in terms of potential explanation remains the same, and it is now asserted that a singular sentence E is explainable by a

---

28. ibid.
29. ibid., pp. 277-278.

theory T if there exists a sentence C such that (T,C) constitutes an explanans for E.

By way of summarizing this section, two points have been singled out for discussion in later sections. First, we have noted the reluctance of Hempel and Oppenheim to rule out what has been called "partial self-explanation", and secondly we have noted that they adopt what might be called a theory of possible verification. Some of the difficulties to be discussed in the subsequent development will be traced to the adoption of these two positions. Before turning to a consideration of a series of proposals and counter-examples to the above described model, a few words must first be said concerning the Hempel and Oppenheim treatment of laws.

1.3. Laws

It is to be noted that the Hempel and Oppenheim conceptions of "law" and "lawlike" are not entirely adequate in their present formulations. The considerations so far presented do not allow for a precise distinction between nomic universals and accidental generalizations. Let me illustrate this distinction by way of example.

Consider

(1) Every man in this room is less than 7' tall.

and

(2) Copper expands when heated.

(1), we should say, is not lawlike, even if true, whereas (2) is both lawlike and true. How, then, do Hempel and Oppenheim distinguish between such cases in general?

They considered two criteria upon which to make the required distinction; viz., (i) (1) makes an assertion about a finite number of objects only, and (ii) it also makes reference to a specific object. Let us consider these two criteria in turn.

First, the fact that (1) makes reference to a finite number of objects does not suffice to distinguish it from bona fide laws; for Kepler's laws refer to a finite number of planets. To obviate this difficulty, Hempel and Oppenheim distinguished between fundamental and derivative laws, requiring that fundamental lawlike sentences satisfy a condition of non-limited scope. Derivative lawlike sentences are sentences which contain essential occurrences of constants and which do not satisfy the non-limited scope requirement. In this manner they believed that Kepler's laws, which they supposed to be derivable from the more comprehensive Newtonian laws, were distinguishable from (1), insofar as (1) is not derivable from a set of fundamental lawlike sentences.

However, Ernest Nagel has pointed out[30] that Kepler's laws are not, strictly speaking, derivable from Newton's

30. Nagel, E., The Structure of Science, p. 58.

laws, and consequently that the concept of "derivative law" proposed by Hempel and Oppenheim is too restrictive.

Hempel, in a Postscript to "Studies in the Logic of Explanation",[31] discusses this problem and is hopeful that the condition of non-limited scope which was imposed on fundamental lawlike sentences may also be extended to apply to derivative lawlike sentences as well. Nagel, in his discussion of laws, requires that lawlike sentences in general should be "unrestricted universals", i.e., "their 'scope of predication' must not fall into 'a fixed spatial region or a particular period of time'".[32]

To understand part of what Nagel has in mind it will be instructive to consider one of his four requirements imposed on laws. This is a requirement to be satisfied outside, as it were, the language L. It is helpful to think of such a condition as one which is necessary for the acceptance of a universal generalization into the set K_L of laws; or, from another point of view, as an epistemological condition which a universal generalization must satisfy in order to be accepted into our body of knowledge K.

To be clear as to the nature of this condition, reconsider for the moment sentence (1). What kind of evidence would justify our accepting (1) as true?

---

31. Aspects, pp. 291-295.
32. ibid., p. 292.

One class of evidence statements could consist of sentences of the form

> individual 'i' is in this room and 'i' is less than 7' tall,

where the examined individuals exhaust the scope of application of (1). If we had reason to believe that there were some individuals in the room not included in the sample, we would not be in a position to assert "everyone in the room is less than 7' tall" as true.

But in this case, we could never use (1) to make a prediction beyond the examined instances which were used to establish (1). Its scope of predication would not extend beyond the examined individuals.

On the other hand, (1) might be accepted on the grounds that only a fair sample of men in the room were examined, from which it is inferred that the unexamined men in the room are also less than 7' tall. Yet here too it is an assumption of the inference that the men in the sample come from a population that is complete and will not be augmented. For example, we would not infer from the observed sample that everyman in the room is less than 7' tall if we knew that the members of the Harlem Globetrotters basketball squad were going to be added to the population. Thus, if we accept (1) on the basis of a sample, we do so on the assumption that the sample comes from a population that will neither be increased nor altered.

Yet there seem to be no analogous assumptions made concerning the evidence upon which sentence (2) is accepted. For consider two cases analogous to the two above. In the first case, if the evidence for (2) completely exhausted the scope of predication of (2), then we would be reluctant to call (2) a law. That is to say, we could never use (2) to predict the behaviour of a yet unexamined specimen of copper, since by hypothesis all specimens have been examined.

However, it is much more likely that (2) will be accepted on the basis of a sample of copper specimens, and thus it can be used to predict the behaviour of as yet unexamined specimens. In this case, unlike that above, there is no assumption that the sample of copper specimens comes from a population that will neither be augmented nor altered. The scope of (2) can be extended to any specimen of copper, in the sample or not. Nagel therefore states that, "for an unrestricted universal to be called a law it is a plausible requirement that the evidence for it is not known to coincide with its scope of predication and that, moreover, its scope is not known to be closed to any further augmentation".[33]

Let us now consider the second criterion advanced by Hempel and Oppenheim for distinguishing laws from accidental

33. Nagel, pp. 62-63.

generalizations. Their second criterion was based upon the observation that (1) makes reference to a specific object, whereas (2) does not. As in the preceeding case, Hempel and Oppenheim required that fundamental laws contain zero constants, and that derivative laws, which do contain names for particular objects, be derivable from a finite set of fundamental laws. This, however, is not sufficient, for two reasons.

First, the generalization

(3) All Uranic objects are spherical

does not contain a constant, yet (3) refers only to a particular object, i.e., the planet Uranus (according to an obvious definition of 'Uranic'). To avoid this, Hempel and Oppenheim required that all the primitive predicates of L be "purely qualitative". A predicate is purely qualitative, in their sense, only if a statement of its meaning does not require reference to any one particular object or spatio-temporal location.[34]

Yet the concept of a "purely qualitative" predicate is, as Hempel is aware, extremely ambiguous. Additionally, he remarks, it essentially just shifts the problem from the object language to the metalanguage, for criteria will have to be specified in the metalanguage for distinguishing between permissible and nonpermissible interpretations of

---

34. Hempel, Aspects, pp. 268-270.

the primitive predicates.[35]

However, the problem is being approached from other directions, where conditions are being sought for determining the "projectibility" or the "entrenchment" of predicates. It is hoped that one or the other of these approaches will finally yield success; for the present I shall rely on the scope requirement and certain other extra-systematic conditions for distinguishing laws from non-laws.

The second reason why the Hempel and Oppenheim account is defective turns on the fact that the definition of derivative law does not do the required job. Kepler's laws contain a designation for a particular object, viz., the sun, and these laws are not consequences of more fundamental laws from which constants are absent.

This brief discussion serves merely as an indication of some of the complexities surrounding the notion of a lawlike sentence. For the present I suggest the procedure of dividing the problem into parts, and of tackling each part singly. Problems relating to lawlikeness itself will not be discussed further in this work, although such problems relating to the logic of explanation will be.

---

35. ibid., pp. 269-270.

Chapter II

Recent Work on Explanation

The account sketched in the previous chapter has been the predominant view in the field since its conception. It seems correct to say that, despite far-reaching problems with the analysis of "lawlike", at least the logical conditions specified by Hempel and Oppenheim were considered to be correct by the majority of philosophers. Debates frequently raged in regard to the assimilation of explanation in the non-natural sciences to the Hempelian model. The model was, nevertheless, accepted as adequate in its domain of natural science, even if it was denied to be applicable to, say, the social sciences.

Yet the model does not even adequately specify the requisite logical conditions in its own domain, as will shortly be pointed out. This chapter will critically assess and evaluate the merits of various proposed revisions of the now defunct Hempel and Oppenheim model.

2.1. The theorems of Eberle, Kaplan, and Montague

R. Eberle, D. Kaplan, and R. Montague have demonstrated[1]

---

1. Eberle, R., Kaplan, D., and Montague, R., "Hempel and Oppenheim on Explanation", Philosophy of Science, 28 (1961), pp. 418-428.

that the Hempelian model is inadequate because it is trivializable. They prove five theorems which show, roughly, that between almost any theory T and almost any true singular sentence E a relation of explanation holds, according to the Hempel and Oppenheim account.

The authors first combine definitions (7.6) and (7.8) of Chapter I, writing the Hempel and Oppenheim conditions in the following form:

> An ordered couple (T,C) of sentences is an explanans for a singular sentence E if and only if
> 1) T is a theory
> 2) T is not equivalent to any singular sentence
> 3) C is singular and true
> 4) E is derivable from the set {T,C}
> 5) There is a class K of basic sentences such that C is logically derivable from K, and neither E nor the negation of T is logically derivable from K.[2]

As a first intuitively undesirable counter-example, the authors assert that, by the above definition, the sentence "the Eiffel Tower is a good conductor of heat" is explainable by the theory "all mermaids are good conductors of heat", in conjunction with the singular sentence "either the Eiffel Tower is a mermaid or the Eiffel Tower is a good conductor of heat"; formally,

(A) T: $(x)(Mx \rightarrow Hx)$
C: $Me \vee He$ $\qquad$ $K = \{Me\}$
E: He

2. ibid., p. 419.

It is obvious that this example satisfies all of the above conditions (if it is assumed that there are no mermaids and that the Eiffel Tower is a good conductor of heat). Note that although the set K contains a false sentence, this is still in accordance with the Hempel and Oppenheim conception of possible verification.

Eberle, Kaplan and Montague claim that the definition is more conclusively trivialized by the five theorems they prove. It will be instructive to consider their first theorem, which is stated as follows:

> Theorem 1. Let T be any fundamental law and E any singular true sentence such that neither T nor E is logically provable and T, E have no predicates in common. Assume in addition that there are at least as many individual constants in L beyond those occurring in E as there are variables in T, and at least as many one-place predicates in L beyond those occurring in T and E as there are individual constants in E. Then there is a fundamental law T' which is logically derivable from T and such that E is explainable by T'.[3]

A sketch of one instance of this theorem and of its consequences for the Hempel and Oppenheim analysis is the following. Let

$$T = (y)(My \rightarrow Cy)$$

and

$$E = E_1 a,$$

in accordance with the hypothesis of the theorem. T

---

3. ibid., pp. 420-421.

has as a logical consequence the fundamental law

$T' = (x)(y)((My \rightarrow Cy) \vee (Jx \rightarrow E_1x))$.

Now we can find a C such that (T',C) constitutes an explanans for E; simply take C to be

$C = ((Mb \rightarrow Cb) \vee \sim Ja) \rightarrow E_1a$.

This sentence is true, since it is a logical consequence of the (assumed true) sentence E. And taking K to be the set

$K = \{Mb, \sim Cb, Ja\}$,

condition 5) is satisfied, and (T',C) satisfies conditions 1), 3) and 4). By extending the argument a little further, it can be shown that 2) is also satisfied, thereby establishing the conclusion of the theorem.

This theorem demonstrates that the Hempel and Oppenheim conditions fail to capture the intuitive requirement that the theory of an explanans be in some sense relevant to its explanandum. For, since T has T' as a logical consequence, and (T',C) is an explanans for E, it follows that (T,C) is an explanans for E. But T and E have no predicates in common and therefore, intuitively, deal with diverse subject matter. This is surely an undesirable consequence.

The four remaining theorems of Eberle, Kaplan, and Montague follow much the same strategy, thus we may concentrate on attempts to vitiate the force of their Theorem 1. There have been several such attempts, ranging from adding conditions to the previous model which rule out the known undesirable forms (such as (A)), to proposals for an entirely new model. These revisions have also been subjected to criticism, and, as matters stand at present, there is no generally received model of deductive explanation. However, there is one proposal which I believe can be defended, and which is capable of extension in a sense soon to be clarified. To this end, we must consider the strengths and weaknesses of all proposals so far advanced. I begin this task with a discussion of an analysis provided by Professor J. Kim.

2.2. Kim's proposed revision of the H-O model

In his paper "On the Logical Conditions of Deductive Explanation",[4] J. Kim adds two conditions to the original Hempel and Oppenheim definition in an attempt to avoid the theorems of Eberle, Kaplan, and Montague. Let me first rewrite these conditions in the form provided by Kim, for ease of reference:

E1. T is a theory
E2. T is essentially generalized
E3. C is singular and true
E4. $\{T,C\} \vdash E$
E5. There is a class of basic sentences K such that
  a) $K \vdash C$
  b) not $K \vdash E$
  c) not $K \vdash {\sim}T$.[5]

---

4. Kim, J., "Discussion: On the Logical Conditions of Deductive Explanation", Philosophy of Science, 30 (1963), pp. 286-291.

5. This formulation is given by Charles Morgan.

Kim finds the difficulty revealed in the trivialization theorems to lie in the fact that they allow, and their proofs depend on, a relation of entailment to hold between the explanandum E and the singular sentence C. A glance at example (A), section 2.1, will verify and illustrate this fact. Kim is thus led to the additional requirement that C be put in conjunctive normal form, deleting inessential occurrences of subsentences, and that E not entail any conjunct C_i of this conjunctive normal form, thusly:

(K1) For all i, $(\sim\vdash C_i \rightarrow \sim\{E\}\vdash C_i)$.

Kim's second condition is needed to rule out the following purported explanation (which will be recognized as being a form of "partial self-explanation"):

T: (x)(Ux→Wx)
C: Ua&Ka   K= {Ua,Ka}
E: Wa&(KavJa)

This form is objectionable because one conjunct of E follows from C alone; it is ruled out by not allowing the derivation of any E_i (the conjuncts of the conjunctive normal form of E) from the set of basic sentences K. Thus Kim adds the condition

E5.d) For all i, not $K\vdash E_i$.

Kim's proposal then is to save the Hempel and Oppenheim model of explanation by blocking the derivation of any of the conjuncts of C from E. In so doing he had hoped to prevent the trivialization of the model embodied in Theorem 1 of the previous section. In the next section we will see that his conditions do not suffice for this purpose.

2.3. The trivialization of Kim's revised model

Charles Morgan, in a paper entitled "Kim on Deductive Explanation",[6] proves that one can block the derivation of any of the C_i's from E as Kim requires and still not satisfy the requirement that the explanans be in some sense relevant to the explanandum. More specifically, he shows that one can prove theorems similar to those of Eberle, Kaplan, and Montague on the basis of the Hempel and Oppenheim conditions supplemented by the Kim conditions.

Morgan first proves a few lemmas, but we need not concern ourselves with these. It will be sufficient for my purposes to state, without proof, Morgan's Theorem M1 (analogous to Theorem 1 of Eberle, Kaplan, and Montague) and an example which illustrates the consequences of this theorem.

> Theorem M1. Let T be any fundamental law and E any true singular sentence, with no predicates in common. Assume in addition that there are at least as many constants in L beyond those in E as there are variables in T, at least as many one-place predicates in L beyond those occurring in T and E as there are constants in E, and at least one more distinct predicate P, which does not occur in E or T. Let E' be any basic sentence of P. Then there is a fundamental law T' which is logically derivable from T and such that EvE' is

6. Morgan, C., "Discussion: Kim on Deductive Explanation", Philosophy of Science, 37 (1970), pp. 434-439.

explainable by T'.[7]

The net import of this theorem is that any instance of the trivialization theorems of Eberle, Kaplan, and Montague which is ruled out by (K1) can be re-instated so as to satisfy all of Kim's conditions by simply disjoining any sentence whatsoever to the original explanandum. As a result, almost any theory will explain almost any singular disjunctive sentence whatsoever.

Following the example given above to illustrate Theorem 1 of Eberle, Kaplan, and Montague, we have

T: $(y)(My \rightarrow Cy)$
T': $(x)(y)((My \rightarrow Cy) \vee (Jx \rightarrow E_1 x))$
C: $((Mb\ Cb) \vee \sim Ja) \rightarrow E_1 a$
E: $E_1 a \vee Pa$

This example shows that almost any disjunctive sentence E will be explainable by almost any theory T, although T and E have no predicates in common. The basic insight, of course, is that by making use of the rule of addition the deduction essential to the Eberle, Kaplan, and Montague theorems still goes through, while at the same time the derivation of C (or some C_i) from E is precluded.

Let me provide one more counter-example to Kim's proposal, similar to the "mermaid" counter-example (A) of section 2.1. This is not to belabor the obvious, but will set the stage for later discussion.

---

7. ibid., pp. 438-439.

Consider then the following argument, the interpretation of which is the same as that of (A), with the predicate 'Ex' being interpreted as "x is a good conductor of electricity":

(B) T: $(x)(Mx \rightarrow Hx)$
C: $\underline{Me \vee He}$ $K = \{Me\}$
E: $He \vee Ee$

That this example satisfies Kim's conditions is one consequence of Morgan's theorem, and it is as equally undesirable as (A). A variation of this example will be considered in the next chapter, where Charles Morgan's own recent proposal is evaluated. It is sufficient to note for the present that we take such examples quite seriously.

In the next section we discuss Professor David Kaplan's proposal (which actually appeared in print before Kim's paper), the significance of which does not seem to be generally recognized. I have discussed Kim's paper first in order to show that the failure of his model is directly attributable to what Kaplan has diagnosed as the source of the ailment in the original Hempel and Oppenheim model. I will also argue that Kim's criticism of Kaplan's proposal is completely misguided, based as it is on a mistaken interpretation of the general character and intent of the original deductive-nomological model. This point is discussed in detail in Chapter VI.

2.4. Kaplan and S-Explanation

In "Explanation Revisited",[8] David Kaplan proposes that several revisions be made in the Hempel and Oppenheim model. Since his cogent proposal has been rather summarily dismissed it will be wise to examine it in some detail. I will, therefore, discuss his revisions in three sub-sections, in order properly to assess each of them.

2.4a. Kaplan's simplification of H-O explanation and three conditions of adequacy

Kaplan first proposes the following simplification of the notion of H-O explanation; the simplification lies in the fact that condition 2) of the Hempel and Oppenheim model, given above, requiring that T not be equivalent to any singular sentence, is shown to follow from the remaining four conditions, and is therefore superfluous. We state his theorem here without proof:

> Theorem 1: An ordered couple (T,C) of sentences is an H-O explanans for a singular sentence E if and only if the following conditions are satisfied:
> 1) T is a theory
> 2) C is singular and true
> 3) $\{T,C\} \vdash E$, and
> 4) there is a class K of basic sentences such that $K \vdash C$ but neither $K \vdash E$ nor $K \vdash \sim T$.[9]

This of course is not a crucial revision which changes

8. Kaplan, D., "Explanation Revisited", Philosophy of Science, 28 (1961), pp. 429-436.
9. ibid., p. 429.

the set of acceptable explanations, but only serves to eliminate a condition which is redundant. But Kaplan's newly proposed criteria of adequacy do lead to a crucial revision, and I turn to these now.

The first two of these requirements are the following:

R1. If a singular sentence is explainable by a given theory, then it is explainable by any theory from which the given theory is logically derivable.

R2. Any singular sentence which is logically derivable from singular sentences explainable by a theory is itself explainable by that theory.[10]

These two criteria require that the relation of explanation be closed under the relation of logical derivability. I wish to withhold any discussion of these criteria until the following chapter. In the meantime, let us adopt R1 and R2 as desirable, i.e., as conditions which should be met by any adequate account of explanation.

Kaplan's third criterion of adequacy is the following:

R3. There is an interpreted language L which contains a fundamental theory T and singular sentences E and E' which are true but not logically provable such that E is explainable by T and E' is not explainable by T.[11]

This criterion is also quite plausible; it certainly seems correct to require of an analysis of explanation that there be some languages containing theories by which certain

10. ibid., p. 430.
11. ibid., p. 431.

singular sentences are explainable and others are not. The effect of this requirement is to prohibit a definition of explanation which allows for the explanation of _any_ singular sentence E by means of _any_ theory T. Such a definition would surely be trivial.

From the results of Eberle, Kaplan, and Montague, Kaplan concludes that if H-O explanation satisfies R3, then it does not satisfy R1 and R2; he also shows that H-O explanation does satisfy R3, leading him to propose a new analysis in which all three conditions are simultaneously satisfied.

2.4b. _Kaplan's second revision_

The second proposal made by Kaplan is in connection with condition 4) of the simplified Hempel and Oppenheim model given in Theorem 1 of section 2.4a. As was pointed out in Chapter I, Hempel and Oppenheim required only that there be a _possible_ verifying class K. Kaplan remarks that this requirement "is not yet strong enough. It must not be merely _possible_ for the verifying class K to exist; there must be an actual verifying class K. That is to say, the members of K must be true."[12] Failure to make this requirement leads to the difficulty which Kaplan's theorem 4 exhibits, where E is said to be _possibly H-O explainable_

12. ibid., p. 432.

by T if there is a sentence C such that (T,C) satisfies all the requirements for an H-O explanans for E with the possible exception of the requirement that C be true.

Theorem 4. If E is possibly H-O explainable by T, then E is H-O explainable by T.[13]

To illustrate one consequence of this theorem, let us recall the first counter-example given by Eberle, Kaplan, and Montague. In this example, the theory "all mermaids are good conductors of heat" was shown to explain, on the Hempel and Oppenheim analysis, the singular sentence "the Eiffel Tower is a good conductor of heat" by taking as the sentence of initial conditions the disjunctive sentence "either the Eiffel Tower is a mermaid or the Eiffel Tower is a good conductor of heat", as the following shows:

$$\begin{array}{ll} T: & (x)(Mx \rightarrow Cx) \\ C: & \underline{\quad Me \vee Ce \quad} \\ E: & \quad Ce \end{array}$$

This "explanation" meets all of the requirements in the Hempel and Oppenheim definition, if we take for K the following set of basic sentences-

K= {Me}.

Even though this is a false sentence, it still meets condition 4) of the definition. And although Me is false, the sentence of initial conditions is true because of the second disjunct, which is the (assumed true) explanandum. But if we were to take as K the set

K= {Ce},

---

13. ibid.

this would violate the condition that

not K⊢ E.

So by requiring that K contain only true basic sentences, the mermaid counter-example, as well as the difficulty embodied in Kaplan's theorem 4, is avoided.

In section 2.5, a theorem directly analogous to Kaplan's theorem 4 will be proven with respect to Hempel-Oppenheim-Kim explanation, showing that the reason for the failure of Kim's model is directly attributable to his maintaining the <u>possible</u> verification condition.

Kaplan also shows that there is another way of making the same requirement. He says:

> ... let the singular sentence C which represents the initial conditions be in disjunctive normal form. Then in order to avoid trivial self-explanation we might require that the explanation does not depend in an essential way on any disjunct of C from which E is logically derivable. That is to say, the explanation can be carried through where the initial conditions are represented by a sentence in disjunctive normal form none of whose disjuncts logically imply E.[14]

Thus, to take the above example, the initial conditions C are already in disjunctive normal form, with the two conjuncts 'Me' and 'Ce', and one of these does have the explanandum 'Ce' as a logical consequence.

In the following very interesting theorem, Kaplan demonstrates that the two proposals above are equivalent,

---

14. ibid.

and also that they are both equivalent to another condition, which is condition (iii) of the theorem.

> Theorem 5. Assume that T is a theory and E is a true singular sentence. Then the following conditions are equivalent:
> (i) There is a singular sentence C'' such that E is logically derivable from the set {T,C''}, and there is a class K of true basic sentences such that C'' is logically derivable from K, and neither E nor the negation of T is logically derivable from K.
> (ii) There is a sentence C' in disjunctive normal form such that E is not logically derivable from any disjunct of C' and (T,C') is an H-O explanans for E.
> (iii) There is a conjunction C of true basic sentences such that E is logically derivable from the set {T,C} and E is not logically derivable from C alone.[15]

In the proof which shows these conditions to be equivalent, the conjunction C mentioned in (iii) is just the conjunction of the elements of K mentioned in (i); (iii) is only claimed to be a more perspicuous way of making the same requirement as (i) or (ii).

Thus, we might consider the following as a proposed revision of the analysis of explanation given by Hempel and Oppenheim:

(A) (T,C) is an explanans for the singular sentence E if and only if the following conditions are satisfied:
1) T is a theory
2) C is a conjunction of true basic sentences
3) $\{T,C\} \vdash E$, and
4) not $\{C\} \vdash E$.

---

15. ibid., p. 433.

By theorem 5, the following conditions could be used in place of 1)-4):

(B) 1') T is a theory
2') $\{T,C\} \vdash E$
3') C is singular and true
4') there is a class K of <u>true</u> basic sentences such that $K \vdash C$ but neither $K \vdash E$ nor $K \vdash \sim T$.

Of course, there is still yet another way of specifying the requisite conditions, which is in terms of (ii) above.

In (A), there is no need of referring to a verifying class K of true basic sentences because the conjunction C of true basic sentences will be the same as the members of K required by (B), and whereas in the original definition it was required that not $K \vdash E$, we have in (A) the condition that not $\{C\} \vdash E$, and whereas before we had not $K \vdash \sim T$, we no longer need any condition other than the truth of C.

Consider the following illustration in terms of conditions (B):

T: $(x)(Fx \vee Sx \rightarrow Gx)$
C: $Fa \vee Sa$
E: Ga

with the following set K of true basic sentences:

K $\{Fa\}$.

K here could be, of course, any of the following three sets:

$\{Fa\}$, $\{Fa, Sa\}$, $\{Sa\}$.

To show how 'Ga' can be explained by the law '$(x)(Fx \vee Sx \rightarrow Gx)$' as required by conditions (A), consider the following:

$$\begin{array}{ll} T: & (x)(Fx \vee Sx \rightarrow Gx) \\ C: & \underline{\quad Fa \quad} \\ E: & \quad Ga \end{array}$$

For C here the sentence 'Sa' or the conjunction 'Fa&Sa' would serve as the conjunction required by (A), depending on whether or not both atomic sentences were true.

The following, then, are equivalent explanations:

<u>Kaplan (with true K)</u>

$$\begin{array}{lll} T: & (x)(Fx \vee Sx \rightarrow Gx) & \\ C: & \underline{\quad Fa \vee Sa \quad} & (K=\{Fa\}) \\ E: & \quad Ga & \end{array}$$

<u>Kaplan (with conjunction C)</u>

$$\begin{array}{ll} T: & (x)(Fx \vee Sx \rightarrow Gx) \\ C: & \underline{\quad Fa \quad} \\ E: & \quad Ga \end{array}$$

After the following sub-section, in which I discuss Kaplan's third revision, I will consider an example similar to the one above which Kim takes to be a counter-example to Kaplan's analysis.

By way of summarizing this section, we have seen that Kaplan revises the original definition of Hempel and Oppenheim in the following way: the set K of verifying sentences must contain only true members, i.e., such a set must actually exist, rather than be merely possible. This requirement is given the more perspicuous formulation provided in (A), above.

2.4c. <u>Kaplan's third revision</u>

The revision of the previous section leads to a concept of explanation which satisfies R1. But this conception does not yet satisfy R2. Remember, R2 was that closure condition which requires that singular sentences which are consequences of singular sentences explainable by a given theory be themselves explainable by that theory. Thus, if we had the following explanation

$$\begin{array}{ll} T: & (x)(Fx \rightarrow Gx) \\ C: & \underline{\quad Fa \quad} \\ E: & \quad Ga \end{array}$$

one logical consequence of the given E is 'GavFa'. But if we tried to construct an explanation of 'GavFa' on the basis of the explanans (T,C) above, we would run afoul of condition 4) of (A), as well as condition 4) of the simplified version of the Hempel and Oppenheim definition given above.

Kaplan states that his intuitions are unclear with respect to R2, but instead of dispensing with it he decides to retain it and to introduce some other conditions which keep 'GavFa' explainable by the given theory. This he does by introducing the concept of <u>S-explanation</u>, or step-wise explanation. He specifies conditions under which a singular sentence E is <u>directly</u> <u>explainable</u> by a theory T and singular sentences C. ('GavFa' will not be directly explainable by the given explanans, although 'Ga' will be.) Then he specifies conditions under which a singular sentence is held to be <u>explainable</u> by a theory in virtue of the fact that such a singular sentence is logically derivable from other singular sentences which are directly explainable by the given theory ('GavFa' qualifies here).

Before stating his final definition, however, we must be clear as to Kaplan's third revision. It is this: he requires that the sentence E which is to be explained be "factored" into its directly explainable components by placing E in conjunctive normal form. We then test each conjunct separately as to whether or not it is directly explainable. Clearly, if each conjunct of the conjunctive normal form of E is directly explainable, then E itself will be explainable. And any sentence E which is not directly explainable, but which follows from singular sentences which are so explainable, is held to likewise be explainable.

Thus, Kaplan is led to the following proposal as a revision of Hempel and Oppenheim explanation:

> E is directly S-explainable by T if and only if there is a sentence C such that the following conditions are satisfied:
> 1) T is a theory
> 2) C is a conjunction of true basic sentences
> 3) E is a disjunction of basic sentences
> 4) E is logically derivable from the set {T,C}
> 5) E is not logically derivable from {C}.
>
> E is S-explainable by T if and only if E is a singular sentence which is logically derivable from the set of sentences which are directly S-explainable by T.[16]

This is not a definition of "S-explanation", but we can readily transform the given definition into one. The

---

16. ibid., p. 435.

usual way of doing this is as follows. We say that the ordered couple (T,C) of sentences is a direct S-explanans for the singular sentence E just in case conditions 1)-5) above are satisfied. Then we say that E is directly S-explainable by T if there is a C such that (T,C) is a direct S-explanans for E. (S-explanation is similar).

Kaplan finally shows that S-explanation satisfies R1, R2, and R3, and thus that this concept of explanation is immune to the criticisms of Eberle, Kaplan, and Montague.

2.5. Kim's criticism of S-explanation

In the paper discussed in section 2.2, Kim devotes a final page to Kaplan's proposal and finds it lacking. Kim first finds that Kaplan did not offer a definition of S-explanation, but rather one of S-explainability; he then proposes that we take the following as a definition of direct S-explanation:

(A) An ordered couple (T,C) of sentences is a direct S-explanans for the singular sentence E if and only if
1) T is a theory
2) C is a conjunction of true basic sentences
3) E is a disjunction of basic sentences
4) $\{T,C\} \vdash E$, and
5) not $\{C\} \vdash E$.

Given what I say above, it can be seen that this reconstruction is perfectly acceptable.

But now Kim asks us to consider the argument

(B) $\frac{(x)(Fx \vee Gx \rightarrow Jx) \quad Fb \vee Gb}{Jb}$

and comments as follows.

> Assume that the law and the singular premiss of this argument are true. Then ... 'Jb' is directly S-explainable by the law '(x)(FxvGx→Jx)', since we know that either 'Fb' or 'Gb' is true. But suppose that we have verified the disjunction 'FbvGb' without verifying either disjunct (e.g. by the use of some law in conjunction with some specific information). Then, although we know that 'Jb' is directly S-explainable by the law above, we cannot actually construct a direct S-explanation for it, since (B) is not a direct S-explananation.... It seems to me, however, that (B) is a perfectly adequate explanation.[17]

In order to get at Kim's objection here, let me first list the three ways in which we may write Kaplan's second revision (the third revision is not applicable to this example, thus I ignore it).

(I) (T,C'') is an explanans for the singular sentence E if and only if all of the following are satisfied:
 (1) T is a theory
 (2) C'' is singular and true
 (3) {T,C''}⊢ E, and
 (4) there is a class K of true basic sentences such that K⊢ C'' but not K⊢ E nor K⊢ ∼T.

(II) (T,C') is an explanans for the singular sentence E if and only if
 (1) T is a theory
 (2) C' is true, singular, and in disjunctive normal form
 (3) no disjunct of C' entails E
 (4) {T,C'}⊢ E, and
 (5) there is a class K of basic sentences such that K⊢ C' but not K⊢ E nor K⊢ ∼T.

---

17. Kim, p. 291.

(III) (T,C) is an explanans for the singular sentence E if and only if
(1) T is a theory
(2) C is a conjunction of true basic sentences
(3) $\{T,C\} \vdash E$, and
(4) not $\{C\} \vdash E$.

Professor Kaplan has shown us that these conditions are equivalent; for example, if (I) is satisfied, it follows that there exists a singular sentence C' such that (II) is satisfied, and also that there exists a singular sentence C such that (III) is satisfied. And if (III) is satisfied, it follows that (I) is satisfied.

Kaplan, you will recall, chooses formulation (III), because it is, intuitively, simplest, and it is to these conditions that Kim objects. In his example, 'FbvGb' is asserted to be true, but just which disjunct is true is <u>unknown</u>. Although he grants that there exists a C such that (III) is satisfied (because one or the other disjunct must be true), he further maintains that we cannot actually produce such a C and actually <u>exhibit</u> the explanation required.

The same objection, it may be noted, would apply to formulation (I), for (under the conditions of Kim's example) we may not be able to specify the true basic sentences which are to be taken as members of the set K, although we are assured that such a set exists. Kim's point, I take it, is that we may not <u>know</u> it exists.

But now consider formulation (II). As is readily seen, the example Kim produces not only satisfies all of these conditions, but it also itself exhibits the explanation he demands be exhibited. This formulation is thus not susceptible to the criticism that it does not allow one to actually produce or construct the desired explanation.

What has gone wrong here? The three formulations above are demonstrably equivalent, but Kim's example seemingly does not satisfy either (I) or (III), while it does satisfy (II).

The trouble, I believe, is this: Kim has here introduced an epistemological consideration where it does not belong. It is clearly irrelevant to this version of the deductive-nomological model to argue that 'FbvGb' may be known to be true without knowing which of the disjuncts is true. This is simply beside the point. The original Hempel and Oppenheim model was designed to be free from such epistemological matters.

For now, we must rest content with the simple assertion that Kim is mistaken in this regard. In Chapter VI, however, we demonstrate this point more conclusively.

Kim might also be understood as rejecting as too strong a requirement that there be an actual verifying class K. In his example, he does not say what the class K is to be, but such a matter is almost trivial when one adopts the potential verification requirement of Hempel and Oppenheim. Yet it is this revision which is so crucial to Kaplan's argument. In fact, we now show that the failure of Kim's model is directly attributable to his rejection of Kaplan's proposal concerning the verification of the singular sentence C of an explanans.

The following theorem is directly analogous to Kaplan's theorem 4. Let us say that a singular sentence E (which is of the form SvS', with at least S being true and S, S' having no predicates in common) is possibly Kim-explainable by T if there is a singular sentence C such that (T,C) is a Kim-explanans for E with the possible exception that C is true.

> Theorem 4*: If E is possibly Kim-explainable by T, then E is Kim-explainable by T.

Proof: assume that E is possibly Kim-explainable by T. Then there is a set of singular sentences C such that if C were true, (T,C) would be a Kim-explanans for E. It remains only to show that (T, CvS) is a Kim-explanans for E, and for this we need only show that CvS is true. But this follows from the assumption of the truth of S.

Kim was therefore mistaken in thinking that his conditions would suffice in place of Kaplan's. In fact, Kim just perpetrated the same fundamental error as the original Hempel and Oppenheim model.

It seems to me that Kaplan's proposal should be adopted as an adequate revision of the Hempel and Oppenheim model. It is not susceptible to the earlier trivialization theorems and seems natural from the viewpoint of the intent of the original model. However, since the particular formulation which Kaplan proposes is not as amenable to the extension we envisage in Chapter IV, we prefer to fix on formulation II, above, for the rest of this work.

In so doing we must first add another condition to accommodate Kaplan's third revision, since so far formulation II satisfies R1 but not R2. In complete form, Kaplan's proposal comes to this:

(C1) (T,C) is a direct S-explanans for the singular sentence E if and only if
(1) T is a theory
(2) C is singular and true
(3) E is not derivable from any disjunct of the disjunctive normal form of C
(4) $\{T,C\} \vdash E$
(5) there is a class K of basic sentences such that $K \vdash C$ but neither $K \vdash E$ nor $K \vdash \sim T$
(6) E is a disjunction of basic sentences

(C2) (T,C) is an S-explanans for the singular sentence E if and only if
(1) (T,C) is a direct S-explanans for each of a set of singular sentences $\{E_1,\ldots,E_n\}$, and
(2) $\{E_1,\ldots,E_n\} \vdash E$.

In the definition of direct S-explanation the rationale for condition (6) has already been discussed. However, there is yet one more objection voiced by Kim which we must answer to.

Kim objects to the fact that 'FbvGb' is S-explainable by the law '(x)(Fx→Gx)', in conjunction with the singular sentence 'Fb'. On the Hempel and Oppenheim analysis, this amounts to self-explanation, and is eliminated on the grounds that 'FbvGb' follows from the sentence of initial conditions alone.

But it is important to note two things. The above example amounts to an S-explanation of 'FbvGb', but not a direct one. Thus, 'FbvGb' is S-explainable by the above law only in virtue of its derivability from 'Gb', which is directly S-explainable by that law. Second, the reason for rejecting self-explanation should be recalled. Clearly, the reason was that the deductive model of explanation was designed to capture the intuitively desirable requirement that an adequate scientific explanation make essential use of laws. Sentences describing initial conditions and sentences describing events to be explained must be mediated by sentences expressing universal laws. But notice the significant respects in which the following "explanation" differs from the example of the preceeding paragraph:

(C) T: (x)(Fx→Rx)
C: Fa
E: FavTa

The law of (C) is not essential for the derivation of the explanandum; the "explanation" can be carried through without it. But this is not the case with the above S-explanation of 'FbvGb'; for, by definition, the law of that example must be made use of.

To illustrate the effect of Kaplan's proposal, consider the following examples which it deals with in an intuitively desirable manner.

1. T: $(x)(Mx \rightarrow Hx)$
   C: $\underline{MevHe}$
   E: He

This is the 'mermaid' example of Eberle, Kaplan, and Montague, and it is eliminated because one disjunct of the disjunctive normal form of C entails E.

2. T: $(x)(Mx \rightarrow Hx)$
   C: $\underline{MevHe}$
   E: HevPa

This is a Morgan-type 'mermaid' example to Kim's proposal. It is eliminated for the same reason as above.

3. The following example is intended to make good a promissory note offered in Chapter I. There I discussed the reasons for Hempel and Oppenheim's reluctance to prohibit partial self-explanation. That is, an argument of the form

(D) T: $(x)(Px \rightarrow Qx)$
    C: $\underline{Pa\&Rab}$
    E: Qa&Rab

was not eliminated by Hempel and Oppenheim, even though one of the conjuncts of E follows from C alone, for the following reason. If such partial self-explanations were prohibited, then, Hempel and Oppenheim argue, we would be committed to the view that only singular sentences logically derivable from a theory would be explainable by that theory. The rationale behind this assertion is the following. Given the perfectly adequate explanation

$$\text{(E)}\quad \begin{array}{ll} T: & (x)(Px \rightarrow Qx) \\ C: & \underline{\quad Pa \quad} \\ E: & \quad Qa \end{array} ,$$

this could be transformed into the logically equivalent formulation

$$\text{(F)}\quad \begin{array}{ll} T: & (x)(\sim Px v Qx) \\ C: & \underline{(PavQa)\&(Pav\sim Qa)} \\ E: & (PavQa)\&(\sim PavQa) \end{array}$$

According to Hempel and Oppenheim, "this reformulation shows that part of the content of the explanandum is contained in the content of the singular component of the explanans and is, in this sense, explained by itself".[18] For one conjunct of E follows from C alone, while the other conjunct follows from T alone.

As mentioned in Chapter I, Hempel and Oppenheim feel constrained to accept (F) in virtue of its equivalence to (E), and thus make no attempt to rule out (D). Clearly, they are of the opinion that if stipulations are introduced to rule out (D), then (F) will be ruled out as well, which, they think, is quite undesirable.

Yet Kaplan's proposal <u>does</u> accord with our intuitions by adequately distinguishing between (D) and (F). For by

---

18. Hempel, C., <u>Aspects of Scientific Explanation</u>, The Free Press, New York, 1965. p. 275.

the definition given, (D) is neither a direct S-explanation nor an S-explanation; (E) on the other hand constitutes a direct S-explanation of 'Qa'. Now let us consider (F). The essential difficulty here can be removed if we first rewrite this explanation with C being put into disjunctive normal form, as follows:

(F') T: $(x)(Px \rightarrow Qx)$
C: Pa
E: $(Qa \vee Pa) \& (\sim Pa \vee Qa)$

Here, one conjunct of E follows from C alone and the other conjunct follows from T alone. But by the definition of S-explanation, E is explainable by the given theory only because each conjunct is a logical consequence of a singular sentence (viz. 'Qa') which is *directly* S-explainable by that theory. Adoption of R2 allows us to put some restriction on partial self-explanation while at the same time avoiding the consequence that, by doing so, only singular sentences logically derivable from a theory are explainable by that theory.

2.6. *Ackermann's generalized model*

Robert Ackermann has recently argued that a wholly new model of explanation is needed, and he has criticized all proposals so far advanced on the grounds that they are too restrictive in so far as they apply only to *singular* explananda. He feels that no set of conditions so far enumerated allows for any obvious generalization to cover non-singular E sentences, thus precluding from the outset explanations of laws or other generalized sentences.

In this section I will discuss Ackermann's paper "Deductive Scientific Explanation"[19] and its sequel "A Corrected Model of Explanation", by Ackermann and Stenner. The first paper, although defective, provides the basic rationale for Ackermann's approach. I will show that both models are faulty, and argue that Ackermann has failed in his task to provide a generalized model of deductive explanation.

2.6a. Ackermann's misgivings over Hempel-Oppenheim-Kim explanation

In view of the trivialization theorems of Eberle, Kaplan, and Montague, Ackermann, at the time his paper appeared in print, considered the then current status of deductive explanation to be the model proposed by Hempel-Oppenheim-Kim (or rather, Kim's amendment to the Hempel and Oppenheim model). Since it is in contrast to this model that Ackermann proposes his own analysis, his criticisms must be made apparent.

The first point of disagreement, as already mentioned, is with the restriction by Hempel and Oppenheim (and Kim) of the range of E to the set of singular sentences only. In general, he is critical of all models proposed thus far

---

19. Ackermann, R., "Discussion: Deductive Scientific Explanation", Philosophy of Science, 32 (1965), pp. 155-167.

in that they "do not allow for any obvious generalization of their conditions which will encompass the full range of all those scientific explanations which must be considered plausible candidates for translation into deductive models".[20]

A second misgiving over H-O-K explanation arises in connection with the following argument form, which is ruled out by their conditions:

(A) T: $(x)(Mx \rightarrow Nx)$
C: $Ma \& Ea$
E: $Na \& Ea$.

Apart from the fact that the condition which disallows this form (i.e., not $E \vdash C_i$) also precludes the trivialization theorems of Eberle, Kaplan, and Montague, Ackermann feels that it is not clear that explanations of this form should be excluded by any model of explanation that is sufficiently general to cover every plausible case of deductive scientific explanation. He argues that the question of whether the form (A) should or should not be permissible ought to remain open. Of course, he then has to provide conditions which are sufficient to exclude the trivialization theorems from applying to his model.

The reason Ackermann gives for wanting to leave this an open question is this. If we consider, he says, the way that scientific theories seem to develop historically, then at one time we may be able to explain 'Ea' by the

20. ibid., p. 155.

following argument:

$$\text{(B)}\quad \begin{array}{ll} T: & (x)(Mx \rightarrow Ex) \\ C: & \underline{\quad Ma \quad} \\ E: & \quad Ea \end{array}$$

and at a latter time we might be able to introduce the following definitions into our language,

$Mx= (Rx\&Sx)$ and $Ex= (Sx\&Tx)$

thus finding upon further analysis that the properties 'M' and 'E' share a certain other property in common. But if so, the following explanation, which follows by definition from (B), is illicit on the H-O-K analysis:

$$\text{(C)}\quad \begin{array}{ll} T: & (x)(Rx\&Sx \rightarrow Sx\&Tx) \\ C: & \underline{\quad Ra\&Sa \quad} \\ E: & \quad Sa\&Ta \end{array} \quad .$$

Ackermann asks for a more detailed justification of any condition (such as condition (5) of H-O-K explanation) which will show why (C) should be ruled out, other than the ad hoc fact that such a condition avoids the theorems of Eberle, Kaplan, and Montague. He feels that his own conditions avoid these theorems while still preserving (C) as an instance of deductive scientific explanation.

A third misgiving Ackermann has over the proposed models concerns their requirement that the explanandum sentence E be true. He sees it as important to be able to ask, for purposes of investigating the explanatory power of a past theory, whether the past theory would explain E, even though E is now known to be false. This objection, it seems to me, carries very little weight, for in such a case as this we would normally proceed by assuming that E were true, and then ask whether a given theory would explain it, under this supposition. I will therefore make no further comment on this point.

2.6b. Ackermann's initial model

In his first paper, Ackermann sets out the following necessary conditions for deductive scientific explanation:

> A set T of sentences $\{T_1, T_2, \ldots, T_n\}$ $(n \geq 1)$ is a deductive scientific explanation of a sentence E (where neither $\vdash E$ nor $\vdash \sim E$) only if:
> (T1) $T \vdash E$
> (T2) not $T \vdash \sim E$
> (T3) some T_i $(T_i \in T)$ is a theory
> (T4) there is no proper subset T' of the set T* of ultimate conjuncts of T such that $T' \vdash E$.[21]

Most of these conditions are self-explanatory, but (T4) requires some comment. We must first define the notion of "ultimate conjuncts", which Ackermann does as follows (where T is some set of sentences):

> If any sentence of T is equivalent in sentential logic to a conjunction R, of at least two non-equivalent sentences (if the logic in which the members of T are formalized is predicate logic, then one or more applications of quantifier distribution rules may be necessary before the test for equivalence is made in sentential logic), then that sentence may be replaced in T by R so as to produce a new set of conjuncts 'T' with at least one more member than T. This process may be carried out recursively until no more conjunctions can be found in the last produced

---

21. ibid., p. 160.

> set of conjuncts. This set is then the set of ultimate conjuncts of T.[22]

As an example of how Ackermann intends this definition to apply, let T be the set

{(x)(Fx→Gx), ∿(∿Fa∨∿Ha)};

then the set T* of ultimate conjuncts of T would be the set

{(x)(Fx→Gx), Fa, Ha}.

The purpose of (T4) is to exclude explanans sets with superfluous elements and to ensure that the theory of T be used essentially. The reason why (T4) requires that there be no proper subset of T*, and not simply that there be no proper subset of T, from which E may be derived can best be illustrated by the following example. If T were the set

{((x)(Mx→Nx)&Tb), Ma}

and E were the sentence

Na,

then the set T meets the requirement that there be no proper subset of T from which E is derivable, but it contains the superfluous element 'Tb'. When T* is introduced,

T*= {(x)(Mx→Nx), Tb, Ma}

it can be seen that there is a proper subset of T* from which E can be derived, and thus the T of this example is ruled out as a suitable explanans.

One may well puzzle about the need for such a condition as (T4). It hardly seems necessary to exclude sentences from

---

22. ibid., p. 161.

an explanans which are superfluous; we are in search of a general definition of "explanation", not what might be called "the most elegant explanation", or "the most complete explanation". Ackermann's reason, though he does not say so, seems to be that he wants to avoid a previously mentioned difficulty. Recall why Hempel and Oppenheim limited their definition to singular E sentences. To explain Kapler's law (K), we could take as an explanans 'K&B' (where 'B' is any other law), and on the basis of the Hempel and Oppenheim model, this would be an acceptable explanation. Ackermann's condition rules this out; but I argue in the subsequent chapter that we can avoid this result without introducing anything like Ackermann's (T4).

Before introducing Ackermann's last two conditions it will be necessary to introduce some additional definitions. We must first understand the intended meaning of the phrase "truth-functional components of T".

> Let T be the explanans set, and let A(T) be the set of truth functional sentential components of T. If any sentence R of T is shown by sentential logic (after quantifier distribution if predicate logic is also involved) to be a simple truth function of two or more other sentences (for example, R may be the disjunction of two sentences, or a conditional with one sentence as antecedent and the other as consequent, or a negation of another sentence), then these other sentences (or this other sentence in the case of negation) are admitted as members to the set T in order to form a superset T' of T. If any sentence of this superset is a simple truth function of other sentences at least one of which is not

> provably equivalent to some member of the superset, then a new superset of T is formed by adding the appropriate sentence as members. When no further sentences can be added to the largest superset of T formed by this procedure, this largest set is A(T).[23]

To illustrate an application of this definition, simply let

$T= \{(x)(Mx),\ Ma \rightarrow Sa\}$,

and then

$A(T)= \{(x)(Mx),\ Ma \rightarrow Sa,\ Ma,\ Sa\}$.

Secondly, we must define "I-development":

> Let W be the set of w distinct individual constants occurring essentially in E, and let T* ... be the set of ultimate conjuncts of T. (If w=0, then some arbitrary individual constant in the range of the quantifiers ... is taken as the sole member of W.) Let Q be any element of T* in prenex normal form with the longest possible initial string of universal quantifiers. If there are n universal quantifiers ($n \geq 0$) in this initial string, let the I-development of Q be the w^n sentences obtainable by dropping the n initial universal quantifiers and replacing the occurrences of the variable of each (which may then become free) by any of the w individual constants in W. The I-development of T is the set obtained by replacing each element of T* with its I-development.... The I-development of any element Q of T* which has no universal quantifier as its first quantifier, or no quantifiers at all in appropriate prenex normal form, is identical with Q. The I-development of any set of sentences with respect to the set W is carried out in a similar manner.[24]

We may now state Ackermann's last two conditions.

> (T5) Let T* be (as above) the set of ultimate conjuncts of T, and let A*(T) be (as above) the set of all truth functional components of T along with their negations. It must not be possible to

---

23. ibid., p. 162.
24. ibid., p. 165.

construct a set C whose elements are chosen from the elements of A*(T) such that not C⊢ E, T* and C are consistent, but some subset P of the union of T* and C satisfies either of the following conditions, A or B:
A. P⊢ E, but every occurrence of E or a sentence equivalent to it in P can be replaced by a sentence S not an element of A*(T) so that S can be derived from P(S), the result of this replacement.
B. Some subset K of the I-development of P with respect to W, the set of individual constants occurring essentially in E, is provably equivalent to the I-development of the set {E}.

(T6) If T⊢ E and A(E) is the set of truth functional sentential components of E, then there must not be a sentence F which is not a member of A(T) which can replace every occurrence of sentences equivalent to that sentence, so that the result of the described replacement can be derived from T.[25]

Some discussion of these final conditions is in order. (T5)A is what Ackermann calls the trivialization principle. This principle rules out the following form:

(D) T: (x)(Mx)
C: $Ma \rightarrow S$
E: S

This ought to be ruled out, according to Ackermann, because if it were an acceptable explanation, then almost any other sentence could be substituted for 'S' with the result being an acceptable explanation.

(T5)B incorporates what Ackermann calls the redundancy principle, and is arrived at by way of the following consideration. In

25. ibid., pp. 165-166.

(E) $\dfrac{(x)(Tx)}{Ta}$

he claims, this form can be augmented to qualify as an H-O-K explanation by taking any singular sentence we please (except Ta) as initial conditions. He then argues as follows:

> But there is surely some question as to whether a purported explanation, whose explanandum is simply an instance of a universally generalized theory occurring in the explanans, may be construed as a candidate for sound scientific explanation. If one takes seriously the view that theories are not legitimately taken to be just logical constructions out of data stateements because they are not simply compendious descriptions of data, some reflection of this in our stipulations should be sought....
>
> ... This I take to be the root of the difficulty. In a loose way, one may express this by saying that if an explanans and an explanandum are logically equivalent in the domain of the individuals specified and essentially mentioned in the explanandum sentence, then the explanans is redundant and the putative explanation faulty.[26]

This principle is also designed to eliminate the "mermaid" example of Eberle, Kaplan, and Montague, i.e.,

$$\frac{\begin{array}{c}(x)(Mx\rightarrow Hx)\\ Me\vee He\end{array}}{He},$$

as the following illustrates. '(x)(Mx→Hx)' is equivalent to 'Me→He' in the domain of singleton 'e', but 'Me→He' and 'MevHe' are together equivalent to 'He' in this domain.

Finally, (T6) is introduced to rule out having 'EvP'

26. ibid., p. 163.

explainable by T, if E alone is explainable by T. This, in part, allows for the elimination of certain arguments that can be constructed by the use of the rule of addition, but which would otherwise contravene (T1)-(T5).

Ackermann asserts that (T1)-(T6) are necessary and sufficient for deductive scientific explanation. Before considering the corrected version given by Ackermann and Stenner, let us see just what inadequacies reside in the present model.

One inadequacy has been pointed out by Alfred Stenner, leading he and Ackermann to make a new proposal. Consider the following example, usually considered to be a paradigm form of explanation by writers on the subject:

$$
\text{(F)}\quad \begin{array}{ll} \text{T:} & (x)(Fx \rightarrow Gx) \\ \text{C:} & Fa \\ \hline \text{E:} & Ga \end{array}
$$

It can be shown that this example violates (T5)A. T* here is the set {(x)(Fx→Gx), Fa}, A*(T) is the four element set {(x)(Fx→Gx), ∿(x)(Fx→Gx), Fa, ∿Fa}, and we may take for C the set {Fa}. Now, not {Fa}⊢ Ga, T*∪C=T*, and choose as P the set {(x)(Fx→Gx), Fa}. (T5)A says that if we can replace in P every occurrence of E by a sentence S not an element of A*(T) (take '(Ex)(Gx)' for S), and such that from P(S) (P with S substituted for E) E can be derived, then the original set T from which T* was constructed is not an explanans set for E, thus ruling out (F).

This argument is also eliminated by (T6), taking F to be '(Ex)(Gx)'.

Charles Morgan, in an as yet unpublished paper and also in several conversations, has pointed out other inadequacies in the Ackermann analysis. In particular, he notes that by the definition of "ultimate conjuncts" given by Ackermann, all of (T1)-(T6) cannot be simultaneously satisfied. For by that definition, there will always be included in the set of ultimate conjuncts some arbitrary logical truth, and thus (T4) will always be violated. As a simple illustration, take the example given above, i.e.,

$$T = \{(x)(Fx \rightarrow Gx),\ Fa\}$$

and look at the first sentence in the definition of "ultimate conjuncts". It says that if any sentence which is in T is equivalent to a conjunction R of two non-equivalent sentences, then that sentence may be replaced in T by R. Now 'Fa' is a member of T, and

$$Fa \equiv (Fa \& (Pa \vee \sim Pa)),$$

so we may replace 'Fa' in T to obtain the following set of ultimate conjuncts:

$$T^* = \{(x)(Fx \rightarrow Gx),\ Fa,\ (Pa \vee \sim Pa)\}.$$

Now if we look at (T4), which requires that there be no subset T' of the set of ultimate conjuncts of T such that $T' \vdash E$, we see clearly that this condition is violated. (In the general case we must assume that T is consistent, so this result should perhaps be stated as follows: If (T1) is satisfied, then by the definition of "ultimate conjuncts" (T4) is violated.)

There are other difficulties with the Ackermann proposal, and these are pointed out by Professor Morgan in the above mentioned unpublished paper. But there is no need to consider these in any more detail. Let us look instead at the 'corrected' version offered by Ackermann and Stenner.

2.6c. Ackermann, Stenner, and a corrected model

Instead of patching up the defective version, Ackermann and Stenner present a revised model which is intended to replace the old one. Before stating the conditions of this model, however, it might be wise to spell out some new definitions provided by the authors, which lead to some ambiguity. In fact, Professor Morgan seems to have misconstrued the intent of these definitions in the aforementioned unpublished work.

The authors first define anew the notion of "truth-functional component", as follows:

> (I) A sequence of statemental wffs $(W_1, W_2, \ldots, W_n)$ of an appropriate language L is a sequence of truth functional components of T if and only if T may be built up from the sequence by the formation rules of L, each member of the sequence being used just once in the application of the rules in question. The W_i are thus construed as tokens, in that (Ga,Ga) but not (Ga) is a sequence of truth functional components of the wff Ga&Ga.[27]

---

27. Ackermann, R., and Stenner, A., "Discussion: A Corrected Model of Explanation", Philosophy of Science, 33 (1966), p. 169.

The next definition is that of "ultimate sentential conjunct";

> (II) A set of ultimate sentential conjuncts T_0 of a sentence T is any set whose members are the wffs of a longest sequence $(W_1, W_2, \ldots, W_n)$ of truth functional components of T such that T and $W_1 \& W_2 \& \ldots \& W_n$ (the conjunction of the W_i) are provably equivalent. It T is a theory, then this definition holds with the stipulation that each of the W_i is also a theory. Further, if T is a set of sentences, the set of ultimate conjuncts T_c of T is the union of the sets of ultimate sentential conjuncts of each member of T.[28]

Morgan claims that from the discussion above and that of (I), it follows that not every sentence has a set of ultimate sentential conjuncts. He argues as follows:

> Let T be PavPb. Although (I) has no clause which literally requires that no member of a set of truth functional components of some sentence be such that it cannot be further broken down, such an assumption is certainly made for the first model and seems reasonable here. Thus it seems that the set of truth functional components of T is (Pa,Pb); but T is not equivalent to Pa&Pb. If we allow as a set of truth functional components of T the sequence (PavPb), then the same sequence could serve as the set of ultimate sentential conjuncts of T. But this set would seem to be ruled out by the requirement that we choose a longest sequence of truth functional components of T, and the fact that (on at least one reasonable interpretation of 'longest sequence') the sequence (Pa,Pb) is longer than the sequence (PavPb).[29]

I agree with Morgan that the discussion of Ackermann and Stenner is confusing, but I think the confusion can be cleared up by taking note of a footnote which Ackermann and

---

28. ibid.
29. Morgan, C., "On Two Proposed Models of Explanation", <u>Philosophy of Science</u>, forthcoming.

Stenner add to the discussion in definition (I). This note runs as follows: "A set of truth functional components of T has as members all wffs of some sequence of truth functional components of T as well as all wffs constructed from these wffs by the formation rules in the construction of T..."[30] I think that Morgan has overlooked this addition. For given this addendum, we can, I think, make sense of the example that Morgan produces. Morgan asks: What is the set of ultimate sentential conjuncts of T= PavPb? Well, by (I) (Pa,Pb) is one sequence of truth functional components of T. But is this sequence going to count as the members of the set of ultimate conjuncts of T (rather, the members of this sequence)? No, for by definition (II), we are instructed to choose that sequence the conjunction of the members of which is equivalent to T. And, as Morgan notes, PavPb is not equivalent to the conjunction Pa&Pb.

Here is another sequence of truth functional components of T, this time taking into consideration the footnote: (Pa,Pb,PavPb). Is this going to count as the set of ultimate conjuncts of T? No, for again the conjunction of the members of this sequence is not equivalent to T. But now consider the sequence (PavPb). As it turns out, this is the longest sequence of truth functional components of T such that the conjunction of this sequence is equivalent to T.

---

30. Ackermann and Stenner, p. 169.

(It happens that it is a conjunction of only one member, but that does not matter.)

I think that this clears up the issue. We might also note that definitions (I) and (II) no longer have the drawback of the original definitions, which led to the requirement that the set of ultimate sentential conjuncts of a set T of sentences contain some arbitrary logical truth.

Now we may turn to the new set of conditions, which I state in one fell swoop:

> Let T be a set of sentences constituting a putative explanans and E a sentence which is a putative explanandum for T. It is assumed that not $\vdash$ E and not $\vdash \sim$E. Construct a set T_c of ultimate sentential conjuncts of T. Then T is an explanation of E if and only if:
> (1) $T_c \vdash E$
> (2) not $T_c \vdash \sim E$
> (3) Some T_{ci} is a theory (T_{ci} is an arbitrary element of T_c)
> (4) There is no proper subset T_c' of T_c such that $T_c' \vdash E$
> (5) It is not possible to construct a set $R = \{R_1, R_2, \ldots, R_n\}$ such that for each sentence R_i of R some T_{ci} is such that either $\sim T_{ci} \vdash R_i$ or $\sim T_{ci} \Vdash \sim R_i$, not $R \vdash E$, but there is a subset T_c'' of T_c such that $T_c'' \cup R$ is consistent, and $T_c'' \cup R$ yields E by sentential logic.
> (6) If E can be appropriately interpreted in a smallest domain of n individuals ($n \geq 1$), and each member of T_c can also be so interpreted in this domain..., then it is not the case that under this interpretation some subset of T_c is provably equivalent to E.
> (7) There is no sentence F such that $F \vdash E$, F (taken as the E of (1)-(6)) satisfies (1)-(6) with respect to T, and for some F_w, $T \dashv\vdash F_w$.[31]

---

31. ibid., p. 170.

Conditions (5) and (7) call for some discussion. Let us take each of these in order.

Condition (5) replaces old (T5)A, the triviality principle. Let us see how this condition functions by considering the example which led to the introduction of the principle in the first place. (T5)A was designed to eliminate the following:

$$\frac{\begin{array}{l}(x)(Mx)\\ \quad Ma\rightarrow E\end{array}}{E}\ .$$

New condition (5) will also rule this out, as the following demonstrates:

$T_c = \{(x)(Mx),\ (Ma\rightarrow Ea)\}$

$R = \{Ma\}$

$\sim T_{ci} \vdash Ma$ (i.e., $\sim(Ma\rightarrow E) \vdash Ma$)

$T_c'' = \{(Ma\rightarrow E)\}$

$T_c'' \cup R \vdash E$.

The counter-example brought forth by Stenner no longer holds (c.f. the preceeding section), as that example makes use of the rule of existential generalization, and the new condition requires that $T_c''\cup R$ yields E _by sentential logic_.

Condition (7) replaces defective (T6). This condition depends upon a prior definition, which is as follows:

> (III) Let W be a truth functional component of F. F_W is constructed by the same sequence of applications of formation rules to a sequence containing W that will construct F except that the negation of W is substituted

> for W in the steps of the construction. If GavTa is F and Ta is W, then Gav∿Ta is F_W.[32]

Now old (T6) was introduced to eliminate the following example:

$$\frac{\begin{array}{c}(x)(Px\rightarrow Qx)\\ Pa\end{array}}{QavTa}\ .$$

It did that indeed, but it also eliminated

$$\frac{\begin{array}{c}(x)(Px\rightarrow Qx)\\ Pa\end{array}}{Qa}\ .$$

To see this, simply take as the F of (T6) the sentence '(Ex)(Qx)', and the above is ruled out.

New condition (7) avoids both difficulties. For example, take the first illustration above. If we take for F the sentence 'QavTa', then

QavTa ⊢ QavTa

QavTa satisfies (1)-(6) with respect to T, and

T ⊢ F_W (i.e., T ⊢ Qav∿Ta).

I want now to show that the revised model fares no better than the first.

1. Condition (7) is inadequate, as it still rules out

$$\frac{\begin{array}{c}(x)(Px\rightarrow Qx)\\ Pa\end{array}}{Qa}\ .$$

Simply take as the F of that condition the sentence 'Qav(x)(Qx)' and then the following hold:

32. ibid., p. 169.

$F \vdash E$

F satisfies (1)-(6) with respect to T, and

$T \vdash Qa \vee \sim(x)(Qx)$ (i.e., $T \vdash F_w$).

2. Charles Morgan has shown that the conditions above cannot all be simultaneously satisfied by <u>any</u> set of sentences T and E. In his unpublished work,[33] he proves a general theorem to the effect that, if conditions (1) and (2) are satisfied, then condition (5) cannot be satisfied. Let me just provide an instance of what Morgan proves in general.

Take, once again,

$$\frac{\begin{array}{c}(x)(Px \rightarrow Qx)\\ Pa\end{array}}{Qa}\ .$$

Conditions (1) and (2) are satisfied; I shall now show that (5) is violated. Let

$R = \{(Pa \rightarrow Qa)\}$.

Now

$\sim Pa \vdash (Pa \rightarrow Qa)$

not $(Pa \rightarrow Qa) \vdash Qa$;

let

$T_c'' = \{Pa\}$;

then

$T_c'' \cup R$ is consistent, and

$T_c'' \cup R \vdash Qa$.

---

33. In "On Two Proposed Models of Explanation".

This result is conclusive proof that the Ackermann-Stenner model is inadequate.

3. In 1. above I have shown that condition (7), in its present form, is inadequate. Perhaps we can amend it, however, and remove the defect. How might this be done? My counter-example depended upon taking for the F of that condition a sentence which was logically equivalent to the explanandum. This suggests revising (7) to read as follows:

(7*) There is no sentence F such that $\{F\} \vdash E$ but not $\{E\} \vdash F$, ...

Morgan has shown that such a revision will not remove the fundamental difficulty contained in condition (7). Again, he proves a general theorem demonstrating this fact, but we need only consider an instance of this theorem.

Consider, then, the following example (due to Charles Morgan).

$$\begin{array}{ll} T\text{:} & (x)((Px\&Qx)\rightarrow Sx) \\ C\text{:} & Pa\&Qa \\ \hline E\text{:} & Sa \end{array}$$

Suppose further that this example satisfies (1)-(6). Now it can be shown that this example fails to satisfy (7*). For let

$F = (Pa\&((Pa \vee Ra)\rightarrow Sa))$,

and let

$W = Ra$;

then (7*) is violated.

2.7. Omer and an 'information' model

In a recent paper entitled "On the D-N Model of Scientific Explanation", I. A. Omer proposes a new model of explanation which supposedly exhausts the insights to be found in the literature and which is also said to be justified in terms of the rationale of scientific inquiry. This rationale is based upon what Omer calls "certain principles of assertive discourse", such as that one should not deliberately make a less informative statement on a topic when a more informative one could be made. To achieve the end of basing a model of explanation on such principles, Omer proposes that the following conditions be satisfied by any sound model of deductive explanation.

> R_I: In the explanans, no sentence which is less informative in the topic should be given when it is possible to give a more informative one.
>
> R_{IS}: No sentence in the explanans should be of less informative content than the explanandum.[34]

In order to compare the informative content of sentences, Omer draws upon the work of Popper and Carnap, stating that for his purposes, the following general conclusions will suffice:

(a) The logical content of a sentence is the class of all sentences which are consequences of (L-derivable from) the given sentence and not logically true.

(b) If an empirical sentence q is derivable from

---

34. Omer, I.A., "On the D-N Model of Scientific Explanation", Philosophy of Science, 37 (1970), p. 419.

an empirical sentence p and not the reverse, the empirical content of q is less than that of p.

(c) If no relation of derivation exists between two sentences then their contents are noncomparable.[35]

On the basis of these results, Omer rewrites his above principles as follows:

R_I': In the explanans no sentence on the topic which is of less logical content should be given when it is possible to give a sentence with more logical content.

R_{IS}': No sentence in the explanans should be of less logical content than the explanandum.[36]

The importance of these principles will be seen immediately. For example, the 'mermaid' example of Eberle, Kaplan, and Montague, i.e.,

$$\begin{array}{ll} T: & (x)(Mx \leftrightarrow Hx) \\ C: & \underline{\quad Me \vee He \quad} \\ E: & \quad He \end{array}$$,

will be ruled out by R_{IS}', in virtue of the fact that $E \vdash C$, which means that a sentence in the explanans is of less logical content than a sentence in the explanandum. Furthermore, Omer claims, this example is not ruled out in the ad hoc manner of Kim, but in terms of certain principles of assertive discourse.

In accordance with these principles and on the basis of examples like the above, Omer is led to propose the following set of conditions (necessary conditions) for a

35. ibid., p. 421.
36. ibid., p. 422.

sound model of deductive explanation:

> The set T of supposedly true sentences constitutes an explanans for the supposedly true sentence E (where neither $\vdash E$ nor $\vdash \sim E$) only if:
> (1) $T_c \vdash E$
> (2) not $T_c \vdash \sim E$
> (3) some T_{ci} $(T_{ci} \varepsilon T_c)$ is a universal law
> (4) for any T_{ci} $(T_{ci} \varepsilon T_c)$, T_{ci} is noncomparable with E.[37]

'T_c' here refers to the set of ultimate sentential conjuncts of T, as defined by Ackermann and discussed previously.

To arrive at his final condition, Omer finds that R_I' must be revised. He argues to the effect that although we want informative sentences, "we don't want the scientist to recite all the knowledge he has irrespective of the fact that the information so conveyed is useful or useless.... Thus a requirement to rule out redundant elements would be in line with the attitude of the scientist assumed in this paper".[38] Omer's final condition, then, will be one which is designed to rule out not only redundant elements of an explanans set, but also redundant information. What Omer means by "redundant information" is the following:

> It has been shown that if a sentence T' is implied by another sentence T, then the informative content of T is greater than that of T'.... From this fact we directly see that if we use the sentence T (in conjunction with some other sentences) to explain a sentence E which is explainable

---

37. ibid., p. 424.
38. ibid., p. 425.

> by T' (in conjunction with those other sentences) then we will have some redundant information.[39]

To see how Omer <u>intends</u> to apply the principle which is proposed here, let us consider some of his examples. In

$$(1^*)\quad \frac{\begin{array}{c}(x)(Fx\rightarrow Gx)\\ Fa\end{array}}{Ga\vee Ha}\;,$$

there is, he says, some redundant information. For we could replace the law of this example with '(x)(Fx→GxvHx)', which is weaker than the law of (1*), and true because derivable from it. Thus, on his account, we must rewrite (1*) in the following way:

$$(2^*)\quad \frac{\begin{array}{c}(x)(Fx\rightarrow Gx\vee Hx)\\ Fa\end{array}}{Ga\vee Ha}\;.$$

Another example of eliminating redundancy is the following:

$$(3^*)\quad \frac{\begin{array}{c}(x)(Fx\rightarrow Gx)\\ (x)(Gx\rightarrow Hx)\\ Fa\end{array}}{Ha}\;.$$

This he feels should be rewritten as

$$(4^*)\quad \frac{\begin{array}{c}(x)(Fx\rightarrow Hx)\\ Fa\end{array}}{Ha}\;.$$

These considerations lead Omer to his last condition, which is the following:

> (5') It is not possible to find sentences $S_i \ldots . S_r$ $(r \geq 1)$ such that for some

---

39. ibid., p. 426.

$T_{ci} \ldots T_{cn}$ $(n \geq 1)$: $T_{ci} \& \ldots \& T_{cn} \vdash S_i \& \ldots \& S_r$, not $S_i \& \ldots \& S_r \vdash T_{ci} \& \ldots \& T_{cn}$, and on replacing $T_{ci} \ldots T_{cn}$ in T_c by $S_i \ldots S_r$, $T_{cs} \vdash E$. (T_{cs} is the result of this replacement.)[40]

He concludes this section of his paper with the claim that (1), (2), (3), (4) and (5') are necessary and sufficient for deductive scientific explanation.

However, the addition of (5') has untoward consequences, not the least of which is the result that there are no explanations at all! That is to say, Omer's conditions are not simultaneously satisfiable; for in satisfying (5'), we will always run afoul of condition (3). Consider the following satisfactory explanation (with a suitable interpretation being assumed):

$$\frac{\begin{array}{c}(x)(Fx \rightarrow Gx)\\ Fa\end{array}}{Ga}\ .$$

This form does not satisfy (5') and is thus rejected; for let

$T_c = (x)(Fx \rightarrow Gx), Fa$,

$T_{ci} = (x)(Fx \rightarrow Gx)$, and

$S_i = (Fa \rightarrow Ga)$.

Then

$T_{ci} \vdash S_i$

not $S_i \vdash T_{ci}$, and

$T_{cs} \vdash E$.

40. ibid.

(5') is clearly undesirable, since its addition to the other conditions makes the set unsatisfiable. In the example above we have found that, to satisfy (5'), we must violate (3).

Presumably, Omer intended the last clause of (5') to read as follows: "... and on replacing $T_{ci} \ldots T_{cn}$ in T_c by $S_i \ldots S_r$, T_{cs} satisfies (1)-(4) with respect to E." Even this, however, does not resolve the fundamental difficulty; for the following is still ruled out.

$$\frac{\begin{array}{c}(x)(Px \rightarrow Qx)\\ Pa\end{array}}{Qa}$$

To see this, take

$T_{ci} = Pa.$

Then

$T_{ci} \vdash Pa \vee (x)(Qx)$ (i.e., $T_{ci} \vdash S$)

not $Pa \vee (x)(Qx) \vdash T_{ci}$

and on replacing T_{ci} in T_c by S, $T_{cs} \vdash E$.

(For this example I am indebted to Charles Morgan.)

There is one further problem here that I would like to point out. It can be seen, I think, that the trouble with (5') lies not in its formulation, but in its rationale. Consider the following example, which is a Morgan-type 'mermaid' counter-example to the Kim model:

$$(C^*) \quad \frac{\begin{array}{c}(x)(Mx \rightarrow Hx)\\ \sim Me \rightarrow He\end{array}}{He \vee Ee} \quad .$$

To refresh the reader's memory, the sentence "the Eiffel Tower is a good conductor of heat or the Eiffel Tower is a good conductor of electricity" is explainable, on Kim's conditions, by the theory "all mermaids are good conductors of heat". Now let us suppose that (5') does not fact the problem pointed out above. Then (C*) satisfies all of Omer's conditions except (5'). According to the justification given for this condition, we should say that (C*) is unsatisfactory _because the law of this example contains some redundant information_! (5') does not _eliminate_ the irrelevant law of (C*), but only instructs us (in the sense in which Omer intended it) to find a law which is a consequence of this law in order to remove what Omer has called "redundancy". The following 'explanation', however, satisfies all of Omer's conditions (given what Omer had intended by these conditions), with (5') being revised as above suggested:

$$(C^{**})\quad \frac{(x)(Mx \rightarrow Hx v Ex) \quad Me v He}{He v Ee}$$

Clearly, the theory "all mermaids are good conductors of heat or good conductors of electricity" does not help to explain why the Eiffel Tower is a good conductor of heat or a good conductor of electricity. Therefore, even ignoring the other difficulties with (5') and just considering its rationale, it still does not do the required job.

Chapter III

Morgan on Explanation

In an unpublished paper[1] Charles Morgan sets forth a "formal language model of explanation" which, he claims, "is both satisfiable and non-trivial". In Chapter II we noted that the Ackermann and Stenner models, as well as the Omer model, were unsatisfactory in the sense that no set of sentences T and explanandum E could satisfy the conditions proposed. We have also discussed two models which were trivial (viz. that of Hempel and Oppenheim and that of Kim), in the sense that a relation of explanation could be shown to exist between almost any theory and almost any true singular sentence. Morgan thus sets himself the task of producing a model which avoids both types of defects

Despite the claims that Morgan makes for his model (e.g., "the model I have presented is (at least) a reasonable first step in a syntactical characterization of explanation- it is a vast improvement over those models

---

1. Morgan, C., _A Formal Language Model of Explanation_. This paper was read as the main paper in one session of the Philosophy of Science section of the 1972 Canadian Philosophical Association meetings. Other participants in the program, beside myself, were Professors Paul Churchland and Leon Ellsworth.

offered by other authors"[2]), it fails to live up to these claims. In this chapter I show that his model is satisfied by certain argument forms which ought not to be granted the status of legitimate explanations, and is in this sense inadequate.

Morgan's error, it will be seen, is due in part to his not paying sufficient attention to David Kaplan's proposal. As argued in Chapter II, Kaplan's model is entirely adequate as a definition of explanation for singular sentences. I show that Morgan's model will not rule out certain examples Kaplan's analysis does rule out, and indeed which ought to be ruled out on intuitive grounds.

3.1. Morgan's model

Morgan attempts to determine the limits of a syntactical characterization of explanation. To be sure, the syntactical restrictions will be motivated by certain semantical considerations, but the final set of conditions which he imposes will all relate to the form of explanatory arguments. There is, therefore, no requirement that the sentences of the explanans be true.

To avoid misunderstanding, let me emphasize that Morgan's approach is simply a preliminary attempt to provide the necessary and sufficient conditions relating to the

---

2. ibid., pp. 31-32.

syntax of explanatory arguments. His account might best be construed as a definition of "potential explanation", in the sense of Hempel. It would thus be premature and beside the point to criticize Morgan for not providing a semantics, i.e., for not requiring that an explanatory argument be at least sound. There are some argument forms, however, which when given an interpretation turn out to be clearly undesirable, and which Morgan's syntactical restrictions license. This will become evident once we get Morgan's model before us.

The language for which the model is defined is assumed to be the language L characterized by Hempel and Oppenheim.[3] Thus, in what follows, terms such as "sentence", "theory", etc., will always be understood as "sentence of L", "theory of L". etc.

Let T be a set of sentences of L, and let E be a sentence of L. In order for T to be an explanans for E, Morgan first proposes that the following two conditions be taken as necessary:

1. Not $\vdash$ E.
2. Not T $\vdash \sim$E.

Condition 1 requires that E not be provable (or logically true), condition 2 requires that E not be inconsistent and that T be consistent. E must therefore be contingent.

---

3. see Hempel, _Aspects_, p. 270.

Morgan's next condition amounts to a restriction on the relation of logical derivability. He argues as follows:

> All of the other models require that E be derivable from T. But E cannot be just any logical consequence of T if T is to be an explanans for E. Suppose that T is an explanans for S_1, but that T says nothing about- is irrelevant to- S_2. It seems somehow odd to say that T is an explanans for $S_1 v S_2$. But, if $T \vdash S_1$, then trivially $T \vdash S_1 v S_2$. Indeed, if we accept arbitrary disjunctions in this fashion, then severe problems are encountered. For example, Kim's model failed because his conditions could be contravened by taking certain arbitrary disjunctions for E.....[4]

Morgan thus introduces the notion of "direct consequence", which I elaborate upon once certain preliminary matters are dealt with.

Morgan first defines a notion similar to one of Ackermann's, namely, the notion of the "truth functional components" of a sentence. This is done as follows. Let S be any sentence of L. Then S_t is to denote the set of what Morgan calls "TF-components" of S, a set which is formed in the following way:

1. Using quantifier distribution on S, make the scope of each quantifier occurring in S, if there are any, as small as possible (in the sense of having as few occurrences of the truth-functional connectives and atomic formulas as possible). Let this new sentence

---

4. Morgan, p. 2.

be S'. If S contains no quantifiers, or if the scope of each quantifier occurring in S is as small as possible, then S' will simply be the same as S.

2. Let $i=1$, and let $S_{ti}=\{S'\}$.
3. Let $i=i+1$.
4. Form S_{ti} as follows: For each member of $S_{t(i-1)}$, say S_j, if S_j is a truth-functional compound of two shorter sentences (shorter in the sense of having fewer truth-fucntional connectives), say S_{j1} and S_{j2}, then: let $S_{j1} \varepsilon S_{ti}$ unless S_{j1} is equivalent to some sentence which has already been admitted to S_{ti}; and let $S_{j2} \varepsilon S_{ti}$ unless S_{j2} is equivalent to some sentence which has already been admitted to S_{ti}. If S_j is a truth-functional compound of one shorter sentence, S_{j1}, then $S_{j1} \varepsilon S_{ti}$ unless S_{j1} is equivalent to some sentence which has already been admitted to S_{ti}. If S_j is not a truth-functional compound of shorter sentences, then let $S_j \varepsilon S_{ti}$ unless S_j is equivalent to some sentence which has already been admitted to S_{ti}.
5. If $S_{ti} \neq S_{t(i-1)}$, go to step 3, above. Otherwise go to step 6 below.

6. Set S_t=S_{ti} and terminate the procedure.[5]

Let me just provide a brief illustration of what this program is designed to do. It is essentially just a more precise way of specifying the same thing as Ackermann had defined and which we have discussed in the preceeding chapter. For example, if we started with the sentence '(x)(Sx&(Pav Pa))', we would arrive at the following set of distinct truth-functional components: {(x)(Sx), Pa}.

Morgan next gives a definition of "the shortest disjunctive normal form of a sentence S", utilizing the above definition of "TF-components":

> A disjunctive normal form of a sentence S with respect to a non-empty set of sentences S_q is a disjunction of sentences $P=P_1 v \ldots v P_m$ where each P_i is a conjunction of members and/or negations of members of S_q and such that P is logically equivalent to S. The set of shortest disjunctive normal forms of S with respect to S_q is a subset S_r of the set of all disjunctive normal forms of S with respect to S_q, say $S_r = \{S_{r1}, \ldots, S_{rk}\}$, such that the S_{ri} all have the same number, say g, of disjuncts and such that if S_p is any disjunctive normal form of S with respect to S_q and $S_p \notin S_r$, then S_p has more than g disjuncts. Now, if S_r has only one member, the shortest disjunctive normal form of S with respect to S_q is that one member. If S_r has more than one member, then we

5. ibid., pp. 3-4.

may choose one member of S_r by employing some Godel numbering and taking the member of S_r with the lowest Godel number. Finally, the shortest disjunctive normal form of S with respect to S_t is the shortest disjunctive normal form of S, as described above, taking S_q to be the set of TF-components of S.[6]

Given these notions, Morgan is now able to introduce the definition of "direct consequence", as follows. Let T be a set of sentences of L, and let E be a sentence of L such that $T \vdash E$. Let $E_1 v \ldots vE_n$ be the shortest disjunctive normal form of E with respect to E_t. E is a direct consequence of T, $T \vdash_d E$, if and only if there is no disjunct E_i of the shortest disjunctive normal form of E with respect to E_t such that not $T \vdash E_i$ but $T \vdash E_1 v \ldots vE_{i-1} vE_{i+1} v \ldots vE_n$.

Morgan's third requirement then, is that E be a direct consequence of T as a necessary condition for "T is an explanans for E".

I want to pause for a moment to consider Morgan's reasons for introducing this notion of "direct consequence", reasons which I have reproduced above. I do not find it odd, as Morgan does, to say that if T is an explanans for S_1, then T is an explanans for $S_1 vS_2$. Morgan provides no argument against this; instead, he simply remarks that relying on such a principle led to the trivialization of

6. ibid., p. 5.

Kim's model. However, there are other conditions which can be introduced to avoid the problems which confronted the Kim model, but which do not call for any restriction such as the one Morgan introduces. I shall elaborate on this point later.

To summarize the net import of the last several paragraphs, it may be wise to simply state Morgan's third condition:

3. $T \vdash_d E$.

I want now to illustrate the effect of this condition. In doing so, recall Morgan's counter-example to Kim's proposal, discussed in Chapter II. For ease of reference, I state this example below.

(A) $$\frac{\begin{array}{c}(x)(Mx \rightarrow Hx)\\ Me \vee He\end{array}}{He \vee Pa}$$

This, of course, is the notorious "mermaid" example of Eberle, Kaplan, and Montague, with an irrelevant disjunct tacked onto the explanandum. By tacking on such a disjunct, Kim's condition that E not entail C is satisfied, and yet a clear cut example of an illicit explanation is now counted as explanatory on his analysis. This restriction, you will recall (i.e., that E not entail C), was introduced to avoid the original "mermaid" example, given below.

(B) $$\frac{\begin{array}{c}(x)(Mx \rightarrow Hx)\\ Me \vee He\end{array}}{He}$$

Now let us see what work is to be done by the notion of "direct consequence" with regard to (A). In this example, 'HevPa' is not a direct consequence of the given explanans, as the following demonstrates. By definition, there must be no disjunct of 'HevPa' such that that disjunct is not a consequence of the explanans set, but such that all the other disjuncts <u>are</u> consequences of that set. Thus we have, not T ⊢Pa, but T ⊢He, and 'HevPa' is not a direct consequence of T. As already noted, this will eliminate (A) simply because it was in light of examples such as this that Morgan was led to introduce the concept. But this is not yet the correct condition. I will show in what respects this is so in the next sub-section, in which my critical remarks are contained.

I would like to comment in passing that this discussion illustrates my earlier point about the semantical motivation for the syntactical restrictions imposed by Morgan. It is because both (A) and (B) have interpretations which make them absurd as patterns of explanation that syntactical restrictions are introduced to eliminate them.

Morgan's fourth condition is introduced to rule out "needless complications in an explanans set T". Morgan argues that if T explains E, and if T' is irrelevant to E, then the union of T and T' is no better an explanans for E, and is intuitively worse because it is less simple. Before arriving at the final formulation of this condition, I must now introduce some other technical notions defined by Morgan.

Morgan defines two new notions, that of "shortest conjunctive normal form" and that of "set of conjunctive components". I give first the definition of the former.

> Let S be a sentence of L, and let S_t be the set of TF-components of S. A conjunctive normal form of S with respect to an arbitrary set of sentences S_q is a sentence $P = P_1\&...\&P_n$ such that each conjunct P_i is a disjunction of sentences from S_q and negations of sentences from S_q such that P is logically equivalent to S. The set of shortest conjunctive normal forms of S with respect to S_q is a subset, S_u of the set of all conjunctive normal forms of S with respect to S_q, say $S_u=\{S_{r1},..., S_{rk}\}$ such that the S_{ri} all have the same number, say g, of conjuncts and such that if S_p is any conjunctive normal form of S with respect to S_q and $S_p \notin S_u$, then S_p has more than g conjuncts. If S_u has only one member, then the shortest conjunctive normal form of S with respect S_q is that one member. If S_u has more than one member, then one member may be picked by employing the same method used with the shortest disjunctive normal form. The shortest conjunctive normal form of S with respect to S_t is obtained as above taking S_q to be the set of TF-components of S.[7]

7. ibid., pp. 7-8.

Next is given the definition of "set of conjunctive components":

> The set of conjunctive components, S_{cc}, of a sentence S, is the set whose only members are the conjuncts of the shortest conjunctive normal form of S with respect to S_t. If T is a set of sentences, then T_{cc} is the set of conjunctive components of the conjunction of the members of T.[8]

Morgan's requirement concerning arbitrary additions to T may now be stated in the following manner:

> 4. There is no subset T_{cc}' of T_{cc} such that $T_{cc}' \vdash E$, and for some sentence S, $T_{cc} \vdash S$ but not $T_{cc}' \vdash S$.

With the addition of the last clause, condition 4 amounts to the requirement that there be no <u>proper</u> subset of T_{cc} which will suffice for the derivation of E. To take an example, let

$$T = \{Tb\&Ma,\ (x)(Mx \rightarrow Nx)\},$$

and form the set T_{cc} in accordance with Morgan's definition, which gives us

$$T_{cc} = \{Tb,\ Ma,\ (x)(Mx \rightarrow Nx)\}.$$

Finally, let E be the sentence 'Na'. T will not serve as an explanans for E on Morgan's conditions, for there is a subset of T_{cc}, namely

---

8. ibid., p. 8.

$T_{cc} = \{(x)(Mx \rightarrow Nx),\ Ma\}$,

which suffices for the derivation of E, and there is a sentence S (namely 'Tb') such that $T_{cc} \vdash S$ but not $T_{cc}' \vdash S$. There is, therefore, a proper subset of T_{cc} from which E may be derived.

The next condition Morgan introduces is meant to deal with cases of "self-explanation" and "partial self-explanation". These types of cases should be familiar to the reader from the preceeding chapter, so I turn directly to Morgan's condition:

5. There is no member T_{cci} of T_{cc} such that both not $\vdash T_{cci}$ and $\{E\} \vdash T_{cci}$.

Although Morgan speaks as if this condition by itself will rule out "partial self-explanation", this is not so. Some cases of partial self-explanation are excluded by condition 3. However, 5 is not sufficient to rule out all the other cases of partial self-explanation not excluded by 3. For instance, the following is not ruled out by condition 5 (or by condition 3, for that matter):

$$\frac{\begin{array}{c}(x)(Ux\&Kx \rightarrow Wx)\\ Ua\&Ka\end{array}}{Wa\&(Ka \vee Ua)}\ .$$

This is a partial self-explanation; one conjunct of the explanandum sentence is logically derivable from the singular premiss of the explanans alone. For the moment, simply pointing out the existence of such examples which are sanctioned by Morgan's conditions will suffice. More detailed discussion must be postponed to the next section, since we do not as yet have all of Morgan's conditions before us.

Morgan's last condition is a refinement of one of Ackermann's conditions. As previously discussed, Ackermann has argued that a sentence E which is simply an instance of some universally quantified sentence occurring in the explanans cannot be an explanandum of that explanans. The problem with

$$\begin{array}{l} T: \underline{(x)(Mx)} \\ E: \quad Ma \end{array} ,$$

according to Ackermann, is that when T is expanded in the domain of individuals mentioned in E, the result is equivalent to E. Thus, so we are told, T adds no explanatory import.

Morgan shows that, given Ackermann's reasons for wanting to rule out the example of the preceeding paragraph, the criterion given (i.e., that E not be equivalent to T when T is expanded in the domain of individuals mentioned in E) is inadequate. For consider the following argument, due to Charles Morgan:

$$\begin{array}{l} T_2: \underline{(x)(Pxa)} \\ E_2: \quad Pba \end{array}$$

This case should be as objectionable as the previous one. Yet the expansion of T_2 in the domain of individuals mentioned in E_2 is not equivalent to E_2. Morgan then tentatively suggests that we take the following criterion: E must not be equivalent to the expansion of T in the domain of individuals mentioned in E but not mentioned in T. Yet this condition can not be used to rule out all objectionable cases, for consider the following example:

$$\frac{T_3:\ (y)(x)(Qaxy)}{E_3:\ Qabc}\ .$$

Again, Morgan claims, this example is as objectionable as the previous two. But in this case the expansion of T_3 in the domain of individuals mentioned in E_3 but not mentioned in T_3 is not equivalent to E_3.

Some other examples need to be considered. For instance,

T_5: (x)(Rx)
T_5': (y)(Sy)
E_5: Ra&Sb

T_6: (x)(Rx)
E_6: RavRb .

Neither of these examples is ruled out by either of the two preceeding criteria.

Extrapolating from these examples and others, Morgan suggests the following condition:

6. There is no member E_{cci} of E_{cc} such that
   a) E_{cci} is equivalent to a sentence that is the result of an instantiation in the prenex normal form of T_{ccj}, some member of T_{cc}, for one or more essentially occurring universal quantifiers in the prenex normal

form of T_{ccj}: or b) E_{cci} is equivalent to a truth function of such sentences.

This condition is introduced to exclude not only all the examples displayed above but also examples like the two following:

(1) T_7: (x)(Px Qx)
T_7': (x)(Rx)
T_7'': Pa
E_7: Rb&Qa .

Here, one conjunct of the explanandum is a simple instantiation of a universally generalized sentence occurring in the explanans, and illustrates the need for introducing the set of conjunctive components of E and requiring that no conjunct which is a member of that set be equivalent to an instantiation of a universally quantified sentence occurring in the explanans.

Condition 6 also rules out the following:

(2) T_8: (Ex)(y)(Pxy)
E_8: (Ex)(Pxa) .

This example accounts for the following comment of Morgan: "... [the condition] should be taken in the sense that any sentence that is the result of an instantiation in the prenex normal form of a member of T_{cc} for one or more essentially occurring universal quantifiers should be ruled out, regardless of what other quantifiers occur. Likewise, any truth function of such instantiations should be ruled out."[9]

This elucidates fairly well the rationale given for the introduction of Morgan's sixth condition. And although it seems to me to be not entirely adequate, I will postpone further comment on this condition to the following section.

This concludes our statement of Morgan's conditions. To be perfectly clear as to what has gone before, let me now just summarize all the conditions that have been introduced:

Definition M: Let T be a set of sentences of L, and let E be any sentence of L. Then T is an M-explanans for E if and only if all of the following conditions are satisfied:

M1. Not $\vdash$ E.

M2. Not T $\vdash$ $\sim$E.

M3. T $\vdash_d$ E.

M4. There is no subset T_{cc}' of T_{cc} such that a) $T_{cc}' \vdash$ E, and b) for some sentence S, $T_{cc} \vdash$ S but not $T_{cc}' \vdash$ S.

M5. There is no member T_{cci} of T_{cc} such that a) not $\vdash T_{cci}$, but b) $\{E\} \vdash T_{cci}$.

M6. There is no member E_{cci} of E_{cc} such that: a) E_{cci} is equivalent to a sentence that is the result of instantiation in the prenex normal form of T_{ccj}, where T_{ccj} is some member of T_{cc} which has essential occurrences of at least one universal quantifier, for one

9. ibid., p. 12.

> or more essentially occurring universal quantifiers in the prenex normal form of T_{ccj}: or b) E_{cci} is equivalent to a truth function of such sentences.

The second section of Morgan's paper is devoted to a discussion of the above model. The first point he makes is that the definition is satisfiable; i.e., he shows that, for some set of sentences T and some sentence E, T is an M-explanans for E according to definition M. This is actually a trivial matter. Morgan shows that the set

$T = \{(x)(Px \rightarrow Qx), Pa\}$

constitutes an M-explanans for

$E = Qa.$

The reason he even bothers to show this is that the Ackermann and Stenner model, surprisingly enough, turned out not to be satisfiable at all. That is, given that their first two conditions are satisfied, it can be shown that their fifth condition can not be satisfied. Morgan simply wants to demonstrate that his model is not susceptible to this same drawback.

Morgan also shows that his definition is not trivially satisfiable by every set of sentences T and sentence E. He thus proves that each condition in turn can be violated by some sentences T and E.

In the third and final section of his paper, Morgan discusses the three adequacy conditions of Kaplan enumerated in the preceeding chapter. Two of these conditions required that the relation of explanation be closed under logical derivability. Morgan "proves" that his model does not satisfy these two conditions, but he argues that this is a defect of the conditions and not the model. His arguments to this effect will be evaluated and dispensed with in the next section.

3.2. Criticism of M-explanation

The model just presented is inadequate for several reasons. As a matter of fact, it is not even an adequate definition of "explanation" relative to singular explananda, and is therefore a regression from the Kaplan model. Hence in regard to providing conditions definitive of "explanation" for <u>any</u> consistent E sentence, singular <u>or</u> general, Morgan's model misses the mark by a wide margin.

The first defect is that Morgan's conditions allow for explanations whose explanans do <u>not</u> contain laws or theories. Consider the following: let

$T = \{(Pa \rightarrow Qa), (Qa \rightarrow Ra)\}$

and let E be the sentence

$E = (Pa \rightarrow Ra)$.

Then the following is an M-explanation:

(a) $$\frac{\begin{array}{l} Pa \rightarrow Qa \\ Qa \rightarrow Ra \end{array}}{Pa \rightarrow Ra}\ .$$

To verify that (A) is indeed an M-explanation, we must show that all of M1-M6 are satisfied by this example. First, we must construct the set T_{cc}, which will be the following set:

T_{cc}= {(∿Pa∨Qa), (∿Qa∨Ra)}.

This is the set of conjunctive components of the shortest conjunctive normal form of the conjunction of the members of T. The shortest disjunctive normal form of E is simply

E= (∿Pa∨Ra).

Now we show that this example satisfies all of Morgan's conditions.

M1. Not ⊢ (∿Pa∨Ra). Hence M1 is satisfied.

M2. Not T⊢ ∿(∿Pa∨Ra). This condition is satisfied.

M3. We must show that there is no disjunct E_i of the shortest disjunctive normal form of E such that E_i is not derivable from T but all the other disjuncts are derivable from T. It is not the case that a) not T⊢ ∿Pa and T⊢ Ra, and it is not the case that b) not T⊢ Ra and T⊢ ∿Pa. M3 is thus satisfied.

M4. Here we must show that there is no proper subset T_{cc}' of T_{cc} such that T_{cc}'⊢ E. There are only two possible candidates for membership in T_{cc}', which are (∿Pa∨Qa) and (∿Qa∨Ra). a) It is not the case that {(∿Pa∨Qa)}⊢ ∿Pa∨Ra, and b) it is not the case that {(∿Qa∨Ra)}⊢ ∿Pa∨Ra. M4. is thus satisfied.

M5. It is clear that there is no member T_{cci} of T_{cc} such that T_{cci} is logically provable. We must now show that there is no T_{cci} such that $\{E\} \vdash T_{cci}$. Again, there are only two cases. a) It is not the case that $\{(\sim PavRa)\} \vdash (\sim PavQa)$, and b) it is not the case that $\{(\sim PavRa)\} \vdash (\sim QavRa)$. M5. is therefore satisfied.

M6. This condition is satisfied, since there is no sentence which is the result of instantiating a universally quantified sentence occurring in the explanans such that that sentence is equivalent to $(\sim PavRa)$.

By the same token, the following is also an M-explanation.

(B) $$\frac{\begin{array}{l} Pa \\ Pa \to Qa \end{array}}{(Ex)(Qx)}$$

Morgan does not, I think, want to be committed to explanations without laws. The fact that his model does allow for such cases is simply an oversight on his part.

To remedy this defect, Morgan might consider adding the following condition:

M7*. Some T_{cci} $(T_{cci} \varepsilon T_{cc})$ is a law or theory.

But to just make this adjustment alone would not save the model. Even with the addition of M7*, troubles still arise.

Morgan's model will eliminate the "mermaid" example of

Eberle, Kaplan, and Montague, i.e.,

(C) $$\frac{(x)(Mx \rightarrow Hx) \quad MevHe}{He}$$

and it will also eliminate what I have called the "Morgan-type mermaid example" which, in one sense, trivialized the Kim model, i.e.,

(D) $$\frac{(x)(Mx \rightarrow Hx) \quad MevHe}{HevPa} \; .$$

This is no surprise, for he introduced conditions specifically for this purpose. But Morgan's conditions will <u>not</u> eliminate a similar example.

That is to say, on Morgan's conditions the sentence "the Eiffel Tower is a good conductor of heat or the Eiffel Tower is a good conductor of electricity" is explainable by the theory "all mermaids are good conductors of heat" in conjunction with the singular sentence "if the Eiffel Tower is not a mermaid, then the Eiffel Tower is a good conductor of electricity". Formally,

(E) $$\frac{(x)(Mx \rightarrow Hx) \quad MevEe}{HevEe} \; .$$

Indeed, any argument of the following form will satisfy Morgan's conditions:

(F) $$\frac{(x)(Mx \rightarrow Hx) \quad Mxv\phi}{Hxv\phi} \; .$$

Normally, the explanans will be required to be true, so we should also stipulate that 'ϕ' be a <u>true</u> sentence; indeed, any true sentence except 'Hx'.

Example (E) is sufficient, in my opinion, to show the inadequacy of Morgan's proposal.

There are additional problems with Morgan's model. Consider the following argument, similar to (B) above, but which satisfies M7* in addition to M1-M6:

$$\text{(G)}\quad \frac{\begin{array}{c}(x)(Px)\\ Pa \rightarrow Ga\end{array}}{(Ex)(Gx)}\ .$$

Is this argument of good explanatory form? It satisfies all of Morgan's conditions, M5 in particular; for $\sim Pa \vee Ga$ is <u>not</u> logically derivable from (Ex)(Gx).

To get clear as to what is involved here, let us recall Hempel's discussion of the following argument form:

$$\text{(H)}\quad \begin{array}{l}T:\\ C:\\ E:\end{array} \frac{\begin{array}{c}(x)(Px)\\ Pa \rightarrow Ga\end{array}}{Ga}\ .$$

(Actually, this is not Hempel's example, but one essentially equivalent to one he discusses.) Hempel has argued that this form should not be counted as potentially explanatory, for the following reason. If the theory T on which the explanation rests, is actually true, then the sentence C, which can also be put into the form $\sim Pa \vee Ga$, can be shown to be true, or can be verified, only by verifying 'Ga', i.e., E. Furthermore, Hempel continued, this peculiarity deprives the proposed explanation (H) of any predictive import, for E cound not be predicted on the basis of T and C because the truth of C can not be established in a way which does not include simultaneous verification of E. It is for this reason, you will recall, that Hempel and Oppenheim added the following condition to their analysis of potential explanation:

> There is a class K of basic sentences such that $K \vdash C$ but neither $K \vdash \sim T$ nor $K \vdash E$.

Morgan's syntactical model does not contain a similar "verification" condition. In cases such as (H) such a condition is not required, for (H) is eliminated by M5. M5 will not, however, eliminate (G). Yet it seems that (G) ought to be ruled out on the same grounds as (H). That is, if the theory of (G) is true, then the sentence of initial conditions can only be verified in such a way as to simultaneously verify E.

The notion of verification defined by Hempel and Oppenheim was restricted to molecular sentences without quantifiers, such as'((Pa&Qa)vRa)', which could be verified either by establishing the truth of 'Ra' or the truth of '(Pa&Qa)'. In other words, verification of a molecular sentence S is defined as the establishment of of the truth of some class of basic sentences which has S as a consequence.

It seems natural to extend this notion of verification to existentially generalized sentences (though not, of course, to universally generalized ones). Let us say that an existentially generalized sentence, e.g., '(Ex)(GxvHx)', can be verified by establishing the truth of some class of basic sentences which has it as a consequence, in this case, 'Ha', for example.

In the case of (G), verification of 'Pa→Ga' brings in its train verification of '(Ex)(Gx)', and should, therefore, be eliminated on exactly the same grounds as (H). This is just one more problem with which Morgan ought to contend.

I have brought these examples to Morgan's attention, and he agrees that the model is inadequate, as it stands. I have also pointed out to him that part of the difficulty he faces, viz., the fact that he is committed to the undesirable explanatory forms (A), (E), and (G), stems from the fact that he ignores some of the insights contained in Kaplan's proposal.

In my discussion of Kaplan's model (cf. Chapter II), I noted that an extension of his conditions to cover generalized E sentences seemed to be the natural thing to do, as his model is the only one which is an adequate account of explanation for singular E sentences (this task I undertake in Chapter IV). One of Kaplan's conditions, the one which I take to incorporate the chief insight of his analysis, was the requirement that E not be a logical consequence of the disjunctive normal form of C. This requirement ruled out the "mermaid" example of Eberle, Kaplan, and Montague, i.e.,

$$\frac{\begin{array}{c}(x)(Mx \rightarrow Hx)\\ Me \vee He\end{array}}{He},$$

and also the "Morgan-type mermaid example", i.e.,

$$\frac{\begin{array}{c}(x)(Mx \rightarrow Hx)\\ Me \vee He\end{array}}{He \vee Pa}.$$

But the interesting point to note is that a similar requirement will rule out (A), (E), and (G).

In discussing this with Morgan, he suggested the following "Kaplanesque" condition:

(K*) Form the shortest disjunctive normal form of the conjunction of the members of T with respect to the TF-components of that conjunction; if the shortest disjunctive normal form has more than one disjunct then there must be no disjunct D_i, such that $D_i \vdash E$.

It is necessary to state this requirement as a conditional, for otherwise we would rule out

$$\frac{\begin{array}{c}(x)(Px \rightarrow Qx)\\ Pa\end{array}}{Qa}.$$

This condition will rule out (E), i.e.,

$$\frac{\begin{array}{c}(x)(Mx \rightarrow Hx)\\ Me \vee He\end{array}}{He \vee Ee},$$

since, given the shortest disjunctive normal form of the conjunction of the members of T, as follows,

$((x)(Mx \rightarrow Hx) \& Me) v ((x)(Mx \rightarrow Hx) \& Ee)$

it is readily seen that both disjuncts have 'HevEe' as a logical consequence.

Example (A) is also ruled out, i.e.,

$$\frac{\begin{array}{l} Pa \rightarrow Qa \\ Qa \rightarrow Ra \end{array}}{Pa \rightarrow Ra} \quad ,$$

which leads us to conjecture that, in light of condition (K*), it is not necessary to invoke condition M7*.

Before settling on this issue, let me first make good on another promissory note. In the previous section I remarked that Morgan's treatment of partial self-explanation did not appear adequate. I now show that, even with the addition of the "Kaplanesque" condition, serious problems ensue.

Consider the following example once again:

$$\text{(I)} \quad \frac{\begin{array}{l} (x)(Ux \& Kx \rightarrow Wx) \\ Ua \& Ka \end{array}}{Wa \& (KavUa)} \quad .$$

This is a partial self-explanation, since one conjunct of the explanandum sentence follows from the singular premiss of the explanans alone. As noted in Chapter II, this is certainly an undesirable state of affairs, and such argument forms should not be taken as explanatory. I now show that (I) satisfies all of M1-M6, (K*), and even M7*.

M1. and M2. These conditions are obviously satisfied.

M3. This condition is satisfied, since there is no disjunct E_i of the shortest disjunctive normal form of E (i.e., (Wa&Ka)v(Wa&Ua)) such that not $T \vdash E_i$ but $T \vdash E_1 v \ldots vE_{i-1} vE_{i+1} \ldots vE_n$. This is so because, for every E_i, $T \vdash E_i$.

M4. This condition is satisfied, as is easily seen. First, form the set T_{cc}.

$T_{cc} = \{(x)(Ux\&Kx \rightarrow Wx), Ua, Ka\}$.

Now we must show that there is no subset T_{cc}' of T_{cc} such that $T_{cc}' \vdash$ (Wa&Ka)v(Wa&Ua) and such that for some sentence S, $T_{cc} \vdash S$ but not $T_{cc}' \vdash S$. Clearly, there is no proper subset of T_{cc} which will suffice for the derivation of E, since every element of the set is essential to the derivation of E. Therefore, this condition is satisfied.

M5. We must show that there is no member T_{cci} of T_{cc} such that not $\vdash T_{cci}$ but $\{E\} \vdash T_{cci}$. First, it is obvious that no member of T_{cc} is logically provable. Now we must show that there is no T_{cci} such that $\{E\} \vdash T_{cci}$. There are three cases to be considered.

Case 1: Obviously not {(Wa&Ka)v(Wa&Ua)}$\vdash$ (x)(Ux&Kx→Wx).

Case 2: Not {(Wa&Ka)v(Wa&Ua)}$\vdash$ Ua, for there is an interpretation which makes Ua false, but (Wa&Ka)v(Wa&Ua) true.

Case 3: Not {(Wa&Ka)v(Wa&Ua)}$\vdash$ Ka, for the same reason as in case 2.

Therefore, there is no member T_{cci} of T_{cc} such that not $\vdash T_{cci}$ and $\{E\} \vdash T_{cci}$. This condition is therefore satisfied.

M6. We must show that there is no member E_{cci} of E_{cc} such that E_{cci} is equivalent to an instantiation of the prenex normal form of some T_{ccj}, where T_{ccj} is a member of T_{cc} such that T_{ccj} has essential occurrences of at least one universal quantifier. First, I specify the set E_{cc}, which is the following:

E_{cc}= {Wa, (KavUa)}.

There is only one member of T_{cc} having essential occurrences of a universal quantifier, namely the formula (x)(Ux&Kx→Wx). We need consider only one instantiation of this formula, i.e., (Ua&Ka→Wa), as any other choice of constants would clearly not be equivalent to any E_{cci}. There are two cases to be considered, one for each member of E_{cc}.

Case 1: Wa is not equivalent to (Ua&Ka→Wa), since the implication from right to left is invalid.

Case 2: KavUa is not equivalent to (Ua&Ka→Wa), for neither implication is valid.

Condition M6 is, therefore, satisfied.

(K*) The shortest disjunctive normal form of the conjunction of the members of T is (x)(Ux&Kx→Wx)&Ua&Ka. Since this normal form is the shortest and does not have more than one disjunct, this condition is vacuously satisfied.

We have found, therefore, that partial self-explanation is not excluded by any of Morgan's conditions nor by the Kaplanesque condition. By itself, Morgan's model is inadequate, and the addition of condition (K*) will not, unfortunately, eliminate all of the defects.

But now let us return to the question raised earlier, as to whether the addition of (K*) to Morgan's model makes M7* unnecessary. That is, will M1-M6 plus (K*) rule out explanations whose explanans do not contain laws? The answer to this question is 'no', as the following example demonstrates:

(J) T: Fa&Gb
E: (Ex)(Fx)&(Ey)(Gy) .

To show that this is indeed an M-explanation, we must show that this example satisfies all of M1-M6 as well as (K*).

M1. and M2. are obviously satisfied.

M3. There is no disjunct E_i of the shortest disjunctive normal form of E such that not $T \vdash E_i$ but $T \vdash E_1 v \ldots vE_{i-1} vE_{i+1} v \ldots vE_n$. This condition, therefore, is satisfied.

M4. Clearly, no proper subset of T_{cc}={Fa, Gb} has E as a consequence. M4. is thus satisfied.

M5. There is no member T_{cci} of T_{cc} such that not $\vdash T_{cci}$ but $\{E\} \vdash T_{cci}$. For example, Fa is _not_ a logical consequence of (Ex)(Fx). M5. is thus also satisfied.

M6. This condition is vacuously satisfied.

(K*) This condition is also vacuously satisfied, since the shortest disjunctive normal form of the conjunction of the members of T does not have more than one disjunct.

Morgan's model will not, therefore, even when supplemented by condition (K*), rule out explanations whose explanans do not contain laws.

3.3. Further comments on M-explanation

Paul Churchland, a participant in the symposium based on Morgan's paper at the 1972 Canadian Philosophical Association meetings, has shown that Morgan can only adopt condition (K*) at the expense of ruling out certain argument forms which are paradigm instances of acceptable explanations. For example, (K*) will rule out the following:

$$\begin{array}{ll} T: & (x)(Px \vee Qx \rightarrow Sx) \\ C: & Pa \vee Qa \\ \hline E: & Sa \end{array} .$$

To see this, we first form the disjunctive normal form of the conjunction of the explanans set, arriving at the following sentence:

$$(x)(Px \vee Qx \rightarrow Sx) \& Pa \ \vee \ (x)(Px \vee Qx \rightarrow Sx) \& Qa.$$

This sentence has more than one disjunct, so (K*) requires that no disjunct of this sentence have E as a consequence. Since E is derivable from both disjuncts, (K*) rules out this example.

There is, therefore, no obvious satisfactory "patch" of Morgan's model which will both eliminate the "mermaid" counter-example which I have produced specifically against Morgan's model and also not have undesirable consequences.

Now, there is one more promissory note to be cashed. This has to do with Morgan's putative proof of the inadequacy of Kaplan's conditions R1 and R2. The final section of the present chapter will be devoted to Morgan's critique of these conditions. For ease of reference, these conditions will be restated here:

R1. If a singular sentence is explainable by a given theory, then it is explainable by any theory from which the given theory is logically derivable.

R2. Any singular sentence which is logically derivable from singular sentences explainable by a theory is itself explainable by that theory.

Morgan purports to show that M-explanation does not satisfy either R1 or R2, but he adds, "... I shall argue that this fact does not count against the adequacy of the model".[10] It is my contention that this second claim is not substantiated; that is, it seems to me that Morgan's "argument" against, for example, R1, amounts simply to the claim that M-explanation does not satisfy this condition of adequacy. His other remarks against R1 are unconvincing. Again, a few preliminaries are in order.

As Morgan is aware, the terms "fundamental theory", "derivative theory", "law" and "theory" are to be understood

10. ibid., p. 24.

in the present context in the sense of Hempel and Oppenheim. These terms were defined as follows: A _fundamental law_ is a true sentence consisting of one or more universal quantifiers followed by an expression without quantifiers or individual constants. A sentence S is called a _derivative law_ if (1) S consists of one or more universal quantifiers followed by an expression without quantifiers, (2) S is not logically equivalent to any singular sentence, (3) at least one individual constant occurs in S, and (4) there is a class K of fundamental laws such that S is logically derivable from K. A _law_ is a sentence which is either a fundamental law or a derivative law.

A _fundamental theory_ is a true sentence consisting of one or more quantifiers followed by an expression without quantifiers or individual constants. A sentence S is called a _derivative theory_ if (1) S consists of one or more quantifiers followed by an expression without quantifiers, (2) S is not logically equivalent to any singular sentence, (3) at least one constant occurs in S, and (4) there is a class K of fundamental theories such that S is logically derivable from K. A _theory_ is a sentence which is either a fundamental theory or a derivative theory.

These definitions have been stated previously, but since they bear a great deal upon what I have to say about Morgan's arguments it is essential that we be quite clear as to their exact statement.

Further, as Morgan is aware, to say that "E is explainable by T" on the Hempel and Oppenheim account is to say that there is a singular sentence C such that

1) T is a theory
2) T is not equivalent to any singular sentence
3) C is singular and true
4) E is derivable from {T,C}
5) There is a class K of basic sentences such that C is logically derivable from K, and neither E nor the negation of T is logically derivable from K.

This definition should also be familiar.

Now it is clear that Morgan's model does not satisfy R1, on account of the following considerations. Suppose T'' and T' are theories, and that T is an M-explanans for E. Let the explanans set T= {T',C}, and let T'' be the theory T_1&T'. Now, T'' has T' as a logical consequence, but T*= {T'',C} will not be an M-explanans for E, since condition M4 is violated. That is, there is a subset of the set of conjunctive components of T* which will suffice for the derivation of E.

But so far this does not constitute an argument against R1; it simply shows that M-explanation does not satisfy it. However, Morgan continues:

> "The unsatisfactory character of R1 is more clearly exhibited when it is altered slightly to a form which is more directly applicable to M-explanation:
>
>> R1'. Suppose a set of sentences T constitutes an M-explanation for a sentence E. Let CT be the conjunction of all of the sentences in T. Then any set of sentences T' such that $T' \vdash CT$ also constitutes an M-explanation for E.
>
> The reason this condition is objectionable is that the characterization of T' is much too general. T' could be contradictory; or T' could be constructed by taking the conjunction of E and T (and any other sentences)- this latter construction leads to the problem of self-explanation discussed above."[11]

Here we at least have an argument, but I am afraid it comes to very little; in fact, Morgan errs on each of the three points made in this passage.

First, to speak in Morgan's terms, R1 does _not_ assert that _any_ set of sentences T' such that $T' \vdash CT$ constitutes an explanans for a given E (assuming, of course, that $CT \vdash E$); it only asserts that any _theory_ which entails the conjunctive component of CT that is itself a theory will serve as the theoretical component in an explanans for E. Thus, Morgan's reformulation R1' is not an adequate rendering of R1.

Furthermore, R1' does not capture the requirement that

---

11. ibid., p. 25.

T' be a theory. As a result, Morgan's claim that "T' could be contradictory" is false. For, as the definitions given above make clear, T' must be true if it is to be a theory- and hence it can not be contradictory.

Nor is Morgan's claim justified that T' could be the conjunction of T and E; for there is no guarantee that such a conjunction will indeed be a theory. This is something Morgan must prove. By the definitions above, there are certain syntactical and semantical conditions to be met in order for a sentence to qualify as a theory. In the case where we have the conjunction of T and E, in prenex normal form, this has the syntactical structure of a derivative theory (in so far as it is an expression which contains individual constants), but this does not suffice to show that such a conjunction is a derivative theory. For one of the other conditions in the definition of "derivative theory" is that there be a class K of fundamental theories such that the derivative theory is logically derivable from K.

Suppose, for the sake of argument, that such a K exists, and that (T&E) is indeed a derivative theory. Then (T&E) constitutes a purely theoretical explanation of E. And this is not tantamount to self-explanation. That is to say, in

$$\begin{array}{ll} T': & (T\&E) \\ C\ : & \underline{P \vee {\sim}P} \qquad K = \{\phi\} \\ E': & E \end{array}$$

E does not follow from C alone, but it does follow from T' alone. However, this is not objectionable to Hempel and Oppenheim (nor to me); purely theoretical explanation is not considered by us to be undesirable. Thus, Morgan's objection must ultimately rest on the fact that R1 is incompatible with M6. But surely this is no argument showing that the fault lies with R1; it might well lie with M6.

Much of what I have said with regard to Morgan's criticism of R1 holds for his criticism of R2. However, Morgan does make one point vis-a-vis R2 which will necessitate my revising it slightly.

As it stands, R2 will countenance tautologies as explainable. For suppose 'Ea' is explainable by a theory T. 'Pav~Pa' is a singular sentence which is logically derivable from 'Ea', and thus R2 sanctions the explainability of 'Pav~Pa' by T. This is certainly undesirable, as Morgan notes.

However, instead of scraping R2 altogether, I suggest replacing it with the following:

> R2*. Any sentence which is not logically provable and which is logically derivable from singular sentences explainable by a theory is itself explainable by that theory.

All of Morgan's other criticisms of R2 amount to showing that explanations which satisfy it may not be M-explanations, i.e., may violate one of the conditions M2-M6. Since Morgan's model has been shown to be inadequate (on grounds quite independent of R2), I feel that all of his comments in this regard come to nought.

Chapter IV

Explanations of Laws

In the two preceeding Chapters we have shown that much of the effort recently expended on the deductive-nomological model of explanation has been an exercise in futility. With the exception of Kaplan's proposal all the other models seem to form a series progressing from bad to worse. Kim's analysis was shown to be faulty, Ackermann's fared no better, Omer's model was unsatisfiable, as was Ackermann and Stenner's 'corrected' version. Finally, Morgan's unpublished proposal still contained elementary errors which Kaplan had already dealt with in a satisfactory manner.

The responsibility for the demise of any clear picture of explanation rests, I believe, on Ackermann's shoulders. He felt that the original Hempel and Oppenheim model would be too restrictive _even if_ the difficulties that Eberle, Kaplan, and Montague had shown to exist could be remedied. In his 1965 paper, he had remarked:

> The most serious of these defects is to be found in the fact that the extant models seem to be formally restrictive in ways that do not allow any obvious generalization of their

> conditions which will encompass the full range of all those scientific explanations which must be considered plausible candidates for translation into deductive models.[1]

Ackermann is here criticizing the Hempel and Oppenheim model because "the adoption of properties and conditions closely tied to singular sentences in a language L is unduly restrictive; for no full analysis of deductive scientific explanation appears to be constructible by generalization on them or addition to them. The Hempel-Oppenheim-Kim model can at best be the prelude to a rather piecemean analysis of scientific explanation, making explanations belonging to different types justified by reference to differing particular deductive models."[2]

In short, the Hempel and Oppenheim definition of explanation, being defined only for singular sentences, does not give us an account of explanation when what one wants explained is a law. Ackermann then presented a definition in which the sentence E was not restricted to singular sentences, and other authors followed him in this endeavor. Unfortunately, all these attempts have failed.

Now the reason I say that Ackermann is responsible for the main confusions surrounding recent proposals for a definition of explanation is this: treating singular and

---

1. Ackermann, R., "Deductive Scientific Explanation", Philosophy of Science, 32 (1965), p. 155.
2. ibid., p. 160.

general explananda on a par, as Ackermann does, tends to telescope several issues together which I believe require separate treatment. Hempel and Oppenheim pointed vaguely to some of the problems involved in defining explanation relative to sentences expressing laws, but were not clear as to the precise solution of these difficulties. Their discussion does, however, make one thing clear: explanations of laws present certain peculiar difficulties in explication, in light of which they feel it best to treat such explanations separately. Clearly, "explanation" defined relative to singular E sentences has proven to be a more difficult task than was initially envisaged; and when a proposed explication attempts to treat of any E sentence, singular or general, the difficulties are multiplied enormously.

Therefore, contrary to Ackermann, I believe that it is best to follow the original Hempel and Oppenheim approach and propose that we define explanation in two stages; first, relative to singular E sentences, and second, relative to universally generalized E sentences. I do not want to claim that Ackermann's approach will never lead to success; indeed, once we define the notion relative to the two different types of explananda, we may perhaps be able to produce one natural set of conditions to cover both types. On the other hand, I also feel that the isolation of different sorts of problems contained in each of the two different types of explanation that my approach makes possible will justify somewhat my treating explanation in this two-fold way. It will be shown that, in the case of universally generalized explananda, there are a few unresolved problems which Ackermann's approach conceals.

In this chapter a definition of explanation relative to universally generalized E sentences will be offered. The definition will provide a set of necessary conditions, but whether the conditions are also sufficient will require discussion. The conditions would appear sufficient if we wished to remain at the rather naive level of Ackermann's approach; but the additional difficulties involved in explanations of laws (which will be revealed in this chapter and the next) legislate against our counting them as sufficient.

Professor Kaplan has already defined explanation for singular E sentences, and we shall take his proposal as satisfactory for the first stage of our two stage enquiry. In defining "explanation" for laws, we will follow as far and as closely as possible the insights of Kaplan's analysis. We shall limit ourselves in this task by treating existentially generalized sentences differently than Ackermann; in particular, an existentially generalized sentence T will be explainable just in case there is a singular sentence

S such that S is explainable in the sense of Kaplan, and such that T is a logical consequence of S. Mixed quantifications will not be dealt with in the present work.

4.1. <u>Kaplan revisited</u>

In this section I want to follow Kaplan's lead and see how far it will guide us in our search for what might be called a generalized model of explanation. Some of Kaplan's conditions will obviously carry over to the present model, e.g., the requirement that the explanandum be logically derivable from the explanans, but other important conditions will not carry over. Thus, new conditions will be required.

Let us first remind ourselves of Kaplan's proposal. I have argued that the following be taken as Kaplan definition of S-explanation:

(K1) The ordered couple (T,C) of sentences is a direct S-explanans for the singular sentence E if and only if

1) T is a theory
2) C is singular and true
3) E is not derivable from any disjunct of the disjunctive normal form of C
4) $\{T,C\} \vdash E$
5) There is a class K of basic sentences such that $K \vdash C$ but neither $K \vdash E$ nor $K \vdash \sim T$.

6) E is a disjunction of basic sentences

(K2) (T,C) is an S-explanans for the singular sentence E if and only if

1) (T,C) is a direct S-explanans for each of a set of singular sentences $\{E_1,\ldots,E_n\}$, and

2) $\{E_1,\ldots,E_n\} \vdash E$.

It should be obvious that some of these conditions will no longer be relevant for generalized explanation. For instance, in explanations of laws, the explanans need not contain any sentences of initial conditions, but only other laws. The notion of "step-wise" explanation will, however, be retained.

Some conditions will no longer be applicable since in moving to generalized explanation we forfeit the ability to distinguish between different sentences of the explanans. In the case of singular explananda, it is easy to distinguish between the sentences T and C in the explanans; one is a theory, or a law, and the other is a singular sentence. One of Kaplan's insights was that some sort of restriction must be placed on C, the sentences of initial conditions. When all the premises are theoretical, however, it becomes rather arbitrary to pick one of these sentences and place a restriction on it. For example, consider the following putative explanation of a law by other laws:

(A) T: $(x)(Px \rightarrow Qx)$

T': $\underline{(x)(Px \rightarrow Qx) \rightarrow (y)(Gy \rightarrow Sy)}$

E: $(y)(Gy \rightarrow Sy)$

This form of explanation is illicit, for precisely the same reason as the following form, discussed by Hempel and Oppenheim and in Chapter II, is illicit:

(B) T: $(x)(Px)$

C: $\underline{Pa \rightarrow Rab}$

E: Rab

Briefly, the point is that if explanatory forms such as (A) were permitted, then, as long as the language contained at least one law T, every other law would be explainable by T; for we are always assured that there is a sentence T' such that T and T' have E as a consequence.

But suppose, however, that we did define some notion of generalized explanation in terms of the ordered pair (T, T') of laws, and placed a condition on T' similar to the disjunctive normal form requirement that Kaplan placed on C. That is, suppose we required that no disjunct of the disjunctive normal form of T' have E as a logical consequence. We take the disjunctive normal form in terms of the sentential components of T', i.e.,

$$T' = \sim(x)(Px \rightarrow Qx) \ \vee \ (y)(Gy \rightarrow Sy)$$

One disjunct does have E as a logical consequence, so this requirement would rule out example (A).

But suppose we interchange T and T', to get the following:

(C) T: $(x)(Px \rightarrow Qx) \rightarrow (y)(Gy \rightarrow Sy)$

T': $\underline{(x)(Px \rightarrow Qx)}$

E: $(y)(Gy \rightarrow Sy)$

Now there is no disjunct of T' which has E as a logical consequence. We therefore end up at the rather absurd conclusion that the order in which we write the sentences of the explanans has a bearing on its capacity for explanation.

Another reason for avoiding this approach is that we may not always be able to put the sentence T' of example (A) into disjunctive normal form. In that example, we might have taken the following as T':

$(x)((Px \rightarrow Qx) \rightarrow (Gx \rightarrow Sx))$

This sentence can not be put into the desired disjunctive normal form, since the quantifier binds the whole subsequent expression.

4.2. New Foundations

We shall leave Kaplan for the time being, returning to his analysis in the next section.

Now since it seems unlikely that we can single out in a non-arbitrary way a certain member of the explanans set, we would not be ill-advised to begin anew, and try this time to follow up some suggestions made by Nagel in *The Structure of Science*. Nagel limits himself in the first few chapters of this book, subtitled "Problems in the Logic of Scientific Explanation", to some very general remarks regarding the nature of explanation. He discusses both the explanation of particular phenomenon and of general laws. In regard to the latter he gives an example of what intuitively is an explanation of a law, and remarks that three points are evident in such an explanation:

1) All the premises are universal statements.
2) There is more than one premise, each of which is essential in the derivation of the explanandum.
3) The premises, singly or conjointly, do not follow logically from the explanandum.[3]

Condition 2) contains a slight ambiguity, concerning which Nagel remarks: "It is always possible to obtain just one premise, by forming the conjunction of several premises. What is intended... is that if there were only a single conjunctive premise, it would be equivalent to a class of logically independent premises in which the class would contain more than one member."[4] If we wished to be more formal at this stage, we could use the definition of "ultimate sentential conjunct" as given by Ackermann. We could then understand condition 3) as requiring that no such conjunct (i.e., premise) be a consequence of E. We

---

3. Nagel, E., The Structure of Science, 1961, p. 34.
4. ibid.

may just assume at this point that some such definition is invoked.

A condition such as 3) would rule out the following example:

(D) T: $\underline{K \cdot B}$

E: K

It was in view of examples like (D) which satisfy the deducibility requirement but which ought not to be accepted as adequate explanations, that Hempel and Oppenheim limited their definition to singular E sentences. Concerning this limitation, the authors remarked:

> The precise rational reconstruction of explanation as applied to general regularities presents peculiar problems for which we can offer no solution at present. The core of the difficulty can be indicated briefly by reference to an example: Kepler's laws, K, may be conjoined with Boyle's law, B, to a stronger law K·B; but derivation of K from the latter would not be considered as an explanation of the regularities stated in Kepler's laws.... The derivation of Newton's laws of motion and of gravitation, on the other hand, would be recognized as a genuine explanation in terms of more comprehensive regularities, or so-call higher-level laws. The problem therefore arises of setting up clear-cut criteria for the distinction of levels of explanation or for a comparison of generalized sentences as to their comprehensiveness. The establishment of adequate criteria for this purpose is as yet an open problem.[5]

I have previously discussed this remark of Hempel and Oppenheim, but since it points to some crucial considerations I have reproduced it once again to emphasize these points. The

---

5. Hempel, C., Aspects of Scientific Explanation, p. 273.

first point to note is that (i) the Hempel and Oppenheim conditions are not intended to resolve the difficulty embodied in example (D), above. That is why they limited their definition to singular explananda. In addition, (ii) they point out that some criteria are needed for distinguishing between the relative comprehensiveness or generality of laws.

Now condition 3) suggested by Nagel resolves the problem mentioned in (i); for the explanandum K of example (D) has one of the premises as a logical consequence. But more importantly, we must determine whether Nagel's conditions (or any to be subsequently introduced) simultaneously provide an answer to problem (ii). That is, does the requirement that E not entail any premise also solve the problem of providing a criterion for distinguishing the relative generality of laws? This question will be discussed more fully in the sequel, in particular in the following Chapter.

Thus far we have remarked that Nagel's suggestion deals in a satisfactory manner with example (D). We might therefore, tentatively adopt the following as our definition of direct explanation for laws:

(K1') The set of sentences $T=\{(T_1, \ldots, T_n)\}$ ($n \geq 2$) is a direct S-explanans for the purely universal sentence E if and only if

1) For all T_i in T, T_i is a fundamental law

2) $T \vdash E$

3) not $E \vdash T_i$.

This is the definition to which Nagel's suggestion leads us. It will be noted that condition 3) of the definition gives the model a sort of "Kim-like" flavour, so our first step in evaluating this model ought to be a check to see if it is susceptible to the same drawbacks as was Kim's model.

We begin this enquiry with an examination of the following argument, which is directly analogous to Theorem 1 of Eberle, Kaplan, and Montague:

(E) T': $(x)(Mx \rightarrow Hx) \vee (x)(Jx \rightarrow (Px \rightarrow Qx))$

T'': $\underline{(x)(((Mx \rightarrow Hx) \vee \sim Jx) \rightarrow (Px \rightarrow Qx))}$

E: $(x)(Px \rightarrow Qx)$

It is here assumed that T' and T'' are the only members of T.

It is desirable that example (E) be ruled out; otherwise, we would be able to explain a given law by any other law. That is, given the law $(x)(Mx \rightarrow Hx)$ with no predicates in common with the law to be explained, namely $(x)(Px \rightarrow Qx)$, we will always be able to construct another law (T'' of the example) such that $(x)(Mx \rightarrow Hx)$ and T'' constitute an explanans for $(x)(Px \rightarrow Qx)$. Definition (K1') precludes this example, in virtue of condition 3). For E has T'' as a logical consequence.

As was the case with Kim's solution to the problems embodied in the trivialization theorems of Eberle, Kaplan, and Montague for singular explananda, Nagel's conditions avoid the generalized analogue to those theorems. Yet it is clear that his conditions are susceptible to the same shortcoming as Kim's. In fact, we may introduce a counterexample to these conditions much like Morgan produced to demonstrate the inadequacy of Kim's proposal. For consider the following:

(F) T': $(x)(Mx \rightarrow Hx) \vee (x)(Jx \rightarrow (Px \rightarrow Qx))$

T'': $\underline{(x)(((Mx \rightarrow Hx) \vee \sim Jx) \rightarrow (Px \rightarrow Qx))}$

E: $(x)(Px \rightarrow Qx) \vee (x)(Sx)$

By applying the inference rule of addition to the explanandum E of example (E), the derivation of E from T' and T'' still holds, but the derivation of T' or T'' from E is precluded. As was the case with Morgan's Theorem 1, this means that almost any disjunctive sentence will be explainable by almost any law.

4.3. Kaplan Again

Nagel's suggestion does not allow us to capture an analogue of Kaplan's insight. Instead, his suggestion goes the way of Kim's. We attempt in this section to find a requirement that will capture Kaplan's insight.

As mentioned previously, there is no obvious non-arbitrary way of singling out one premise of the explanans in order to place some restriction upon it. Even if we could so select one premise, an analogue of Kaplan's disjunctive normal form requirement will not do the required job. For consider example (F) again, this time with the premise T'' written in what Quine calls the <u>canonical normal form</u> for monadic formulae:

(F') T': $(x)(Mx \rightarrow Hx) \vee (x)(Jx \rightarrow (Px \rightarrow Qx))$

T'': $\underline{\sim(Ex)(Px \& \sim Qx \& \sim Mx) \& \sim(Ex)(Px \& \sim Qx \& Hx) \& \sim(Ex)(\sim Jx \& Px \& \sim Qx)}$

E: $(x)(Px \rightarrow Qx) \vee (x)(Sx)$

There is no single conjunct (or any combination of conjuncts) of T'' which has E as a logical consequence. Therefore, we can not follow Kaplan's lead faithfully, since a comparable disjunctive normal form requirement to the one advanced by Kaplan will not suffice.

However, we can always single-out the explanandum sentence E in a non-arbitrary way, so perhaps a comparable requirement placed on this sentence will be satisfactory. Just glancing at example (F'), we see immediately that this example is excluded if we require that no disjunct of the disjunctive normal form of E have any premise T_i as a logical consequence.

This idea is, I believe, basically sound but it must be modified somewhat. As an illustration of the need for some modification, consider the following:

(G) T': $(x)(Mx \rightarrow Hx) \vee (x)(Jx \rightarrow (Px \rightarrow Qx))$

T'': $\underline{(x)\ (((Mx \rightarrow Hx) \vee \sim Jx) \rightarrow (Px \rightarrow Qx))}$

E: $(x)((Px \rightarrow Qx) \vee Sx)$

The explanandum of this example is a simple logical consequence of the explanandum of (F'), and ought not to be held explainable on the basis of the given explanans. Now in (G) the scope of the quantifier occurring in the explanandum ranges over the entire subsequent expression, and is not amenable to the truth-functional analysis required by the disjunctive normal form requirement. The point of this example is that T' and T'' still have E as a consequence, but since the quantifier in the explanandum binds the whole following expression we are prohibited from applying such a requirement. For the disjunctive normal form requirement is defined in terms of sentential logic and sentential components and it no longer applies to parts of a generalized expression, but only to the whole expression. In short, there is no disjunct of the disjunctive normal form of E in example (G) which has some T_i as a consequence.

To remedy this defect, I want to introduce what may be called "the sub-law requirement". The requirement is best exemplified by means of an illustration. Consider the following sentence:

S. $(x)((Px \rightarrow Qx) \vee Sx)$

The expression is explainable according to the conditions so far discussed by the explanans of example (G), since $(x)(Px \rightarrow Qx)$ is derivable from those premises, thereby making the whole expression explainable. The added disjunct, however, precludes the derivation of any premise from S, hence we can not eliminate sentences like this from satisfying our definition in a trivial way. We need to be able to single out a part of this expression, such as (x)(Px→Qx), in order to make the requisite restriction.

I suggest doing this as follows. Consider all the law-like expressions formed from the predicates occurring in S, such that S is derivable from those laws. For example, S is derivable from each of the following "sub-laws":

1. (x)Qx
2. (x)∼Px
3. (x)Sx
4. (x)(Px→Qx)
5. (x)((Px→Qx) v Sx)

Only the last two expressions are derivable from the explanans of (G). The fourth is the one that causes the trouble, since it is invirtue of that expression that the whole sentence S is explainable.

Therefore, what seems to be required is the following: There must be no sublaw SL of E such that SL ⊢ E, T ⊢ SL, and SL ⊢ T_i. The rationale for such a condition will be discussed more fully in connection with the "verification" condition of Hempel and Oppenheim. But first let us summarize the conditions that we have so far found to be essential components in our definition of direct explanation:

(D1) A set of sentences $T = \{T_1, \ldots, T_n\}$ $(n \geq 2)$ is a direct S-explanans for the purely universal sentence E if and only if

1) For all T_i in T, T_i is a purely universal law
2) $T \vdash E$
3) There is no sublaw SL of E such that $SL \vdash E$, $T \vdash SL$, and $SL \vdash T_i$.

In comparing definition (D1) with definition (K1), we find that there are, in essence, two more conditions in (K1) than in (D1). An analogue to condition 6) of (K1) will be added shortly, but we must first determine whether an analogue of condition 5) of (K1) is needed.

This condition, which may be called the 'verification condition', reads as follows: There is a class K of basic sentences such that $K \vdash C$ but neither $K \vdash E$ nor $K \vdash T$. Hempel introduced this condition in discussing the following problem. Without such a condition, he argued,

> ... any given particular fact could be explained by means of any true lawlike sentence whatsoever. More explicitly, if E is a true sentence- say, 'Mt. Everest is snowcapped', and T is a law- say 'All metals are good conductors of heat', then there always exists a true singular sentence C such that E is derivable from T and C, but not from C alone;.... Indeed, let T_s be some arbitrarily chosen particular instance of T, such as 'If the Eiffel Tower is metal, it is a good conductor of heat'. Now since E is true, so is the conditional $T_s \rightarrow E$, and if the latter is chosen as the sentence C, then T,C,E satisfy (the relevant conditions).[6]

Hempel's example in this passage may be exhibited as follows:

(H) T: $(x)(Mx \to Cx)$

C: $\underline{(Me \to Ce) \to S_m}$

E: S_m

In discussing this example further, Hempel argued that if such forms were accepted as explanatory, then C could only be verified by ascertaining the truth of E, and such explanations would be deprived of any predictive import. That is to say, E could not be predicted on the basis of T and C. Therefore, he introduced the 'verification' clause to insure that it be possible to verify C independently of E.

Yet quite apart from the argument about predictive import, and the development of his verification condition, Hempel was already in possession of a good reason for rejecting forms such as (H). In the first sentence of the passage above quoted, Hempel remarks that if such forms were allowed, then "any given particular fact could be explained by means of any true lawlike sentence whatsoever." This reason alone is sufficient for rejecting (H). But notice that this result is forthcoming only because we can always construct a C such that T and C together entail E. This C is constructed in such a way that it is a logical consequence of E, the assumed true explanandum. Therefore, if we disallow the entailment of C from E, on the grounds that if there were no such restriction then every true singular sentence would be explainable, then examples like (H) that depend on such an entailment will be excluded.

Now notice that this is precisely the effect of, and the rationale for, the "sub-law" requirement that I have introduced. This requirement precludes the truth of C from being dependent on, because it is a consequence of, the true sentence E. I therefore conclude that conditions (D1) do not require the addition of a "verification" condition.

We must now consider an analogue in condition 6) of (K1). Condition 6) reads as follows:

6) E is a disjunction of basic sentences.

This condition was introduced to simplify Kaplan's model. He required that a given explanandum sentence be reduced to conjunctive normal form, and that we then consider each conjunct separately for purposes of direct explanation. We shall follow Kaplan in this approach, and hence our final condition is this:

E is a disjunction of truth-functional components.

This statement is a little ambiguous. What is intended is that we put the sentence to be explained into a form similar to what Morgan has called "the shortest conjunctive normal form". We omit the qualification that such a normal form be the 'shortest'. What we arrive at by following this procedure is, where possible, a sentence which is a truth-functional disjunction whose disjuncts are purely universal laws. For example,

$(x)(Fx \rightarrow Gx)$

$(x)(Fx \rightarrow Gx) \vee (x)(Px \rightarrow Qx)$

$(x)(Fx \rightarrow Gx) \cdot (x)(Hx \rightarrow Gx)$

are all in sentential conjunctive normal form.

In summary, our analysis of "explanation" for general laws has led to the following set of conditions:

C1. A set of sentences $T = \{T_1, \ldots T_n\}$ $(n \geq 2)$ is a direct S-explanans for the purely universal sentence E if and only if

1) For all T_i in T, T_i is a universal law
2) $T \vdash E$
3) There is no sublaw SL of E such that $SL \vdash E$, $T \vdash SL$, and $SL \vdash T_i$.
4) E is a truth-functional disjunction

C2. A set of sentences $T = \{T_1, \ldots, T_n\}$ $(n \geq 2)$ is an S-explanans for the purely universal sentence E if and only if

1) T is a direct S-explanans for each of a set of laws $\{E_1, \ldots, En\}$, and
2) $\{E_1, \ldots, E_n\} \vdash E$.

Chapter V

Generality in Explanation

In the preceeding chapters we have discussed specific proposals for various models of scientific explanation stemming from the early work of Hempel and Oppenheim. We have also spoken of their more general conception of explanation as the deductive subsumption of sentences to be explained under more general principles, a view which the particular proposals were designed to capture.

It is to be noted that Hempel and Oppenheim were not the first to advance this general notion. Mill, for example, stated that "an individual fact is said to be explained by pointing out its cause, that is, by stating the law or laws of causation of which its production is an instance".[1] In regard to the explanation of laws or general regularities, he said: "a law of uniformity in nature is said to be explained when another law or laws are pointed out, of which that law itself is but a case, and from which it could be deduced."[2] Cohen and Nagel have advanced a similar view. Concerning the explanation of laws, they remarked: "Laws

---

1. Mill, J.S., A System of Logic, Book 3, Chap. 12, Sec. 4.
2. ibid.

themselves may be explained, and in the same manner, by showing that they are consequences of more comprehensive theories."[3] Finally, Karl Popper, in Conjectures and Refutations, remarked: "It can be said without paradox that scientific explanation is ... the reduction of the known to the unknown. In pure science, ... explanation is always the logical reduction of hypotheses to others which are of a higher level of universality."[4]

In this chapter we continue the analysis of the last chapter concerning the explanation of laws, but with the emphasis falling on a slightly different area. The last chapter dealt in a very general way with the syntactical structure of explanations of general regularities, while the present chapter attempts to deal with what we might call the content of such explanations.

The quotations above make reference to the fact that, in the words of Cohen and Nagel, "laws themselves may be explained ... by showing that they are consequences of more comprehensive theories." Indeed, as we have already seen, Hempel and Oppenheim themselves have recognized that mere deduction of a given law from other laws is not sufficient for explaining the given law; the laws

---

3. Cohen, M.R., and Nagel, E., An Introduction to Logic and Scientific Method, 1934, p. 397.

4. Popper, K., Conjectures and Refutations, 1962, p. 63.

constituting the explanans must in some sense be <u>more general</u> or more <u>comprehensive</u> than the law to be explained. Their discussion of the following example should be recalled:

(A) T: $\frac{K\&B}{K}$
E: .

The authors noted two points in relation to this example. (1) Their conditions were not intended to resolve the difficulty embodied in this example. As a result, they restricted their model to apply only to <u>singular</u> explananda. (2) In order to deal adequately with explanations of laws, they emphasized that some criteria are needed for distinguishing the relative comprehensiveness of laws. Our concern in this chapter will fall mainly on this second point.

5.1. <u>Some misconceptions</u>

Before dealing with some issues relating to (2), above, some remarks concerning (1) are in order.

Point (1) simply gives the reason for restricting sentences to be explained to singular sentences. There is no obvious way of dealing with example (A) short of giving an analysis of "more general" as this expression applies to laws.

Yet the fact that the D-N model is explicitly restricted to singular explananda does not seem to be generally recognized. For example, a recent paper by Baruch Brody finds fault with the covering-law model on the grounds that the following example of an explanation of a law meets all the requirements of that model but is clearly, Brody argues, not an adequate explanation.

(B) 1. Sodium normally combines with Bromine in a ratio of one to one.
2. Everything that normally combines with Bromine in a ratio of one to one normally combines with Chlorine in a ratio of one to one.
3. Therefore, Sodium normally combines with Chlorine in a ratio of one to one.[5]

Concerning this example, Brody remarks: "This purported explanation meets all of the requirements laid down by Hempel's covering-law model for scientific explanation.... After all, the law to be explained is deduced from two other general laws which are true and have empirical content. Nevertheless, this purported explanation seems to have absolutely no explanatory power.... Why is it that it is not as good an explanation as the explanation of that law in terms of the atomic structure of sodium and chlorine and the theory of chemical bonding?"[6]

Given that Hempel explicitly limits his model to singular explananda, Brody's comment that example (B) "meets

---

5. Brody, B., "Towards an Aristotelian Theory of Scientific Explanation", Philosophy of Science, (39) 1972, p. 20.
6. ibid.

all of the requirements laid down by Hempel's covering-law model" is perplexing. Upon checking his reference, I found Brody to be referring to Hempel and Oppenheim's <u>conditions of adequacy</u>. Now as I understand it, conditions of adequacy are usually proposed to insure that a definition captures certain intuitively desirable requirements; the adequacy conditions may be necessary, sufficient, or both. In the case of Hempel and Oppenheim, they propose the following conditions of adequacy:[7]

I. Logical conditions of Adequacy:

(R1) The explanandum must be a logical consequence of the explanans; ...

(R2) The explanans must contain general laws, and these must actually be required for the derivation of the explanandum....

(R3) The explanans must have empirical content; ...

II. Empirical condition of Adequacy:

(R4) The sentences constituting the explanans must be true....

These conditions are what Brody refers to as "the requirements laid down by Hempel's covering-law model for scientific explanation". And <u>if</u> these conditions were offered as <u>sufficient</u> conditions for explanation in general, then Brody's

---

7. Hempel, C., <u>Aspects of Scientific Explanation</u>, pp. 247-249.

point against any definition of explanation which satisfied such conditions would be well taken. For, as Brody points out, example (B) satisfies all of these conditions (rather, any definition of explanation that satisfies these conditions of adequacy would count (B) as explanatory).

Yet one should suspect that something is amiss once it is recognized that Hempel's own example (A) also satisfies these adequacy conditions. What has gone wrong, I believe, is Brody's reading of Hempel.

There seems to be at least two points on which Brody is in error. First, these adequacy conditions are laid down very early in the Hempel and Oppenheim paper (in part I) after they give some examples of the sorts of explanations they intend their model to encompass. They emphasize that this section is _elementary_[8] and that the adequacy conditions are formulated in a slightly vague manner.[9] In a later section (part III) they explicitly acknowledge that the criteria (R1)-(R4) are insufficient for the explanation of laws _as well as for explanation of particular events_. Their example (A), i.e.,

$$\begin{array}{ll} T: & \underline{K\&B} \\ E: & K \end{array}$$

satisfies (R1)-(R4), and it is partly for this reason that

---

8. ibid., p. 247.
9. ibid.

they limit their analysis to singular explananda. And the authors give several examples of explanations of singular events which also satisfy (R1)-(R4), but which they regard as inadequate. Therefore, Brody has taken Hempel and Oppenheim's preliminary statement of their adequacy conditions as their final definition, ignoring the subsequent revisions proposed by those authors.

Secondly, Brody misunderstands what is the true status of these conditions. _Even if_ he were correct in assuming the conditions (R1)-(R4) to be the final statement of the requirements laid down by the covering-law model, it is evident that the authors intend these conditions to be only _necessary_, and not sufficient, conditions. Brody might have been misled by the fact that the statement of the conditions neglected to explicitly introduce them as necessary conditions, but in the paragraph immediately following the introduction of these conditions the authors state that "the same formal analysis, including the four _necessary conditions_, applies to scientific prediction as well as to explanation."[10] Indeed, as we have already noted, when Hempel and Oppenheim later propose a more refined analysis in part III of their paper, they are even more

10. ibid., p. 249.

explicit on this point. They say:

> In analogy to the concept of lawlike sentence, which need not satisfy a requirement of truth, we will first introduce an auxiliary concept of potential explanans, which is not subject to a requirement of truth; the notion of explanans will then be defined with the help of this auxiliary concept. The considerations presented in Part I suggest the following initial stipulations:
> (7.5) An ordered couple of sentences, (T,C), constitutes a potential explanans for a singular sentence E _only if_
> 1) T is essentially generalized and E is singular.
> 2) E is derivable in L from T and C jointly, but not from C alone.
>
> (7.6) An ordered couple of sentences, (T,C), constitutes an explanans for a singular sentence E if and only if
> 1) (T,C) is a potential explanans for E
> 2) T is a theory and C is true.[11]

Hempel and Oppenheim continue:

> (7.6) is an explicit definition of explanation in terms of the concept of potential explanation. On the other hand, (7.5) is not suggested as a definition, but as a statement of necessary conditions of potential explanation. These conditions will presently be shown not to be sufficient and additional requirements will be discussed by which (7.5) has to be supplemented in order to provide a definition of potential explanation.[12]

Therefore, we conclude that it is not the case that any set of sentences which satisfy (R1)-(R4) is an adequate explanation. The claim is simply that if such a set of sentences _does not_ satisfy (R1)-(R4), then it is not an adequate explanans. Brody's claim vis-a-vis his example (B), that

---

11. ibid., p. 273.
12. ibid.

it meets all of the requirements laid down by Hempel's covering-law model, is simply false. For the requirements he mentions are only necessary conditions, and not sufficient ones, and even when Hempel adds another condition to make them jointly sufficient, he explicitly restricts the resulting definition to singular explananda. Brody has clearly misrepresented the Hempel and Oppenheim analysis, and has attributed to them a position which they did not in fact hold.

Before leaving this section let us return briefly to the main issue, that of the restriction of the Hempel and Oppenheim model to singular explananda. In the last chapter and at the beginning of this chapter we have noted that Hempel's main reason for limiting his definition to singular explananda was that he was not able at that time to produce criteria for a comparison of generalized sentences as to their comprehensiveness. Let us examine his example in somewhat more detail to insure that we are clear on this issue.

It is possible to 'explain' Kepler's laws, K, according to (R1)-(R4), by conjoining them with Boyle's law, B, to form the stronger law K&B. But the derivation of K from the latter would not be considered as an explanation of the regularities stated in Kepler's laws. On the other hand, derivation of Kepler's laws from Newton's laws of motion and of gravitation (assuming for the moment, for illustrative purposes, that such a derivation is possible) would be recognized as a legitimate explanation in terms of more comprehensive regularities.[13] Thus, one other plausible requirement for satisfactory explanations of laws appears to be that the explanans contain at least one law which is in some sense "more general" than the law to be explained.

Now although the derivation of K from the stronger law K&B is eliminated by the model that I have so far suggested (since the explanandum has one of the premises as a logical consequence), it is clear that such a model does <u>not</u> deal satisfactorily with the demand that the explanans be more general than the explanandum. For instance, an argument of the form

$$\frac{\begin{array}{l}(x)(Px \rightarrow Qx \\ (x)(Qx \rightarrow Rx)\end{array}}{(x)(Px \rightarrow Rx)}$$

(which is the form of Brody's argument (B)), satisfies our earlier model but there is no guarantee that one of the premises is "more general" than the explanandum. Indeed, this is the real point behind Brody's example, although he does not seem to see this.

It is clear, therefore, that we must talk directly to the issue of generality.

---

13. See Hempel, C., <u>Aspects of Scientific Explanation</u>, p. 273.

5.2. Preliminary analysis of "more general"

At the first of this chapter we noted that several authors have stated that explanations of laws involve other laws in the explanans which are more "comprehensive", more "universal", or more "general. The sense in which one law is said to be more general than another is usually indicated by way of example. For instance, it is claimed that Archimedes' law- i.e., the buoyant force of a liquid upon a body immersed in it is equal to the weight of the liquid displaced by that body- is more general than the law that ice immersed in water floats, because the former asserts something of all liquids and not only of water, and of all bodies immersed in liquids and not only ice. But to just cite examples is not sufficient to explicate a concept.

Therefore, let us try to be a little more precise, and investigate what could be meant by the claim that "statement S_1 is more general in L than a second statement S_2".

First, as Nagel notes,[14] the claim in not intended to mean that S_1 must logically imply S_2. In other words, it is not a necessary condition for one statements being more general than another, since a relation of implication does

---

14. Nagel, E., The Structure of Science, 1961, p. 37.

not hold between Archimedes' law and the law that ice floats in water, despite the fact that the former is more general than the latter. Nor is it a sufficient condition, according to Nagel, "since 'all planets move on elliptic orbits' logically implies 'all planets move on orbits which are conic sections', but presumably the first of these is not more general than the second."[15]

Karl Popper, in Logic of Scientific Discovery, has offered a more precise account of generality. Consider for the moment only laws which are capable of being stated as universal conditionals of the simplest form, i.e., $(x)(Px \rightarrow Qx)$. Let S_1 be a sentence of the form

$$S_1 \quad (x)(Ax \rightarrow Bx)$$

and S_2 a sentence of the form

$$S_2 \quad (x)(Cx \rightarrow Dx).$$

According to Popper's analysis, S_1 is more general (or "more universal") than S_2 just in case '$(x)(Cx \rightarrow Ax)$' is analytically true, but its converse is not.[16] For example, let S_1 be Archimedes' law and S_2 the law that ice immersed in water floats. Now this particular S_1 is more general than S_2 since the antecedent of S_2 logically implies that of S_1, but not vice versa; i.e., since "Ice immersed in water is an object

---

15. ibid.
16. Popper, K., Logic of Scientific Discovery, pp. 122-123.

immersed in a liquid" is true by virtue of the meanings associated with its terms, while the converse is not.

What this requires is simply that the class of objects in the extension of the antecedent of S_2 be included in the class of objects in the extension of the antecedent of S_1. However, before settling on the status of this definition, let us first consider an objection put forth by Nagel.

Nagel argues that the requirement that two logically equivalent sentences should be equally general seems reasonable, so that if S_1 is more general than S_2, and S_1 is logically equivalent to S_3, then S_3 is also more general than S_2. He then points out that this requirement is not satisfied by Popper's definition of "more general". Consider the following three sentences meeting the conditions just described:

S_1: All living organisms are mortal.

S_2: All human beings are mortal.

S_3: All non-mortals are non-living organisms.

S_1 is more general than S_2, and S_3 is equivalent to S_1, but S_3 is not more general than S_2 since "all human beings are non-mortal" is certainly not analytically true.

Nagel comments extensively on this issue, as follows:

> These difficulties are not necessarily fatal to the proposed explication of the notion of greater generality- but to avoid them one must drop the seemingly plausible requirement that logically equivalent statements must be equally general, and

> adopt the position that the comparative generality of laws is relative to the way they are formulated. It might be objected, however, that such a course opens the door to unlimited arbitrariness in classifying laws according to their generality, since for a given statement there are an infinite number of logical equivalents differing in their formulation.[17]

However, Nagel also points out the following:

> Nevertheless, the arbitrariness may not be as serious as it looks at first sight. For the actual formulation of a law frequently indicates what is the range of things that are the subjects of predication in given contexts, where this identification of the intended scope of the law is controlled by the nature of the particular inquiry. But there is nothing especially arbitrary in this, other than the arbitrariness inherent in dealing with one set of problems rather than with another set. Accordingly, insofar as the subject term in the statement of a law indicates the intended scope of the law in a concrete context (or class of contexts) of its use, the assertion that one law is more general than another is not fatally arbitrary- even if in some other context a different comparative judgement is required. For example, the law that ice floats in water is commonly so used that its range of application is the indefinitely large class of instances of ice which are (or have been or will be) immersed in water. The law is rarely if ever so used so that its range of application is taken to be the miscellaneous collection of things which do not float in water (whether in the past, present, or future). Indeed, it is a plausible claim that were the law used in this latter way in some context, its customary formulation would in that context be appropriately modified. At any rate, there appears to be a tacit reference to contexts of use in the actual formulations of laws.[18]

In essence, Nagel suggests that we give up the requirement that logically equivalent sentences be equally general in

---

17. Nagel, E., Structure, p. 39.
18. ibid.

favor of the view that the notion of "more general" be relativized to contexts, to what he calls "contexts of use". We agree that the notion should be relativized to contexts, but we hesitate to follow him in choosing contexts of _use_ as the relevant contexts. And even though we wish to follow Nagel in his general approach, we also wish to make more precise the notion of "range of application" of a law to which Nagel alludes in the above quotation.

The importance (for our purposes) of making this notion more precise is obvious. To take our previous example of the sentences S_1, S_2, and S_3, we should like to find a general way of characterizing the "range of application" of the generalization "all living orgamisms are mortal" so that "living organisms" rather than "non-mortals" is singled out as what we shall call the _relevant_ reference class. This done, we could than say that S_1 is more general than S_2 in the relevant reference class.

For this task, we rely on some recent results in confirmation theory, on which we will now briefly digress.

5.2.a. _Recent work on the "paradoxes" of confirmation_

Our task is to provide a general method of determining the relevant reference class of a generalization. For instance, we would like this method to "pick out" the reference class of "living organisms" as the relevant reference class relative to the generalization "all living organisms are mortal" as opposed to the reference class "non-mortals". Or to take another example mentioned by Nagel in the quotation above, we want our method to pick out the reference class "ice immersed in water" as the relevant reference class of the generalization "all ice immersed in water floats" as opposed to the reference class "non-floating things".

Once we have such a general method, we will then have a precise way of relativizing the notion "more general" to contexts, i.e., to the relevant reference class, and we may then say that one generalization is more general than another in the relevant reference class. What's more, these contexts will be the intuitively 'natural' ones, such that the choice of any other reference class would be highly artificial. For example, the generalization "all living organisms are mortal" is intuitively about living organisms, and not about non-mortals. Thus, we shall say that "all living organisms are mortal" is more general than "all human beings are mortal" in the relevant reference class of living organisms, and that the choice of any other reference class, such as the class of non-mortals, in which the relation of "more general" does not hold, is not the relevant reference class, and is therefore inappropriate.

Our examples already show an affinity with some problems surrounding the "paradoxes" of confirmation. These "paradoxes" arise in the following way. Consider generalizations of the form "All A are B". We can divide the things of which A and B can be meaningfully predicated into four exhaustive and mutually exclusive classes:

(1) A&B

(2) A&$\bar{B}$

(3) $\bar{A}$&B

(4) $\bar{A}$&$\bar{B}$

On what is known as the "Nicod Criterion" of confirming instance, <u>only</u> (1) is a confirming instance of "All A are B". For example, only a black raven confirms "All ravens are black".

Now consider "All non-B are non-A". By the Nicod Criterion, $\bar{B}$&$\bar{A}$ confirms this, but this is (4) above. Since "All non-B are non-A" is equivalent to "All A are B", it is reasonable to think that (4) also confirms "All A are B". This requirement on the notion of a confirming instance is usually called "the Equivalence Condition". But almost any sort of observation is permitted under the $\bar{A}$&$\bar{B}$ heading, for example, in the raven case, white shoes, red chairs, and perhaps, according to Suppes, even non-black thoughts. Thus, the Nicod Criterion and the Equivalence Condition, when used together, have the consequence that a white shoe or a non-black thought confirms the generalization "All ravens are black".

In a fairly recent article,[19] G.H. Von Wright introduces the phrase "range of relevance of a generalization" and argues for the following thesis: All things in the range of relevance of a generalization may constitute genuine confirmations or disconfirmations of the generalization. Things outside the range of relevance are irrelevant to the generalization- they can't confirm it "genuinely". Since, he argues, they do not disconfirm it either, we may "by courtesy" say that they confirm it, though only "paradoxically".[20]

Given the notion of "range of relevance", Von Wright advances the following question: Is it possible to confirm genuinely the generalization that all ravens are black through the observation of white shoes? His answer is that this is possible or not, depending upon which is the range of relevance of the generalization, "upon what the generalization is 'about'".[21]

The only problem I find with Von Wright's thesis is that he has little constructive to say concerning the range of relevance. In order to tell what the range of relevance is, he says we have to <u>specify</u> the range. When the range of relevance of a generalization, say, $(x)(Ax \rightarrow Bx)$,

---

19. Von Wright, G., "The Paradoxes of Confirmation", in Hintikka and Suppes, <u>Aspects of Inductive Logic</u>.
20. ibid., p. 211.
21. ibid., p. 215.

is not specified, he then understands it to be the class of things which fall under the antecedent term A- which he calls the natural range of relevance.

In summarizing his view, Von Wright raises anew the question of the status of the Equivalence Condition. He asks, do we wish to deny that the generalization "All A are B" is the same generalization as "All non-B are non-A"? He answers:

> We do not wish to deny that "all A are B" as a generalization about things which are A expresses the very same proposition as "all non-B are non-A" as a generalization about things which are A. Generally speaking: when taken relative to the same range of relevance, the generalization all A are B and the generalization all non-B are non-A are the same generalization. But the generalization that all A are B with range of relevance A is a different generalization from all non-B are non-A with range of relevance B. If we agree that, range of relevance not being specified, a generalization is normally taken relative to its "natural range", then we should also have to agree that all A are B and all non-B are non-A normally express different generalizations. The generalizations are different because their "natural" ranges of relevance are different.[22]

However, Von Wright's discussion of the paradoxes does not really help us in our task of settling on a relevant reference class. According to him, taken relative to their "natural" ranges of relevance the generalization "all living organisms are mortal" is different from the generalization "all non-mortals are non-living organisms". What we are in search of is a method which allows us to fix on one of these

22. ibid., p. 217.

reference classes, either "living organisms" or "non-mortals", as being in some sense more relevant than the other. That is, take the raven example once again. Is there any good reason for taking the generalization "all ravens are black" to be "about" ravens, rather than non-black things? I think that there is.

Patrick Suppes, in "A Bayesian Approach to the Paradoxes of Confirmation",[23] provides some considerations in light of which we are able to answer this question in an unambiguous manner. His approach to the paradoxes can be perhaps best described in the following fashion.

Starting with the generalization (x)(Rx→Bx), we divide once again the things of which R and B can be meaningfully predicated into four mutually exclusive and jointly exhaustive categories, as above:

(1) Rx&Bx

(2) $Rx\&\bar{B}x$

(3) $\bar{R}x\&Bx$

(4) $\bar{R}x\&\bar{B}x$

In just which way we choose the universe of objects is not crucial for our purposes. We simply require that it be any broadly chosen set of objects.

To the sets corresponding to (1)-(4), we assign prior probabilities. Using Suppes' notation, we have

---

23. In Hintikka and Suppes, <u>Aspects of Inductive Logic</u>.

$P(\{x: Rx\&Bx\})=p_1$

$P(\{x: Rx\&\bar{B}x\})=p_2$

$P(\{x: \bar{R}x\&Bx\})=p_3$

$P(\{x: \bar{R}x\&\bar{B}x\})=p_4$

It is assumed that these probabilities sum to 1, and that all are non-zero. It is clear that, in any broadly chosen universe of objects, p_4 will be much larger than any of the other probabilities. That is, a random selection from the universe of objects at large ought to yield with greater likelihood a non-black non-raven as opposed to, say, a black raven.

Suppes then attempts to justify the intuitive assumption that we should look at randomly selected ravens and not randomly selected non-black things in testing the generalization that all ravens are black. That is, we are interested in the following conditional probabilities:

(i) $P(B/R)$

and

(ii) $P(\bar{R}/\bar{B})$,

which are read, respectively, as "the probability that an object is black, given that it is a raven" and "the probability that an object is a non-raven, given that it is non-black".

To evaluate these probabilities, we first require the probabilities of $R=\{x:Rx\}$ and $B=\{x:Bx\}$, which are:

(iii) $P(R)=p_1+p_2$

since, by the theorem on total probability,

$P(R)=P(R\&B)+P(R\&\bar{B})$,

and

(iv) $P(B)=p_1+p_3$,

for the same reason.

We may now evaluate (i) and (ii), as follows:

$$\text{(i')}\ P(B/R)=\frac{P(B\&R)}{P(R)}=\frac{p_1}{p_1+p_2}$$

$$\text{(ii')}\ P(\bar{R}/\bar{B})=\frac{P(\bar{B}\&\bar{R})}{P(\bar{B})}=\frac{p_4}{p_2+p_4}$$

The sampling rule we want to justify is this: look at R's rather than non-B's if $P(B/R)<P(\bar{R}/\bar{B})$, that is, if the following inequality holds:

$$\text{(v)}\ \frac{p_1}{p_1+p_2} < \frac{p_4}{p_2+p_4}\ .$$

We note that (v) is true if and only if $p_1<p_4$.

That is to say, our sampling rule is this: to test the generalization that all R are B, we look at R's rather than non-B's if the probability that an object is black, given that it is a raven, is less than the probability that a randomly selected object is a non-raven, given that it is non-black. Our sampling rule thus comes to this: look at ravens rather than non-black things when $p_1<p_4$.

Suppes justifies this rule by the following argument. He says: "In sampling objects to confirm or disconfirm the general law (x)(Rx→Bx), we want to test the law. This, I take it, means that we want to sample items with a higher prior probability of disconfirming the law".[24] To show that our sampling rule requiring that R's be selected rather than non-B's for testing the generalization when $p_1<p_4$, consider the following argument showing that, when $p_1<p_4$, sampling R's has a higher prior probability of disconfirming the law.

The probability that an object is non-black, given that it is a raven, is

$$\text{(vi)}\quad P(\bar{B}/R)=\frac{P(B\&R)}{P(R)}=\frac{p_2}{p_1+p_2}\ .$$

The probability that an object is a raven, given that it is non-black, is

$$\text{(vii)}\quad P(R/\bar{B})=\frac{P(R\&\bar{B})}{P(\bar{B})}=\frac{p_2}{p_2+p_4}\ .$$

According to these equations, the selection of an R has a higher prior probability of disconfirming the law than a selection of a non-B just when

$$\text{(viii)}\quad P(R/\bar{B})<P(\bar{B}/R)$$

i.e., just in case

---

24. ibid., pp. 199-200.

$$\frac{p_2}{p_2+p_4} < \frac{p_2}{p_1+p_2} ,$$

which holds if and only if

$$p_1<p_4.$$

This clearly demonstrates that ravens ought to be examined, as opposed to non-black things, when $p_1<p_4$.

Suppes goes on to point out that "the adoption of a rational rule for what to observe or sample does not follow from the prior probabilities alone. Some other ingredient must be added, but the rule that tells us to select an R rather than a non-B when $p_1<p_4$ follows from any number of more general principles".[25] His discussion of these more general principles, e.g., a principle of minimizing cost or effort, is not important for our purposes. The interested reader is referred to Suppes' paper on this matter.

We now have a precise way of picking out what I shall call the "relevant reference class" of a generalization. The relevant reference class, in the example of the raven generalization, will just be the "test" class of ravens, in the sense of Suppes, and for exactly the reasons he provides. It is to such a reference class that we shall relativize our definition of "more general". Our earlier dilemma

---

25. ibid., p. 200.

concerning the generalizations "all living organisms are mortal" and "all non-mortals are non-living organisms" will be relativized to the relevant reference class of living organisms, since this is the class in which we have the greatest chance of falsifying, i.e., testing, the generalization.

To be perfectly clear, let us recall our earlier problem. According to Popper's criterion, where

S_1= All living organisms are mortal

S_2= All human beings are mortal,

S_1 is more general than S_2. But when we note that

S_3= All non-mortals are non-living organisms

is logically equivalent to S_1, we find that S_3 is <u>not</u>, on the proposed criterion, more general than S_2. Yet, despite the equivalence of S_3 to S_1, the relevant reference class of these two generalizations, the "test" class in Suppes' terms, is the set of living organisms. If we define "more general" relative to the relevant reference class, then S_1 is more general than S_2 in this class. The reference class (i.e., the antecedent of S_3) of non-mortals is not the relevant one, so our procedure instructs us to choose S_1 as the generalization against which S_2 is to be compared (since the reference class determined by the antecedent of S_1 is just what we have called the <u>relevant</u> reference class).

This is, I think, a precise way of explicating Nagel's suggestion that the notion of "more general" be relativized to certain contexts. The context we have settled for is the test class of a generalization.

In our raven example, once we have settled on what is the relevant reference class, we choose that form of the generalization whose antecedent is the set of things in the relevant reference class. That is, given that the relevant reference class of the generalization "all ravens are black" is the class of ravens, and not the class of non-black things, we are instructed to adopt the formulation "all ravens are black" rather than "all non-black things are non-ravens" for purposes of comparing the generality of this hypothesis with the generality of another.

5.3. Generality in explanation

We now have a definition of "more general" and we ought to amend our provisional definition of explanation given in the preceeding chapter. Before doing this, however, we should emphasize that our treatment of "more general" is more programmatic and tentative rather than final. By this it is meant that our analysis is severely restricted insofar as we have only treated of laws of the most elementary form, i.e., simple universal conditionals of the form $(x)(Ax \rightarrow Bx)$. Needless to say, our analysis is unable to provide criteria for the comparison of complex scientific theories, or even laws involving mixed quantification, as to their relative generality.

Our revised definition of explanation can be stated as follows, with condition 1a) incorporating the results of the previous section:

(C1') A set of sentences $T \div \{T_1, \ldots, T_n\}$ is a direct S-explanans for the purely universal sentence E if and only if

1) Every T_i in T is a universal law

1a) Some T_i in T is more general than E

2) $T \vdash E$

3) There is no sublaw SL of E such that $SL \vdash E$, $T \vdash SL$, and $SL \vdash T_i$.

4) E is a truth-functional disjunction.

Clause (C2) of the previous definition remains the same.

It seems that we have now at least in part satisfied Hempel's demand for a criterion for assessing the relative generality of laws. We have required that at least one member of the explanans be a law which is more general than the explanandum, and we have specified under what conditions a sentence S_1 is more general than another sentence S_2 in the relevant reference class.

Let us now reconsider Brody's counter-example mentioned at the beginning of this chapter, concerning the explanation of why sodium combines with chlorine in a certain ratio. This example has the following structure:

(B')

1. (x)(Sx→Bx)
2. (x)(Bx→Cx)
3. (x)(Sx→Cx)

with the obvious scheme of abbreviation. Now let us see if this example satisfies the revised definition (Cl').

In order to satisfy (Cl'), it needs to be shown that one of the premises is more general than the explanandum, since all the remaining conditions are obviously satisfied. Consider the two possible cases. First, consider premise 1. This premise is not more general than the explanandum, for by our definition, relative to the relevant reference class "things which are sodium", it is not the case that (x)(Sx→Sx) is analytically true <u>and</u> its converse is not. Second, consider premise 2. This premise is also not more general than the explanandum, for (x)(Sx→Bx) is not analytic.

The alternative explanation alluded to by Brody (but not provided) in terms of the atomic structure of sodium, etc., will obviously involve a premise that is more general than the explanandum, and will therefore qualify as an explanation on our account.

We can conclude, therefore, that Brody's criticism of the D-N model is not well-founded. His claim that example (B') above meets all the requirements of the covering-law model is simply false, for the covering-law model is not defined for explananda which are laws. And when the requirements laid down by Hempel are supplemented with a condition which Hempel had required but had not specified, Brody's counter-example clearly does not satisfy the resulting set of requirements. His counter-example does point out certain difficulties, but these difficulties were already acknowledged to exist by Hempel and Oppenheim.

In the following section, we evaluate the model which Brody has offered to resolve the difficulty embodied in example (B'), and we argue that his proposed solution is inadequate.

5.4. Criticism of Brody's Aristotelian analysis

In Brody's paper alluded to at the beginning of this chapter, he proposes that two new conditions be added to the "Hempelian model" in order to avoid the difficulties embodied in example (B) and in order to make the resulting set of conditions sufficient. I say "Hempelian model" because the model with which Brody takes issue is not Hempel's. For as previously mentioned, Brody mistakes Hempel's statement of adequacy conditions as his final position on explanation, *even for laws*, which Hempel explicitly denies. These conditions of adequacy were satisfied by Hempel's definition of the necessary conditions for a *potential* explanans, but he realized that other conditions were required in order to have a definition of a plausible set of necessary and sufficient conditions for explanation. Even by adding these other conditions, however, the resulting model was inadequate. But the major error committed by Brody is to be found in his claim that Hempel's model of explanation for laws is inadequate; this because Hempel offered no such model.

Yet in order to properly evaluate Brody's own proposal, let us assume that the conditions (R1)-(R4) do constitute Hempel's account of explanation for laws, and then let us see if Brody's proposed solution does the required job. We argue that it does not.

Recall again Brody's counter-example concerning the combining of sodium, chlorine, and bromine, i.e.,

(A) 1. $(x)(Sx \rightarrow Bx)$

2. $\underline{(x)(Bx \rightarrow Cx)}$

3. $(x)(Sx \rightarrow Cx)$

This example is contrasted with one sketched in terms of the atomic structure of the different elements. Concerning the given example, we have noted that Brody claims that "this purported explanation meets all of the requirements laid down by Hempel's covering law model for scientific explanation.... After all, the law to be explained is deduced from two other general laws which are true and have empirical content."

Suppose, contrary to fact, that Brody's remark is well taken. How does he attempt to solve the problem? He states that examples such as (A) led him to believe that there was something fundamentally wrong with the whole covering-law model, but he still felt that this model, which fits so many cases and seems so plausible, just could not be junked entirely. He then claims that Aristotle had already recognized these problems and had solved them in a satisfactory manner.

Brody's (Aristotle's?) solution consists in adding the two following requirements to the "Hempelian" model:

1.a. A D-N explanation of a particular event is a satisfactory explanation of the event when (besides meeting all of Hempel's requirements) its explanans contains essentially a description of the event which is the cause of the event described in the explanandum.

1.b. A D-N explanation of a law is a satisfactory explanation of that law when (besides meeting all of Hempel's conditions) every event which is a case of the law to be explained is caused by an event which is a case of one (in each case, the same) of the laws contained essentially in the explanans.

2.a. A D-N explanation of a particular event is satisfactory when (besides meeting all of Hempel's requirements) its explanans contains essentially 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 when at least one object involved in the event described in the explanandum is a member of that class of objects.

2.b. A D-N explanation of a law is a satisfactory explanation of that law when (besides meeting all of Hempel's requirements) each event which is a case of the law which is the explanandum, involves an entity which is a member of a class (in each case, the same class) such that the explanans contains a statement attributing to that class a property which each of its members have essentially (even if the statement does not say that they have it essentially).[26]

Let us take an example of how one of these conditions is supposed to function. Consider 2.b. This is designed to eliminate example (A) and retain the alternative explanation in terms of the atomic structure of sodium, etc. Example (A), according to Brody, does not contain a premise ascribing an essential property (in the sense of Aristotle) to a class of objects. The explanation he mentions (but never exhibits) in terms of the atomic structure of sodium,

---

26. Brody, pp. 23-27.

etc., <u>does</u>, he claims, ascribe an essential property to a class of objects, namely, a premise ascribing a certain atomic number to sodium. In such a way, (A) is to be rejected in favor of an explanation in terms of the atomic structure of the given elements. Let's very crudely construct this alternative explanation as follows (this is a very crude example, but it only serves to illustrate a point, and nothing turns upon our particular formulation):

$(x)(Sx \rightarrow Ax)$

$(x)(Ax \rightarrow Bx)$

$\underline{(x)(Bx \rightarrow Cx)}$

$(x)(Sx \rightarrow Cx)$

This argument is to receive the same interpretation as example (A), with the exception that the first premise is intended to ascribe a certain atomic number to sodium. The point of this example is simply to illustrate that the explanans must, in Brody's view, contain a statement to the effect that a certain class of objects has <u>essentially a</u> certain property. In Brody's view, the atomic number of sodium is an essential property of that substance.

According to 2.b., this argument is satisfactory if the explanans attributes an essential property to a class of objects (which premise 1. does) <u>and</u> when one other condition is satisfied: namely, when <u>a case</u> of the explanandum involves an entity which is a member of the class mentioned in the premise which ascribes an essential property to the same class. We might understand "a case" of the explanandum as being any instantial instance of it, say, Sa→Ca, and this 'case' does involve a member (viz., 'a') of a class to which a property is predicated essentially (viz., the class determined by 'S').

Brody goes on to argue that this analysis is immune to the criticism embodied in Theorem 1 of Eberle, Kaplan, and Montague, which he claims is a subtle form of self-explanation. He remarks:

> "As Hempel recognized, we need some additional requirement to rule out such obvious self-explanations as
>
> $$\frac{\begin{array}{c}(x)(Px)\\ Qa\end{array}}{Qa}$$
>
> and
>
> $$\frac{\begin{array}{c}(x)(Px\&Qa)\\ Qa\vee\sim Qa\end{array}}{Qa} \quad .$$
>
> He proposed a simple solution to that problem but Eberle, Kaplan, and Montague showed that it wouldn't do. Consider, they said, the following example of a bad explanation that meets all of Hempel's requirements, of an objects having a property H. Take any law of the form (x)(Fx) (no connection between F's and H's). From that

law it follows that (where G is any third unrelated property)

(1) $(x)(y)(Fx \vee (Gy \rightarrow Hy))$.

It also follows from Ha, the fact to be explained, that

(2) $(Fb \vee \sim Ga) \rightarrow Ha$.

But from these two statements, we can derive Ha, and this derivation, a subtle form of self-explanation, meets all of Hempel's requirements, so Hempel still had not solved the problem of self-explanation.[27]

Brody claims that his theory offers a simple, non ad-hoc solution to this problem. The derivation of Eberle, Kaplan and Montague does not meet either of his two conditions, since, he claims, (a) neither (1) nor (2) describe the cause of Ha, and (b), neither (1) nor (2) ascribe an essential property to a certain class of objects of which 'a' is a member.

Therefore, Brody feels that by adding these two new conditions to the Hempelian requirements certain advantages accrue, in particular that such a model provides a satisfactory solution to problems that the covering law model can not handle.

---

27. ibid., pp. 28-29.

Yet there is reason to believe that Brody's analysis is not entirely satisfactory, even granting him his use of "cause" and ignoring certain problems with his "essential properties" view. Let us look once again at the example which forced Hempel to restrict his definition of explanation to singular explananda. According to the "Hempelian" requirements for explanations of laws, as Brody sees it, the following is satisfactory:

$$\begin{array}{ll} T: & \underline{K\&B} \\ E: & K \end{array}$$

Hempel recognized the fact that this would satisfy his conditions (R1)-(R4), hence he left the solution of such difficulties as an open problem (the difficulty here is that this is a case of self-explanation). Does Brody's analysis constitute a resolution of this problem?

To be perfectly clear, let us take another example, wherein one premise ascribes an essential property to a certain class of objects, say, the law that sodium has such-and-such an atomic number. This is Brody's own example of an ascription of an essential property to a class of objects. Call this premise S_1, i.e.,

$$S_1 = (x)(Sx \rightarrow ANx).$$

Now conjoin S_1 with any other law, say S_2. Then, according to Brody's analysis, the following is a satisfactory explanation:

$$\frac{T:\ (x)(Sx \to ANx) \& S_2}{E:\ (x)(Sx \to ANx)}$$

This example satisfies condition 2.b. in Brody's definition of explanation for laws. To see this, take a 'case' of the law which constitutes the explanandum, say (Sa ANa). By 2.b., this sentence must involve an entity (Namely, 'a') which is a member of a class such that the explanans contains a premise attributing to that class (here, the class determined by the extension of S) a property which each of the members of that class have essentially. But, by assumption, all the members of S have the "essential" property 'AN', thereby satisfying Brody's condition. Therefore, despite Brody's claim to the contrary, his analysis does not solve the problem of self-explanation, at least not in cases where the sentence to be explained is a law.

Furthermore, the type of counter-example to which Brody's analysis is susceptible is the same as the type which led Hempel to restrict his definition to singular explananda in the first place. Thus, Brody has ignored Hempel's request for a criterion to characterize the relative comprehensiveness of laws.

Of course, Brody should have been aware of the fact that Hempel did not define explanation relative to laws at all (i.e., where the explanandum is a law).

5.5. Solutions to other problems: explanation and confirmation

In this section we make use of our account of explanation to follow up a point made by Brody in a somewhat different context.

In an earlier article entitled "Confirmation and Explanation",[28] Brody questions an argument advanced by Hempel which shows that if a qualitative confirmation function satisfies the following two conditions, then any evidence that confirms any sentence confirms every sentence:

1. Special-consequence condition: If a statement confirms a hypothesis H, it also confirms any consequence of H.

2. Converse-consequence condition: If a statement confirms a hypothesis H, it also confirms every hypothesis that entails H.

In light of this result, various authors have suggested dropping one or even both of these adequacy conditions. Hempel, in particular, gives up the converse-consequence condition.

Brody argues that there are several reasons why we ought not to simply drop one or both of these conditions. One reason is that there are many acceptable inferences that are justified if one's qualitative confirmation function satisfies these conditions. In any event, he feels it desirable to preserve the import of these conditions, if at all possible.

---

28. In Brody, B., Readings in the Philosophy of Science, 1969.

In order to isolate the problem, let us consider Hempel's argument:

1. Suppose E confirms H_1.
2. H_1&H_2 entails H_1.
3. E confirms H_1&H_2, by the converse-consequence condition.
4. H_1&H_2 entails H_2.
5. E confirms H_2, by the special-consequence condition.

Since H_2 is arbitrary, any statement is confirmed by evidence that confirms any other hypothesis.

Brody notes that this result depends upon the following entailment (there are other sources of dependency, but this is the only one he feels may reasonably be questioned):

H_1&H_2 entail H_1,

and he suggests that the converse-consequence condition be replaced by the following:

Brody's rule: If E confirms H_1, and H_2 explains H_1, then E confirms H_2.

Brody suggests that this avoids the result embodied in Hempel's argument and retains the intuitively desirable effects of the original converse-consequence condition. Given our account of explanation, we are now able to evaluate this claim.

It is obvious that Brody's rule resolves the paradox involved in Hempel's argument. For even though H_1&H_2 entails

H_1, that conjunction does not explain H_1, since the explanandum has one of the premises as a logical consequence, thus violating one of our conditions for explanations of laws. But to just make this requirement does not resolve all of the difficulties.

First, we ought to consider some remaining issues left open by Brody. He asks: "Should we replace the special-consequence condition by the following? If E confirms H_1, and H_1 explains H_2, then E is evidence for H_2".[29] Then he continues:

> When we discussed previous modifications of our conditions, we did not raise the question of whether we should modify both of the conditions or whether it was sufficient to modify one (since Skyrms' paradox involved only the converse-consequence condition, it would have to be that condition). But had we raised that question, it would have been very unclear as to what to say since we had no account of why we were doing what we were doing. With our new proposed change, however, intuitive considerations might indicate that we should modify only the converse-consequence condition. In the case of consequences of a hypothesis, we are not, after all, giving evidence for an explanation; we are merely pointing out that this evidence for the whole is also evidence for part of the whole. Nevertheless, this matter probably needs further consideration.[30]

For his analysis to be adequate, Brody must modify the special-consequence condition in the way he suggests. Otherwise, difficulties arise, as I now hope to make clear.

---

29. ibid., p. 424.
30. ibid.

Let us go back to the revised converse-consequence condition, and consider an example:

$$\begin{array}{ll} H_2: & (x)(Ax \to Bx) \\ & \underline{(x)(Bx \to Cx)} \\ H_1: & (x)(Ax \to Cx) \end{array}$$

Suppose the premises H_2 explain H_1. We have noted that the conjunction of H_1 and H_2 does not explain H_1. But now, notice that the following also explains H_1:

$$\begin{array}{ll} & (x)(Ax \to Bx) \\ H_3: & (x)(Bx \to Cx) \\ & \underline{(x)(Px \to Qx)} \\ H_1: & (x)(Ax \to Cx), \end{array}$$

where H_3 is like H_2 with the exception that H_3 contains a superfluous premise. This is allowed by our definition of explanation as well as the account given by Brody. We have not required that there not be a proper subset of an explanans which itself will suffice for the explanation of a given explanandum, as Ackermann had required. That is, we have not required that an explanans not contain any superfluous elements.

However, if superfluous elements are allowed, the special-consequence condition will have to be altered as suggested by Brody. For otherwise we can bring the following argument to bear:

1. Suppose E confirms $(x)(Ax \to Cx)$ (i.e., H_1). Then, since H_2 by assumption explains H_1, by the converse-consequence condition E confirms H_2.

2. But H_3 also explains H_1. Thus E confirms H_3, for the same reason as in 1.

3. By the special-consequence condition, H_3 entails $(x)(Px \rightarrow Qx)$, thus E confirms $(x)(Px \rightarrow Qx)$.

Yet $(x)(Px \rightarrow Qx)$, being a superfluous premise in the explanans, is completely arbitrary, with the exception that it must not be a logical consequence of H_1. Therefore, E will confirm almost any sentence, given that E confirms some sentence.

However, if the special-consequence condition is revised as Brody suggests, then the premises of H_3 do not constitute an explanans for $(x)(Px \rightarrow Qx)$, and therefore E does not confirm it. Thus, given our analysis of explanation (and Brody's analysis as well), he must revise the special-consequence condition in the manner which he has suggested, if he wishes to avoid the above result.

One might argue that the source of the difficulty is our definition of explanation, and maintain that an explanans should not contain superfluous elements. This is fair enough, and I don't quite know how to respond to such a challenge. One response could be that it is not only our definition of explanation which countenances superfluous premises. In particular, both Hempel's and Brody's remarks on explanation do not contain such a requirement. However, I realize that this is not a knock-down argument in favor of allowing superfluous elements.

Perhaps a better argument for allowing such premises is this: to be given an explanans which contains more information than is actually required in the explanation of a certain explanandum does not, for this reason alone, require us to consider such an explanation as inadequate. For example, to explain why 'a' conducts electricity, we might offer the following as an explanans: all copper conducts electricity, 'a' is copper, 'a' is heavy, 'a' has such and such an atomic composition, etc. To have at our disposal such "extra" facts does not itself dictate that the above is not a perfectly adequate explanation.

5.6. Evaluation

This section will be devoted to a brief evaluation of the model just developed. However, I want first to continue the discussion of the preceeding section, concerning the desirability of excluding superfluous elements in an explanans. This issue will lead us into a discussion of just what notion is being explicated.

5.6.a. Superfluous premises

In their early attempt to explicate the notion of explanation, Hempel and Oppenheim proposed a set of conditions that could be satisfied by sentences T and C which, however, were not "minimal" sentences. That is, something of the following form

Perhaps a better argument for allowing such premises is this: to be given an explanans which contains more information than is actually required in the explanation of a certain explanandum does not, for this reason alone, require us to consider such an explanation as inadequate. For example, to explain why 'a' conducts electricity, we might offer the following as an explanans: all copper conducts electricity, 'a' is copper, 'a' is heavy, 'a' has such and such an atomic composition, etc. To have at our disposal such "extra" facts does not itself dictate that the above is not a perfectly adequate explanation.

5.6. Evaluation

This section will be devoted to a brief evaluation of the model just developed. However, I want first to continue the discussion of the preceeding section, concerning the desirability of excluding superfluous elements in an explanans. This issue will lead us into a discussion of just what notion is being explicated.

5.6.a. Superfluous premises

In their early attempt to explicate the notion of explanation, Hempel and Oppenheim proposed a set of conditions that could be satisfied by sentences T and C which, however, were not "minimal" sentences. That is, something of the following form

(A) T: (x)(Fx→Gx)
C: Fa&Ba
E: Ga

$$\begin{array}{ll} \text{T:} & (x)(Fx \to Gx) \\ \text{C:} & \underline{\quad Fa \& Ba \quad} \\ \text{E:} & \quad Ga \end{array}$$

satisfies their conditions even though it contains a component which is not essential for the derivation of E. Clearly, the authors were explicating the notion of _an_ explanation; there was no attempt on their part to limit the definition to the smallest set of sentences which would serve as the T and C of their analysis. The idea, I suggest, was to establish conditions definitive of an explanation in general, with other distinctions, such as "more complete" or "most parsimonious" explanation to be made later. That is to say, we could compare two explanations which otherwise met the conditions liad down as to which was simpler or more complete than the other.

On the other hand, Ackermann introduced conditions which were designed in such a fashion to be satisfied only by the smallest set of sentences T (where T is the explanans set). For example, he claimed that the set {Tb, (x)(Mx→Nx), Ma} ought not to be an explanans set for 'Na', since the set contains the superfluous sentence 'Tb'. But he nowhere produces an argument as to why this should be the case, other than remarking that if superfluous premises were countenanced, then "the sheer complexity of the explanans may seemingly destroy relevance".[31]

---

31. Ackermann, R., and Stenner, A., "A Corrected Model of Explanation", _Philosophy of Science_, (33), 1966, p. 169.

Yet this approach, it seems to me, runs into the following difficulty. On the reasonable assumption that there can be more than one explanation of a particular phenomenon, for example

(B) $$\frac{(x)(Cx \rightarrow Ex) \quad Ca}{Ea}$$

(C) $$\frac{(x)(Mx \rightarrow Ex) \quad Ma}{Ea},$$

read respectively as "if all copper conducts electricity and specimen 'a' is copper, then specimen 'a' conducts electricity" and "if all metals conduct electricity and specimen 'a' is a metal, then specimen 'a' conducts electricity", then, although we can say (according to Ackermann) that both (B) and (C) are adequate explanations, we must deny that the following is an acceptable explanation:

(D) $$\frac{(x)(y)((Cx \rightarrow Ex) \& (My \rightarrow Ey)) \quad Ca \& Ma}{Ea}$$

Yet it seems odd to hold that two perfectly acceptable explanations of a particular phenomenon cease to be so when conjoined together.

As an aside, it seems more reasonable to argue the point in a direction opposite to that taken by Ackermann. That is, to argue that subsumption of a particular event under just <u>one</u> law which is part of some theory is <u>not</u> an explanation. Let me explain myself further by means of an example.

Suppose that the following theory is given (i.e., the axioms of a theory), where the T's are theoretical and the O's observational (note: nothing crucial turns on our making this distinction):

Theory T

1. $(x)(T_1x \rightarrow T_2x)$
2. $(x)(O_1x \rightarrow T_1x)$
3. $(x)(T_2x \rightarrow O_2x)$
4. $(x)(T_1x \rightarrow O_3x)$

The empirical generalization $(x)(O_1x \rightarrow O_2x)$ is a logical consequence of the axioms, and hence is part of the theory. Suppose that $(x)(O_1x \rightarrow O_2x)$ is used, along with suitable statements of initial conditions, to explain O_2a. One might well reject this explanation on the following ground. Only theories as a whole are meaningful units. The axioms of a theory "implicitly define" the terms occurring therein, in the sense that they specify relations in which such terms may stand to other terms. Indeed, Carnap and Hempel held (and still may hold) the view that theories are really only "partially interpreted" calculi, in the sense that no interpretation is assigned to theoretical terms, only to observation terms; and that correspondence rules associate certain theoretical constructs with observational ones, the latter, however, not constituting explicit definitions.

On such a view, the theory as a whole must be used in any explanation, if the explanans is required to be true or cognitively significant. But notice, on this view we are led to a position quite different from Ackermann's. He calls any premise "superfluous" if it is not needed for the derivation of a particular explanandum. Yet a premise (i.e., a component of a theory) could be quite inessential in this sense, and yet be quite essential in another sense, i.e., in the sense that such a component is essential to either the truth or the meaning of the theory as a whole, and hence, not eliminable.

I don't want to argue this point any further, since it would lead us into the enormously complex question of theoretical meaning. I would like to note, however, that our analysis is adaptable enough to provide for the possibility of a more adequate account of theory and theory meaning. We could, for example, require that explanation take place by deductive subsumption under theories (in the sense of "theory" alluded to above, not in Hempel's sense), rather than simply subsumption under laws, which, normally, will only be components of theory. In any event, it is certainly not obvious that Ackermann's exclusion of premises in an explanans which are inessential to the derivation of the explanandum is a completely unobjectionable requirement.

5.6.b. Counter-examples to other models

It is not necessary to show again that all the counter-examples brought forth against previous models fail to count as counter-examples to Professor Kaplan's proposal. In fact, it was by utilizing his proposal that we found counter-examples to other models, in particular to the model of Charles Morgan.

My chief objective in this chapter has been to emphasize that different models are required in providing explanations for particular events and for general regularities. Professor Kaplan's analysis is free from the difficulties which plagued the Hempel and Oppenheim model as well as the defects of more recent proposals. Whether the conditions he provides are indeed sufficient for explanations of singular events remains an open question. It is this author's opinion that they require supplementation. But the main point is this: Kaplan's cogent proposal provides a coherent framework within which further reasearch may be carried on. All the other proposals simply muddied the water and confused the issues.

In regard to our discussion of explanation relative to general regularities, we have essentially concentrated on repairing Popper's account of generality. Explanation of laws, we feel, requires a better account of "theory" than has so far been provided. Yet we do believe that the conditions that we have argued for at least call attention to certain argument forms which clearly ought to be excluded by any, yet to be advanced, model.

Chapter VI

Rationally-Acceptable Explanations

In Chapter I, we remarked that Hempel's early attempt to explicate the notion of explanation was concerned solely with the logical structure of adequate explanations in science. In this chapter, we spell out some of the consequences of such an approach (and some accompanying criticisms of these consequences), determine the general character and intent of the model, and finally, we offer some suggestions for a slightly liberalized version of the Deductive-Nomological model in the light of these criticisms.

6.1. Syntactic, semantic, and epistemic conditions

In the early paper written with Oppenheim, Hempel remarks that in the analysis of the logical structure of explanatory arguments, the requirement that the explanans must be true may be disregarded.[1] Thus, his definition of *potential* explanation was stated in terms of the following syntactical conditions:

(1) T is essentially generalized and C is singular.

(2) E is derivable in L from T and C.

1. Hempel, C., and Oppenheim, P., "Studies in the Logic of Explanation", in *Aspects of Scientific Explanation*, p. 249, footnote 3.

(3) T is compatible with at least one class of basic sentences which has C but not E as a consequence.[2]

The definition of explanation proper (or correct explanation, as Hempel calls it) adds the following semantic condition:

(4) T is a theory and C is true.

Finally, Kaplan showed that requirement (3) had to be replaced by the following semantic condition:

(3') T is compatible with at least one class of true basic sentences which has C but not E as a consequence.

In this original version of the D-N model, there are no epistemic conditions. That is, it is not required that the explanans be known, be confirmed, or be (justifiably) believed. Although we later introduce a model which does carry the requirement that an explanans need only be justifiably believed, thus introducing an epistemic condition, we still believe that it is essential to keep these models distinct from true explanation. It is our opinion that certain confusions in the literature derive from a conflation of these models

To be more precise, let us say that "epistemic condition

---

2. ibid., p. 278.

for explanation" means any condition which requires of an explanatory model that it be relativized to some class K of statements representing a particular knowledge situation.

One more preliminary. Our use of "knowledge" is best construed as "rational belief". For when we speak of a model of explanation which is relativized to a class of sentences K representing a body of scientific knowledge at a given stage, we do not mean to imply that such sentences are true.[3]

To summarize, the original D-N model is not relativized to any epistemic condition, to what we are justified in believing. That this is indeed Hempel's view is made clear by the following passage from his most recent publication on explanation ("Aspects of Scientific Explanation"), in which he discusses some differences between the inductive-statistical model and the D-N model. He concludes this long essay with the remark: "I would like to stress here once more that there are profound logical differences between those two modes of explanation.... Another difference, which so far does not seem to have received attention, lies in what I called the epistemic relativity of

---

3. See Richard Jeffrey, "Probable Knowledge", in Lakatos, The Problem of Inductive Logic, North-Holland, 1968, for an attempt to make the concept of belief do the work that philosophers have generally assigned to the concept of knowledge.

probabilistic explanation, i.e., the fact that we can significantly speak of a probabilistic explanation, even a potential one, only relative to some class K of statements representing a particular knowledge situation. The concept of deductive-nomological explanation requires no such relativization."[4]

Before turning to an elaboration of some of the consequences of such a view, we should note that Hempel's position on the epistemological relativity of D-N explanation is by no means unambiguous. For example, in a review of the book containing the essay from which we have taken the above citation, Prof. Howard Smokler has remarked: "A clear point of view about science emerges from Hempel's work. Science is an activity whose primary function is the explanation and prediction of phenomena on the basis of beliefs which have already been rationally accepted."[5]

Now this remark is true if we understand "explanation" in this context to refer to inductive-statistical explanation, and it is perplexing if we understand it to refer to deductive-nomological explanation. It is perplexing because we have just noted that the original D-N model is _not_ relativized to what it is _rational to believe_, as

---

4. _Aspects_, p. 488.
5. _Synthese_, (16) 1966, pp. 110-122.

Smokler asserts. Indeed, we have just quoted (see footnote 4) from the closing pages of Hempel's most recent contribution to the analysis of D-N explanation stating this fact. Yet between the opening and concluding pages of this long essay, Hempel <u>does</u> introduce a sketch of a model which <u>is</u> relativized to a body of rational beliefs, so we must not assume that Smokler is misrepresenting the intent of the D-N model. Our solution to this difficulty will be to show that the original D-N model has been amplified in certain respects.

The perplexing situation above will be remedied in the remaining sections of this chapter. I want now to consider some criticisms of the "epistemologically-free" version of the D-N model adumbrated above. In so doing, we hope to provide a rationale for moving to a deductive model of explanation which of necessity involves a relativization to our body of accepted beliefs.

In a recent book, Gerald Radnitzky has criticized the "logical empiricist" school, among whose members he includes Prof. Hempel. He is intent on arguing for a view of science which concerns itself with the production and growth of scientific knowledge, as opposed to a concern with the finished products of scientific activity. Some of his criticisms will be important to the later development, thus we turn to these now.

Put briefly, Radnitzky argues that the logical empiricist's global program for the philosophy of science is to articulate an ideal of science, and that its various key themes, such as unified science, empirical significance, confirmation, and explanation, may be looked upon as spelling out various features of this ideal.[6] The ideal of knowledge of this school is pictured as comprising a deductive system, expressed in an artificial and to some extent idealized language. In particular, he views their theory of explanation as the "spelling out of what explanations in Ideal-Unified-Science look like."[7]

One result of such a view of science, Radnitzky feels, is that workers in the field become more intent on working out problems internal to the models themselves, with the question of the applicability of the model being shoved into the background. Carnap, for one, distinguished between inductive logic proper and the 'methodology of induction', and explicitly restricted most of his efforts to the former area. As Radnitzky remarks: "Logical empiricists do not even appear to care whether their results are relevant for the active researcher or not."[8]

---

6. Radnitzky, G., Contemporary Schools of Metascience, 1970, p. xvi.
7. ibid.
8. ibid., p. x.

Yet his main difference of opinion is with the logical empiricist's concentrating exclusively on the language in which the finished products of science are couched, or in which they think such products ought to be presented. Their goal is to make the language more precise and perspicuous, and perhaps to propose improved languages, and also to develop criteria for appraising the finished products with respect to formal adequacy. Examples may be provided, "but they will be of a sketchy variety, often taken from everyday physics rather than from science proper... They will serve to illustrate a thesis rather than provide material for testing a hypothesis about, say, explanation in physics."[9]

In sum, Radnitzky is critical of the fact that the models proposed by people such as Hempel establish conditions which, however, are so ideal as to be incapable of applying to any actual examples of explanations taken from the history of science. He does recognize a certain liberalization of the model, namely, the move to inductive-statistical explanation and another type to be distinguished in subsection 6.3. In the next section, we wish to evaluate the general intent and character of the original D-N model from Hempel's point of view.

---

9. ibid., p. x.

6.2. General character and intent of the model

In the essay "Aspects of Scientific Explanation", Hempel concludes the work by remarking that the D-N model is not to be construed as simply descriptive of the explanations offered in empirical science. For one reason, there does not appear to be any generally accepted presystematic understanding as to what ought to count as a scientific explanation. His construal is intended as an *explication*, an attempt to replace a familiar but vague notion by a more precisely characterized and illuminating one.[10]

Thus, it is no surprise that Hempel does not undertake "case-studies" of actually proffered explanations (for lack of which Radnitzky criticizes him). Unless we arrived at some understanding of what an adequate explanation should look like, i.e., unless we had a theory of explanation, we would be incapable of drawing any conclusions at all concerning the adequacy of proffered explanations. Without a model, we would have no criterion for evaluation.

Besides, even if there were no explanations offered which fit precisely the requirements of the model, that might just show that such examples are defective, or that

---

10. Hempel, *Aspects*, p. 489.

obvious premises were suppressed, and once made explicit, could be brought more in line with the proposed model. Thus, I find no fault with Hempels' failing to generate the model by abstracting its features from already proffered explanations.

In the same essay, Hempel explicitly states that his models are not meant to describe how working scientists actually formulate their explanatory accounts.[11] He compares the concepts of explanation which he advances to the concept of mathematical proof as construed in metamathematics. Both the metamathematical model of proof and the models of explanation are selective. For example, metamathematical proof theory is concerned only with the notion of proof in mathematics;[12] and this is not to say that every sense of 'proof' must conform to this model, e.g., as it is used in "the proof of the pudding is in the eating".

By the same token, the models of explanation do not purport to reflect the sense of 'explain' when we speak of explaining the rules of a hockey game or of a contest. Hempel's interest is in providing a model of explanation which involves explaining natural phenomena by subsuming them under general laws, showing that they were indeed to

---

11. ibid., p. 412.
12. ibid.

be expected, and that they fit into a nomic nexus.

Besides the fact that both models are selective, Hempel remarks that neither purports to be descriptive. Thus, "the formulations that mathematicians actually offer will usually depart to some extent from that called for by rigorous, and, as it were, 'ideal' metamathematical standards. Yet those standards may be said to exhibit the logical structure and the rationale of mathematical demonstration and to provide criteria for the critical appraisal of particular proofs that might be proposed."[13] Thus, if a proposed proof differs from a theoretical standard only in minor ways, for example, by assuming obvious intermediate steps, we might say that the proof is eliptically stated. In other cases, such ommissions may be crucial.

Most importantly, in addition to providing standards for appraisal, Hempel notes that the development of a rigorous proof theory led to other "far-reaching and unexpected results concerning provability, decidability, and definability in mathematical systems of specified kinds."[14]

Now Hempel feels that "analytic models of scientific explanation ... can serve similar purposes, if on a much

---

13. ibid., p. 414.
14. ibid.

more modest scale".[15] He mentions in this connection the possibility of utilizing the results established by Ramsey for the possible dispensability of theoretical terms in the context of scientific explanation. The main point is, however, that such results could not even be possible without reference to a precisely formulated conception of scientific explanation.

So far in our sketch of Hempel's statement of the general intent of the models he offers we notice that almost every point of criticism advanced by Radnitzky is expressly admitted by Hempel. His attempt is not one designed to contribute to the growth of scientific knowledge, but rather it is an attempt to provide a sufficiently precise account of a notion that may lead to certain theoretical results, e.g., the above mentioned possibility of utilizing the results provided by Ramsey in the context of explanation. For Hempel clearly intends his models to be ideal to a certain extent. And I think that he deserves credit for this undertaking, as he (at least) initiated research aimed at providing a precise explicatum for explanation. This task, I think, has been partially completed by Kaplan.

On the other hand, one point Radnitzky makes still lingers in the background, demanding a response. Namely,

---

15. ibid.

the question of the applicability of the model as a tool for appraising actually proffered explanations. For example, we have above noted Hempel's claim that the models are not intended to be descriptive, but rather are intended to provide standards of critical appraisal. Yet it seems that very few, if any, examples can actually meet the conditions of the model. Here I have in mind the requirement that the explanans be _true_. And notice that Hempel does not mean by this _accepted as_ true or (justifiably) _believed_ to be true.

Yet if we adopt, as I think we should, the view that scientific activity provides us with a body of knowledge (i.e., a body of rational beliefs), however tentative and subject to revision, it seems natural to ask how the D-N model fares with respect to an explanans which need be only highly confirmed or _accepted_ (in a sense to be clarified shortly). Indeed, several authors, notably Scheffler and Radnitzky, seem to think that Hempel at times wants the truth requirement and at other times _replaces_ it by the requirement of high confirmation. Such is not the case, as will be argued in the following section. However, Hempel has considered liberalizing the original model in such fashion, and I will turn to this matter directly.

First, we note that the D-N model as proposed by Kaplan need not remain unaltered if we make the requirement that the explanans be only <u>accepted</u> as true. This might seem surprising- but once we introduce such epistemological considerations, we find that certain alterations must be made (again, in a manner to be discussed shortly).

Hempel presumably intended his model to serve in the appraisal of proffered explanations in some such way as the following. First, test it to see if it meets the requirements for a potential explanation. If it does, test it to see if its explanans are highly confirmed. Of course, an empirical generalization, in an infinite domain, can never definitely be <u>known</u> to be true. Therefore, in this situation, we can speak of an explanation as probably adequate, or justifiable, or that the explanans are true so far as we have evidence for believing. In short, an explanation will be appraised as adequate if its explanans contains laws and sentences of initial conditions which are acceptable as worthy of our belief.

But if this is how we are to appraise actual examples, it seems natural just to introduce a condition to the effect that the explanans be rationally <u>accepted</u> as true. Hempel did in fact suggest something like this. We turn now to consider this view, and we discuss the notion of "acceptance" alluded to above at the end of the next section.

6.3. <u>Rationally-acceptable explanations</u>

In his 1962 essay "Deductive-Nomological vs Statistical

Explanation", which antedates "Aspects of Scientific Explanation" but postdates "Studies in the Logic of Explanation", Hempel liberalizes the original D-N model in two ways. (1) He relaxes the deducibility requirement and introduces the notion of an inductive-statistical explanation, in which the explanandum need not follow deductively from the explanans, but only with a certain degree of probability. (2) He also in this paper recognizes D-N explanations in which the explanans need not be true, but rather only well-confirmed. Thus, he distinguishes between the following:

(i) potential explanation

(ii) explanation with highly confirmed explanans

(iii) explanations with true explanans.[16]

Now with the introduction of type (ii) explanations, we must reconsider our quotation above concerning the epistemic relativity of I-S (but not D-N) explanations. What Hempel says there must be qualified in light of what he says earlier on in the same essay. After discussing the epistemic relativity of I-S explanation, he remarks: "... this danger never arises for deductive explanations. Hence, these are not subject to any such restrictive condition, and the notion of a potential deductive explanation (as

---

16. Hempel, C., "Deductive-Nomological vs Statistical Explanation", in Minnesota Studies in the Philosophy of Science, vol. III, pp. 98-169.

contradistinguished from a deductive explanation with well-confirmed explanans) requires no relativization with respect to K. As a consequence, we can significantly speak of true D-N ... explanations: they are those potential D-N explanations ... whose premsies (and hence also conclusions) are true- no matter whether this happens to be known or believed, and thus no matter whether the premises are included in K."[17] (my emphasis)

This indicates that we now have three distinct models. Besides the original D-N models containing only syntactic and semantic conditions, we now have the basis for a third model which of necessity includes epistemic conditions.

Furthermore, type (ii) explanation seems important for the following reason: it makes the D-N model applicable to actually proffered explanations. That is, we now have a sketch of a model which countenances explanations by appealing to what we are rationally justified in believing, or to what is confirmed. For this reason, I call type (ii) explanations "rationally acceptable", or I speak of them as satisfying the "rationally acceptable model of scientific explanation".

We note that Prof. Smokler's comment (cited above) seems best understood as referring to this model, although it is doubtful that Hempel intends this to be the only or

---

17. Hempel, Aspects, p. 402.

the primary model. We note also that there is no question of _replacing_ the truth requirement by that of high confirmation; that is, there are three distinct models, only one of which requires explicit relativization to our body of scientific beliefs. What is true is the fact that type (iii) explanations have been Hempel's major concern.

But it does seem desirable to exploit type (ii) explanations. This model allows us to give explanations in terms of our body of scientific knowledge, which, of course, need not be infallible. In fact, what is acceptable as an explanation at one time, relative to what we are justified in believing at that time, may not be acceptable as an explanation at another time. And this seems perfectly in accord with the view of scientific knowledge as growing and tentative, changing with new discoveries.

However, the major advantage of type (ii) explanations is that the D-N model is now applicable to actual examples, rather than simply being concerned with the products of science, i.e., finished systems.

But what more can be said about the model itself? Hempel does not have much to say in regard to its specific character, but he does make the following remark:

> ... we say that a given potential explanation is more or less highly confirmed by a given body of evidence according as its explanans is more or less highly confirmed by the evidence

> in question. If the explanation is formulated in a formalized language for which an adequate quantitative concept of degree of confirmation or of inductive probability is available, we might _identify the probability of the explanation relative to e with the probability of the explanans relative to e_[18] (my emphasis)

This description of type (ii) explanations calls for some comment. To be more specific, let 'i' be the statements of initial conditions, 'a' the explanandum sentence, 'h' a general hypothesis, and 'e' evidence statements relevant to 'h'. In addition, let 'a' be a deductive consequence of 'h' and 'i', thus

(A) h

$\underline{\text{i}}$

a .

Since we are considering the type (ii) model, we say that the explanans are well-confirmed, rather than true. Hempel counsels us to identify the probability of the explanation (relative to e) with the probability of the explanans (relative to e). Let us suppose the statements of initial conditions have probability 1; then the probability of the explanans reduces to P(h,e).

Now does Hempel intend the following to be the form of type (ii) explanations?

(B) P(h,e)= r

$\underline{\text{i}\qquad}$

a

18. _Minnesota III_, p. 103.

(Presumably, we are to say that this explanans 'probably' (to degree r) explains 'a'.)

This interpretation seems unlikely, for the following reason. In schema (B), it is not the case that 'a' is a logical consequence of the given explanans. Thus, Hempel could adopt the above schema only by giving up the deducibility requirement.

The point here is this. If we can not detach hypotheses from their evidence statements, then it seems like we are committed to including degree of confirmation statements among our explanans (in the type (ii) case), given Hempel's statement quoted above. Yet this commits us to schema (B) and the surrendering of the deducibility requirement. And this result seems very unsatisfactory.

But perhaps we can understand Hempel's remarks concerning the probability of a given explanans in another way. Namely, high probability is to be taken as a criterion for admission to our body of justifiably accepted scientific beliefs. To be more specific, let us consider Hempel's account of what he calls the "accepted-information model of scientific knowledge".

Let K_t (or simply, K) be the set of our total body of scientific knowledge at a given time t, i.e., a set all of whose elements are sentences accepted as true by the scientific community at t. Membership in K will of course change from time to time, with new members being admitted and old, now disconfirmed ones being rejected. Membership in K is either by _direct_ acceptance or _inferential_ acceptance, the latter being either deductive or inductive inference. Hempel then remarks:

> This schematic model does not require, then, that the statements representing scientific knowledge at a given time be true; rather, it construes scientific knowledge as the totality of beliefs that are accepted at a given time as warranted by appropriate scientific procedures. I will refer to this model as the _accepted-information model of scientific knowledge_.[19]

Type (ii) explanations as construed in this context do not face the same defects as pointed out in connection with example (B), above. Assuming for the moment the existence of some suitable acceptance rule (the issue of acceptance to be discussed shortly), we can detach hypotheses from their evidence statements and accept them as members of our body of scientific knowledge K, to be used for any purpose that we see fit. Thus, an accepted-information schema of type (ii) explanations will be of the following form:

(C) h

$$\frac{i}{a}$$

(Here, 'a' _is_ a deductive consequence of the explanans.)

---

19. ibid., p. 150.

In this case, explanations are relativized to our body of reasonable beliefs, rather than to evidence e. And we need no longer attach probabilities to our accepted sentences- for they are accepted outright.

Thus, we suggest that this latter account of type (ii) explanations is more reasonable than the first account. Of course, membership in K may well be controlled partially by a requirement of high probability, but this is not the same as asserting that an explanans must contain degree of confirmation statements.

Yet Hempel never develops the type (ii) model any further. In his most recent essay, as already noted, he still distinguishes between the three types of deductive-nomological explanation- potential, highly confirmed (i.e., acceptable), and true- but any detailed discussion of the accepted-information model, or the highly confirmed version, is absent. Instead, Hempel simply ends the essay with the remark we have quoted at the beginning of this chapter, i.e., by affirming that D-N explanation does not require relativization to a class K of statements representing a particular knowledge situation.

It is difficult to understand exactly why Hempel drops what I have called the "rationally-acceptable" model. I suspect that the difficulties that he ran into in his attempt to formulate a particular rule of acceptance had something to do with it. But clearly, it may be profitable to examine the structure of such a model independently of any particular rule of acceptance. This we do in the following section.

First, however, we must elaborate on our use of "acceptable". Recently, much work in inductive logic has been devoted to establishing a rule of acceptance which would allow for the detachment of inductive conclusions from particular premises. Such an inductive acceptance rule might read as follows, where 'c(h,e)' denotes the confirmation of h, given e:

> Accept or reject h, given e, according as c(h,e) is greater than 1/2 or c(h,e) is less than 1/2; when c(h,e) is equal to 1/2, h may be accepted, rejected, or left in suspense.

(We do not mean to endorse this particular rule, rather we introduce it here only for illustrative purposes.)

The plausibility and desirability of such rules has not been generally agreed upon. For example, Carnap has long rejected the use of such rules in inductive logic. He has argued that the main task of inductive logic concerns the determination of a suitable confirmation function which assigns certain numerical values to (singular) hypotheses on the basis of evidence. Such numerical values can be used for determining rational bets, but can <u>not</u> be used as a basis for accepting, even tentatively, any hypothesis.

On the other hand, Prof. H. Kyburg has argued very forcefully for the use of such rules in inductive logic.[20] He points out that any inductive logician of Carnapian stamp is committed at least to an inductive acceptance rule for evidence statements. For example, in the rule stated above, we must have accepted 'e' prior to our computation of 'c(h,e)'.

Moreover, Wesley Salmon has argued that without inductive acceptance rules covering-law theorists of explanation are committed to schema (B) above, which as we have already noted, has the consequence that the deducibility requirement can not be met. Thus we must conclude, that without an acceptance rule, we are unable to appraise actually proffered explanations in science, since, as we have also noted earlier, the truth requirement can not be met by actual explanatory examples.

Assuming the existence of a suitable acceptance rule,[21] we turn now to examine the structure of rationally-acceptable explanations.

---

20. Kyburg, H., "The Rule of Detachment in Inductive Logic", in Lakatos, _The Problem of Inductive Logic_.

21. Different rules are possible. For recent contributions to this area see Levi, I., _Gambling with Truth_, Hintikka, J., "Induction by Enumeration and Induction by Elimination", in Lakatos, _The Problem of Inductive Logic_, Hilpinen, R., _Rules of Acceptance and Inductive Logic_, and the paper by Kyburg in the above footnote.

6.4. The structure of rationally-acceptable explanations

Preliminary to our discussion of the model, we undertake a reconsideration of Kim's objection to Kaplan's proposal which we think will illuminate certain conditions that we will eventually impose on the accepted-information model. We will also better understand why Kaplan's proposal, though unharmed by Kim's counter-example, needs to be altered slightly in the present context.

Kim's objection can perhaps best be understood as an objection to the accepted-information model, or to what we have called type (ii) explanations; of course, Kaplan's proposal is of type (iii), and thus the objection is not well founded. That is to say, Kim's objection rests on the fact that he confuses type (ii) and type (iii) explanations.

For consider his objection once again. He produces the following example,

$$\begin{array}{ll} T: & (x)(Fx \vee Gx \rightarrow Jx) \\ C: & \underline{\quad Fb \vee Gb \quad} \\ E: & \quad Jb \end{array}$$

and comments as follows (my paraphrase). Suppose that the explanans of this example are true. Then by Kaplan's definition (i.e., the version where C is required to be a conjunction of true basic sentences), 'Jb' is directly S-explainable by the given law, since, Kim remarks, we

<u>know</u> that either 'Fb' or 'Gb' is true (note that it is not a matter of what is <u>known</u> at all). But, Kim continues, suppose that we have verified the disjunction 'FbvGb' without verifying either disjunct (for example, by the use of some law in conjunction with some specific information). Then the conjunction required by Kaplan's proposal can not be produced, since we do not <u>know</u> which disjunct is true, thereby ruling out the above example. And this example is surely unobjectionable; thus, Kim infers, Kaplan's proposal is inadequate. So goes the objection.

In a footnote, Kim adds the parenthetical remark to the effect that 'FbvGb' might be ascertained (does he mean accepted?) on the basis of statistical information. But clearly, this objection involves the confusion we have already mentioned above between two types of explanation. We grant, with reference to the accepted-information model, that 'FbvGb' may indeed be acceptable while neither 'Fb' nor 'Gb' alone is acceptable. But Kaplan's model is intended as a revision of the original Hempel and Oppenheim model, i.e., a type (iii) model which is not in any way relative to what is known, confirmed, or accepted. Thus, we dismiss Kim's objection as irrelevant to Kaplan's proposal.

I suggest that Kim was confused by the shift from the truth requirement to the requirement that the explanans be highly confirmed. If this move is thought of as a replacement, then Kim's objection would be well taken. Of course, no such replacement was intended. But Kim's objection does show that the accepted-information model will not be directly parallel to Kaplan's model. We turn now to examine the structure of a model of rationally acceptable explanations (or, alternatively, of an accepted-information model).

Supposing again the existence of a suitable acceptance rule, suppose that we were to construe an accepted-information model along the following lines:

(T,C) is a direct S-acceptable explanans for the (already accepted as true) singular sentence E if and only if

1) T is a theory which is rationally acceptable.
2) C is a conjunction of singular sentences, each rationally acceptable as true.
3) E is a disjunction of basic sentences.
4) $\{T,C\} \vdash E$, and
5) not $\{C\} \vdash E$.

This proposal is an exact parallel to Kaplan's proposal. Will it suffice as an explication of "rationally acceptable explanation"? Kim's remarks show us that it will not.

To elaborate somewhat more on this point, let us note a few parallels between the concept of acceptability and the concept of truth. The following condition of truth,

(CT1) h_1 is true and h_2 is true iff $h_1 \& h_2$ is true

corresponds to the following condition of acceptability:

(CA1) $Ac(h_1/e)$ and $Ac(h_2/e)$ iff $Ac(h_1 \& h_2/e)$.

(CA1) is not a condition restricted to the concept of inductive acceptability; for example, in Hintikka's system in <u>Knowledge and Belief</u>, the following is a logical truth:

Ba(p) and Ba(q) iff Ba(p&q).

Yet we must also note certain differences; for example, the following rule of truth,

(CT2) $h_1 v h_2$ is true iff h_1 is true or h_2 is true

has no parallel in terms of acceptability. That is, the following condition is considered to be too strong a condition to impose on the concept of acceptability:

(CA2) $Ac(h_1 v h_2/e)$ iff $Ac(h_1/e)$ or $Ac(h_2/e)$.

(We note that the following formula is not valid in Hintikka's system:

Ba(pvq) iff Ba(p) or Ba(q).)

In these terms, we might say that while Kaplan's proposal depends upon (CT2) for its adequacy, Kim's objection is based upon a confusion between (CT2) and (CA2).

Now in the above definition of direct S-acceptable explanation we have required that C be a conjunction of sentences accepted as true. Kim's example will not meet that requirement, and thus it will be ruled out. That is, so long as we were working with the truth of the explanans, we were guaranteed by condition (CT2) (and also by the fact that the language L is assumed to have been given an interpretation) that either 'Fb' or 'Gb' of Kim's example is true, and thus that that example could be shown to conform to Kaplan's account. But we can not make a similar claim with respect to acceptability. Thus Kaplan's account must be revised in this context.

But now let us recall the fact that Kaplan has shown us three ways of constructing a model of type (iii) explanations. The main features of these three were:

(1) we could require there to be a verifying class K of true basic sentences.

(2) we could require that no disjunct of the disjunctive normal form of C have E as a logical consequence, or

(3) we could require that C be a conjunction of true basic sentences.

Now Kaplan chose (3), since it is somewhat simpler than the other versions. But we should note that Kaplan's proof showing that the three sets of conditions are equivalent will no longer hold when the shift is made to explanans which need only be acceptable as true. In fact, both (1) and (3) are discredited in this regard by Kim's counterexample. Yet something like one of these conditions is absolutely essential to a "rationally acceptable" model. Without such a condition, the model will have all the flaws of the original Hempel and Oppenheim model.

The obvious move at this point would be to adopt formulation (2), since it avoids Kim's counter-example as well as all th counter-examples which could be produced similar to those brought forth against the earlier models. Thus, I propose the following as a model of rationally-acceptable explanation:

(K1') The ordered couple (T,C) of sentences is a direct S-acceptable explanans for the singular sentence E (accepted as true) if and only if the following conditions are satisfied:

1) T is a theory, accepted as true.
2) C is singular and accepted as true.
3) E is not logically derivable from any disjunct of the disjunctive normal form of C.
4) there is a class of basic sentences K such that $K \vdash C$ but neither $K \vdash E$ nor $K \vdash \sim T$.
5) $\{T,C\} \vdash E$
6) E is a disjunction of basic sentences.

(K2') (T,C) is an <u>S-acceptable</u> explanans for the singular sentence E (accepted as true) if and only if the following are satisfied:

1) (T,C) is a direct S-acceptable explanans for each of a set of singular sentences $\{E_1,\ldots,E_n\}$, and
2) $\{E_1,\ldots,E_n\} \vdash E$.

We are now in a position to evaluate explanations offered on the basis of what we are rationally justified in believing. It remains, of course, to study how the model changes, if it does at all, according to the specific acceptance rules that are adopted as a criterion of membership for our body of scientific knowledge K. There is much activity in this area at present, and many old, established principles are being challenged (e.g., whether the class K ought to be closed under the consequence relation). Hintikka, Kyburg, Hempel and Levi (among others) have produced various acceptance rules and logics of rational belief, which often differ in various respects. Our model of rationally-acceptable explanations is neutral with respect to the specific rules which one adopts for admission into K. Of course, further work is needed in this regard.

6.5. Discussion

We conclude with a brief comparison of what we have called type (ii) and type (iii) explanations. Type (ii) explanation requires that the explanans statements be rationally-acceptable, whereas type (iii) explanations are those in which the explanans must be true- independently of whether the explanans statements are known, believed, or thought of. The latter account provides us with a theory of explanation in an idealized sense, being relative to what is the case, not to what is known or believed to be the case. The advantage of clarifying such a model is that it enables us to arrive at some conclusions regarding the logical structure of explanation. For example, in Hempel's early discussion of type (iii) explanation, he was led to introduce the notion of a class K of potentially verifying basic sentences from which C was derivable but from which E was not derivable. Such a condition is shown by the trivialization theorems of Eberle, Kaplan and Montague to be inadequate. Kaplan was later to improve upon this version, requiring that the class K of verifying sentences be an actual class, i.e., by requiring that K contain only true basic sentences.

Although the type (iii) model provides us with a theory of explanation relative to what is the case, it is not directly applicable to actually proffered examples of explanation whose explanans are only believed to be true. Clearly, the truth requirement is too strong a condition to impose upon our attempts to explain singular statements in terms of a body of knowledge which we accept provisionally at a certain stage of inquiry. Indeed, it was in order to provide for such considerations that Hempel was led to introduce the type (ii) model.

In his early paper written with Oppenheim, Hempel was led to require that the explanans statements be true by the following consideration. Suppose that we required only that the explanans be highly confirmed by all the relevant available evidence. Then it might happen that the explanans of an explanatory argument was well confirmed at an early stage of scientific research, but strongly disconfirmed by the evidence available at a later time. In this case, we would have to say that the explanandum was correctly explained by the given argument at the earlier stage, but not at the later one. And this, Hempel remarked, "seemed counterintuitive, for comman usage appeared to construe the correctness of a given explanation as no more time dependent than, say, the truth of a given statement".[22]

---

22. Hempel, C., "Deductive-Nomological vs. Statistical Explanation", in Minnesota Studies in the Philosophy of Science, III, p. 102.

In a later article, Hempel came to amend the above line of reasoning, arguing that the notion of correctness can be understood in two different ways, "both of which are of interest and importance for the logical analysis of science: namely, as truth in the semantical sense, which is independent of any reference to time or to evidence; or as confirmation by the available relevant evidence- a concept which is clearly time dependent".[23] In this way, Hempel was led to distinguish between type (ii) and type (iii) explanations.

Above we have considered type (ii) explanations in the context of Hempel's "accepted-information model of scientific knowledge". Hempel himself did not construe it in this fashion but rather in somewhat the following way. Recall again the conditions imposed on schema (B) in section 6.3. Hempel could also be interpreted as intending the following construal of this schema:

An argument of the form $\frac{\begin{matrix} h \\ i \end{matrix}}{a}$

is correct to degree r if and only if

p(h&i/e) if greater than or equal to r

(and the other conditions are met).

We argue that such a construal still requires an acceptance

---

23. ibid., p. 102.

rule, for the following reason. Surely not <u>any</u> value of 'r' for the explanans is sufficient to render 'a' explainable on the basis of that explanans. Clearly what is intended is that 'r' be sufficiently high, i.e., that 'r' be high enough to justify our asserting the explanans, in order to render such an argument explanatory. In other words, that 'r' be high enough to justify the <u>acceptance</u> of the explanans. Notice that this is precisely what we have argued for above, in connection with the accepted-information model. An acceptance rule governs the admission of a sentence S into our body of knowledge K. Once admitted to K, S may then serve in the explanans of type (ii) explanations.

Hempel envisaged a unified treatment of explanation types (i)-(iii) in the following way. He wrote:

> First, we define a potential explanation (of deductive-nomological form) as an argument ... which meets all the requirements indicated earlier, except that the statements forming its explanans and explanandum need not be true....
> Next, we say that a given potential explanation is more or less highly confirmed by a given body of evidence according as its explanans is more or less highly confirmed by the evidence in question....
> Finally, by a true explanation we understand a potential explanation with true explanans- and hence also with true explanandum.[24]

Now we show how Kaplan's analysis leads to a unified

24. ibid., pp. 102-103.

treatment of these three notions. First, we repeat that the version of Kaplan's proposal for type (iii) explanation requiring that C be a conjunction of true basic sentences does not remain the desired formulation in the context of type (ii) explanation. But the interesting point to note is that Kaplan has provided us with other ways of making the desired requirement, and that one of his other formulations does provide just such a unified account. That account is as follows. Where K is our body of accepted-information, as before, we say:

> The ordered couple of sentences (T,C) is a potential explanans for the singular sentence E if and only if (T,C) is a potential direct S-explanans for E or (T,C) is a potential S-explanans for E.
>
> (T,C) is a potential direct S-explanans for the singular sentence E if and only if
> 1) T is essentially generalized
> 2) C is singular
> 3) E is not derivable from any disjunct of the disjunctive normal form of C
> 4) $\{T,C\} \vdash E$
> 5) There is a class K of basic sentences such that C is derivable from K, but neither E nor the negation of T is derivable from K.
> 6) E is a disjunction of basic sentences.

(T,C) is a potential S-explanans for the singular sentence E if and only if

1) (T,C) is a potential S-explanans for each of a set of singular sentences $\{E_1,\ldots,E_n\}$, and
2) $\{E_1,\ldots,E_n\} \vdash E$.

Next, we say that

(T,C) is a rationally acceptable explanans for the singular sentence E if and only if

1) (T,C) is a potential explanans for E
2) T is a member of K
3) C is a member of K.

Finally,

(T,C) is a true explanans for the singular sentence E if and only if

1) (T,C) is a potential explanans for E
2) T is a theory
3) C is true.

Although we display here only the logical conditions of explanation, it should be clear that rationally-acceptable explanations involve reference to an epistemological condition of acceptance. The advantage of having such a model should be clear: it allows us to appraise explanations as to their formal adequacy in those cases where the explanans is worthy of our belief, though which we may be willing to reject at some later time.

BIBLIOGRAPHY

Ackermann, R. "Deductive Scientific Explanation." Philosophy of Science, 32 (1965), pp. 155-167.

Ackermann, R., and A. Stenner. "A Corrected Model of Explanation". Philosophy of Science, 33 (1966), pp. 168-171.

Aronson, J. "Explanations without Laws." The Journal of Philosophy, 66 (1969), pp. 541-557.

Braithewaite, R.B. Scientific Explanation Cambridge Press, 1955.

Brody, B. "Towards an Aristotelian Theory of Scientific Explanation." Philosophy of Science, 39 (1972), pp. 20-31.

Brody, B., ed. Readings in the Philosophy of Science. Englewood Cliffs: Prentice-Hall, 1970.

Carnap, R. "On Rules of Acceptance". I. Lakatos, ed. The Problem of Inductive Logic. Amsterdam: North-Holland Publishing Co., 1968, pp. 146-150.

Eberle, R., Kaplan, D., and Montague, R. "Hempel and Oppenheim on Explanation." Philosophy of Science, 28 (1961), pp. 418-428.

Hempel, C., and Oppenheim, P. "Studies in the Logic of Explanation." Philosophy of Science, 15 (1948) pp. 135-175. (Reprinted in C. Hempel, Aspects of Scientific Explanation. New York: The Free Press, 1965.)

Hempel, C. "Aspects of Scientific Explanation". C. Hempel. Aspects of Scientific Explanation. New York: The Free Press, 1965, pp.

Hempel, C. Aspects of Scientific Explanation. New York: The Free Press, 1965.

Hempel, C. "Deductive-Nomological vs Statistical Explanation." H. Feigl and G. Maxwell, eds. Minnesota Studies in the Philosophy of Science. vol. III. Minneapolis: University of Minnesota Press, 1962, pp. 98-169.

Hintikka, J., and Suppes, P., eds. Aspects of Inductive Logic. Amsterdam: North-Holland Publishing Co., 1966.

Hintikka, J. Knowledge and Belief. Ithaca: Cornell University Press, 1962.

Hintikka, J. "Induction by Enumeration and Induction by Elimination". I. Lakatos, ed. The Problem of Inductive Logic. Amsterdam: North-Holland Publishing Co., 1968, pp. 191-216.

Hintikka, J., and Hilpinen, R. "Knowledge, Acceptance, and Inductive Logic". J. Hintikka and P. Suppes, eds. Aspects of Inductive Logic, 1966, pp. 1-20.

Hilpinen, R. Rules of Acceptance and Inductive Logic. Acta Philosophica Fennica, 21. Amsterdam: North-Holland Publishing Co., 1968.

Hilpinen, R., and Hintikka, J. "Knowledge, Acceptance, and Inductive Logic". J. Hintikka and P. Suppes, eds. Aspects of Inductive Logic. Amsterdam: North-Holland Publishing Co., 1966, pp. 1-20.

Jeffrey, R.C. "Probable Knowledge". I. Lakatos, ed. The Problem of Inductive Logic. Amsterdam: North-Holland Publishing Co., 1968, pp. 166-180.

Kaplan, D. "Explanation Revisited." Philosophy of Science, 37 (1961), pp. 429-436.

Kaplan, D., Eberle, R., and Montague, R. "Hempel and Oppenheim on Explanation." Philosophy of Science. 28 (1961), pp. 418-428.

Kim, J. "On the Logical Conditions of Deductive Explanation." Philosophy of Science. 30 (1963), pp. 286-291.

Kyburg, H.E. "The Rule of Detachment in Inductive Logic." I. Lakatos, ed. The Problem of Inductive Logic. Amsterdam: North-Holland Publishing Co., 1968, pp. 98-119.

Kyburg, H.E. Probability and Inductive Logic. London: Collier-Macmillan, 1970.

Kyburg, H.E. "Recent Work in Inductive Logic". American Philosophical Quarterly. Vol. 1, 1964.

Lakatos, I., ed. The Problem of Inductive Logic. Amsterdam: North-Holland Publishing Co., 1968.

Levi, I. Gambling with Truth. London: Routledge and Kegan Paul, 1967.

Miettinen, S.K. "On Omer's Model of Scientific Explanation". Philosophy of Science. 39 (1972), pp. 249-251.

Montague, R., Eberle, R., and Kaplan, D. "Hempel and Oppenheim on Explanation". Philosophy of Science. 28 (1961), pp. 418-428.

Morgan, C. "Kim on Deductive Explanation". Philosophy of Science. 37 (1970), pp. 434-439.

Morgan, C. "On Two Proposed Models of Explanation". (unpublished manuscript). Recently published in Philosophy of Science. 39 (1972), pp. 74-81.

Morgan, C. "Omer on Explanation". (unpublished manuscript).

Morgan, C. "A Formal Language Model of Explanation". (unpublished manuscript). Read at the Canadian Philosophical Association Meetings, Montreal, June, 1972.

Oppenheim, P., and Hempel, C. "Studies in the Logic of Explanation". Philosophy of Science. 15 (1948), pp. 135-175.

Omer, I.A. "On the D-N Model of Scientific Explanation". Philosophy of Science. 37 (1970), pp. 417-433.

Nagel, E. The Structure of Science. New York: Harcourt, Brace, and World, 1961.

Popper, K. The Logic of Scientific Discovery. London: Hutchinson and Co., 1959.

Popper, K. Conjectures and Refutations: The Growth of Scientific Knowledge. New York: Harper Torchbooks, 1963.

Quine, W.V.O. From a Logical Point of View. Cambridge, Harvard University Press, 1953.

Radnitzky, G. Contemporary Schools of Metascience. Goteborg: Akademiforlaget, 1968.

Rescher, N. Scientific Explanation. New York: The Free Press, 1970.

Salmon, W. "Who Needs Inductive Acceptance Rules?" I. Lakatos, ed. The Problem of Inductive Logic. Amsterdam: North-Holland Publishing Co., 1968, pp. 139-144.

Scriven, M. "Explanations, Predictions, and Laws." H. Feigl and G. Maxwell, eds. Minnesota Studies in the Philosophy of Science. vol. III. Minneapolis: University of Minnesota Press, 1962, pp. 170-230.

Suppes, P. "A Bayesian Approach to the Paradoxes of Confirmation." J. Hintikka and P. Suppes, eds. Aspects of Inductive Logic. Amsterdam: North-Holland Publishing Co., 1966, pp. 198-207.

Suppes, P., and Hintikka, J., eds. Aspects of Inductive Logic. Amsterdam: North-Holland Publishing Co., 1966.

Von Wright, G.H. "The Paradoxes of Confirmation." J. Hintikka and P. Suppes, eds. Aspects of Inductive Logic. Amsterdam: North-Holland Publishing Co., 1966, pp. 208-218.