Transactions of the Nebraska Academy of Sciences and Affiliated Societies

Nebraska Academy of Sciences

1981

Problems for the Covering-Law Model of Explanation

Eric Russert Kraemer
University of Nebraska-Lincoln

1981. *Transactions of the Nebraska Academy of Sciences*, IX:91–94.

PROBLEMS FOR THE COVERING–LAW MODEL OF EXPLANATION

Eric Russert Kraemer

Department of Philosophy
University of Nebraska–Lincoln
Lincoln, Nebraska 68588

The standard covering-law model of explanation sets forth a formal, deductive account of explanation. The account faces two kinds of formal objections. The first is the problem of explanatory relevance. This concerns the specification of formal conditions for relevant logical derivations. The second difficulty is the problem of "self"-explanation. This involves specifications which govern the role that a given statement is allowed to play in an explanatory deduction of itself. A revision of the standard covering-law model provides a natural way to avoid both of these problems. And, it also has the virtue of *not* being an *ad hoc* solution.

† † †

INTRODUCTION

The deductive-nomological (DN) or covering-law model for the explanation of singular events has come under extensive criticism. This article attempts to defend this model from several of the charges that have been made against it and shows that these criticisms may be accommodated by reasonable additions to the model. The criticism that this model is insufficient for scientific explanation of singular events is questioned. (It is not argued here that all explanations fit the DN-model.)

Two formal sorts of criticisms of the DN-model for the explanation of singular events are considered. The first of these is called the problem of relevance, the second, the problem of "self"-explanation.

THE PROBLEM OF RELEVANCE IN DN–EXPLANATION

In their classic discussion of the covering-law model, Hempel and Oppenheim (1948) introduced something like the following as a plausible definition of DN-explanation of singular events:

> Definition (I): An ordered couple of sentences, (T,C), constitutes an explanans for a singular sentence E iff. the following conditions are satisfied: (i) T is a theory, (ii) C is singular and true, and (iii) E is derivable from T and C jointly, but not from C alone.

Although these authors abandoned this definition for a more complicated one, definition (I) will be sufficient for illustrating the problem of relevance, one of two formal problems facing the DN-model.

What is meant by the problem of relevance? Hempel (1966) answered when he listed the following as a "basic requirement for scientific explanation":

> *The requirement of explanatory relevance:* the explanatory information adduced affords good grounds for believing that the phenomenon to be explained did, or does indeed occur.

In a subsequent passage, Hempel made clear what the criteria are for the "good grounds" for belief afforded by DN-explanation:

> DN-explanations satisfy the requirement of explanatory relevance in the strongest possible sense: the explanatory information they provide implies the explanadum sentence deductively and thus offers logically conclusive grounds why the explanandum phenomenon is to be expected.

However, logical deducibility does not seem to be enough to insure that the requirement of explanatory relevance is met. Consider the following case:

(1) All chickens have two legs.
John Doe has as many legs as some chicken.
____
John Doe has two legs.

The conclusion of (1) follows deductively from the theory and initial conditions cited. But, there is something very odd about a theory which deals with chickens claiming to give a scientific explanation of John Doe's bipedal state. Clearly, the conclusion is irrelevant to the theory cited. Whether John Doe has two legs does not affect the theory that all chickens naturally do have two legs.

Hempel's criterion for explanatory relevance as it stands above is inadequate. A suggestion for a better one comes from Hilpinen's (1971) discussion of the problem of relevance in epistemic justification. He remarked:

> If *g* is irrelevant to the justification of *h*, then *h* is justified no matter whether we assume that *g* or *-g* is true, that is, no matter whether *g* or *-g* is added to our original evidence *e*.

In other words, if the conjunction of *e* and *-g* calls *h* into question, then *g* is relevant to the justification of *h*.

To apply this suggestion to DN-explanation, let us substitute theories (*T*'s) for *h*'s, sets of initial conditions (*C*'s) for *e*'s, and explananda (*E*'s) for *g*'s. A definition of the concept of a direct-singular contradiction is needed.

> *D* is a direct-singular contradiction of *T* = Df. (i) *D* is singular, and (ii) existential quantification over *D* may produce something logically equivalent to not *T*.

For example, ∿Fa is a direct-singular contradiction of (x)(Fx), since (Ex) (-Fx) is logically equivalent to -(x) (Fx). Similarly, (Fa & -Ga) is a direct-singular contradiction of (x) (Fx→Gx), since (Ex)(Fx & -Gx) is logically equivalent to -(x) (Fx→Gx); but, ⟦[(Fa→Ga)→Ha] & -Ha⟧ is not a direct-singular contradiction of (x) (Fx→Gx), since even the most likely existential version of the former, namely (Ex) ⟦[(Fx→Gx)→Ha] & -Hx⟧, is not logically equivalent to -(x) (Fx→Gx). Using this machinery, an appropriate modification of definition (I) to include only relevant DN-explanation is as follows:

> Definition (II): The ordered pair, (*T*,*C*), relevantly DN-explains *E* iff.: (i) *T* is a theory, (ii) *C* is singular and true, (iii) *E* is derivable from *T* and *C* jointly, but not from *C* alone, and (iv) (*C* & -*E*) is logically equivalent to a direct-singular contradiction of *T*.

Definition (II) does not meet all formal difficulties, but it does accommodate many problem cases. Observe first how it fits the paradigm case of DN-explanation of singular events:

(2) (x) (Fx→Gx)
Fa
―――――――
Ga

Since (Fa & -Ga) is a direct-singular contradiction of the theory cited in (2), namely (x) (Fx→Gx), by definition (II) case (2) is a relevant DN-explanation. The definition also excludes case (1) as irrelevant. Further, the definition also handles one of the cases which forced Hempel and Oppenheim to abandon definition (I), namely:

(3) (x) (Fx→Gx)
(Fa→Ga)→Ha
―――――――
Ha

In this case, *Ha* is clearly irrelevantly deduced; and, definition (II) says just this. Not surprisingly, definition (II) also handles cases equivalent to (3), such as:

(3′) (x) (Fx→Gx)
(Fa→& -Ga) v Ha
―――――――
Ha

Again, (*C* & -*E*) is not logically equivalent to a direct-singular contradiction of the theory cited.

Although definition (II) accommodates many problem cases, it is not completely satisfactory, for it excludes instances which should be allowed as perfectly acceptable DN-explanations, such as:

(2.1) (x) [(Fx v Gx)→Hx]
Fa
―――――――
Ha

Here (Fa & -Ha) is not a direct-singular contradiction of the theory cited. To meet this objection, the concept of a *trivial logical consequence* (*TLC*) is introduced:

> *P* is a trivial-logical consequence of *Q* = Df. Either (i) Q is logically equivalent to a conjunction of which *P* is a conjunct, or (ii) *P* is logically equivalent to a disjunction of which *Q* is a disjunct.

According to this definition, *P* is a *TLC* of (P & Q), (P v Q) is a *TLC* of *P*, and *P* is a *TLC* of *P* [since *P* is logically equivalent both to (P & P) and to (P v P)].

Clause (iv) of Definition (II) may now be modified as follows:

> (iv′) Either (*C* & -*E*) is logically equivalent to a direct-singular contradiction of *T*, or the conjunction of -*E* and a trivial-logical consequence of *C* is logically equivalent to a direct-singular contradiction of *T*.

Since (Fa v Ga) is a *TLC* of *Fa*, and [(Fa v Ga) & -Ha] is a direct-singular contradiction of (x) [(Fx v Gx)→Ha], deductions such as (2.1) will be included as relevant DN-explanations under this modification of definition (II). It might also be noted that deductions of the following sort are now included as good DN-explanations according to this modification:

$$
\begin{array}{ll}
(2.2) & (x)\,(Fx \rightarrow Gx) \\
 & \underline{Fa \,\&\, Ha} \\
 & Ga
\end{array}
$$

The sort of relevance discussed here differs from that mentioned by Hilpinen (1971) because the projects are different. Hilpinen (1971) was concerned with finding a means for picking out those items of evidence which are relevant to the justification of a given hypothesis, *h,* on a certain evidence base, *e.* I am concerned with finding a criterion for picking out those logical deductions which are relevant to a given theory being used explanatorily in accordance with the DN-model.

For Hilpinen (1971) there was no important restriction on the sort of thing to be considered for relevancy to the justification project. Whatever is such that either it or its negation affects the justification of h on e qualifies as relevant. With DN-explanation, however, there are two restrictions. First, in accordance with the basic guidelines of the DN-model, only those events which follow logically from a given theory and set of initial conditions may be considered as *prima facie* candidates for relevantly explained explananda in terms of that theory. Secondly, not just any logical deduction will do. The task of this section has been to specify those deductions which actually fall within the province of the theory cited.

If it is required only that $(C \,\&\, -E)$ entail a contradiction of the given theory, then, as demonstrated by cases (1), (3), and (3′), theories will be placed in the position of having to DN-explain events of a kind that are completely unrelated to those theories. If, on the other hand, the stronger requirement is demanded, that $(C \,\&\, -E)$ or $(-E \,\&$ a TLC of $C)$ be a direct-singular contradiction of the theory cited, then it seems possible to limit the events that are DN-explainable by a given theory to those events of the kind mentioned in that theory. This latter restriction does not seem to be merely an *ad hoc* maneuver. For it seems well in accordance with the rationale of scientific explanation to require that theories dealing with events of a given kind be able to DN-explain only particular instances of that kind.

THE PROBLEM OF "SELF"-EXPLANATION

Although irrelevant attempts at DN-explanation can be successfully accommodated by the modified version of definition (II), there is another serious, formal problem facing DN-explanation, that of "self"-explanation. By "self"-explanation is meant that the set of initial conditions (C) may contain some part (or all) of the explanandum (E), and that part is required for the logical deduction of that part of the explanandum. According to Kim (1963), instances of partial "self"-explanation such as the following satisfy Hempel and Oppenheim's definition (I):

$$
\begin{array}{llcll}
(4) & (x)\,(Fx\ Gx) & \qquad & (4') & (x)\,(Fx\ Gx) \\
 & \underline{Fa \,\&\, Ra} & & & \underline{Fa \,\&\, Ra} \\
 & Ga \,\&\, Ra & & & Ta \,\&\, (Ra \vee Ja)
\end{array}
$$

Both (4) and (4′) are held to be irrelevant deduction by the modified clause (iv′) of definition (II).

"Self"-explanation seems completely foreign to the concept of DN-explanation. For as Hempel and Oppenheim (1948) said, "Scientific explanation makes essential use of generalized sentences." Any instance of even partial essential "self"-explanation contradicts this maxim. This maxim is understood to say that the whole of an explanandum and not merely some proper part of it must be explained by making essential use of generalized sentences. For this reason it is somewhat disconcerting that in the same paper Hempel and Oppenheim stated:

> In every potential explanation in which the singular component of the explanans is not dispensible, the explanation is partly explained by itself.

They based this claim on the logical fact that the paradigm case of DN-explanation of singular events, that is, case (2), may be rewritten in the following equivalent way:

$$
\begin{array}{ll}
(2') & (x)\,(-Fx \vee Gx) \\
 & \underline{(Fa \vee Ga) \,\&\, (Fa \vee -Ga)} \\
 & (Fa \vee Ga) \,\&\, (-Fa \vee Ga)
\end{array}
$$

They continued that if (Fa v -Ga) is omitted from the initial conditions of (2′), then (2′) is reduced to two parts: (i) an instance of theoretical explanation, (x) (-Fx v Gx) entailing (-Fa v Ga), and (ii) an instance of "self"-explanation, *i.e.,* (Fa v Ga) entailing itself. On the basis of (2′) they refrained from "introducing stipulations to prohibit partial self-explanation" on the grounds that such prohibition would "mean limiting explanation to purely theoretical explanation."

Hempel and Oppenheim's conclusions from (2′) seem confused. The most that follows from (2′) is that in every potential explanation in which the singular component of the explanans is not dispensible, the explanandum *may* be interpreted as being partly explained by itself. However, the derivation in (2′) does not *require* that some part of the explanandum be deduced from itself. For example, the initial conditions cited there may still be used to derive *Fa*; this, when combined with the theory (x) (-Fx v Gx), yields *Ga*, which in turn entails the explanandum given in (2′), without any "self"-explanation. Cases such as (2′) do not necessitate the use of partial "self"-explanation, unlike cases such as (4) and (4′).

The DN-explanation schema given in (2′) is perfectly acceptable. First, according to definition (II), (2′) is a relevant deduction. Second, it should be pointed out that the predicate G, which introduces the possibility of partial "self"-explanation in C of (2′), occurs *inessentially* there. Following Kim (1963), the locution "A occurs inessentially in S" is used as follows:

> A occurs inessentially in S = Df. There exists a sentence S' such that S' is logically equivalent to sentence S, and A does not occur in S'.

Clearly the C of (2′) can be rewritten in an equivalent form in which the predicate G does not appear, *e.g.*, (Fa v Ha) & (Fa v -Ha). Since the explanation can be derived without any "self"-explanation, it is concluded that any partial "self"-explanation that may occur in (2′) is incidental, and not a necessary part of the deduction of the explanandum.

Consider also what happens if (Fa v -Ga) is omitted from the C in (2′) as Hempel and Oppenheim (1948) suggested. The argument then becomes:

(2″) (x) (-Fx v Gx)
Fa v Ga
―――――
(Fa v Ga) & (-Fa v Ga)

This attempt at DN-explanation will not satisfy the modified definition (II), for (2″) cannot meet the relevancy requirement.

Therefore, even though Hempel and Oppenheim thought it impossible to draw a non-arbitrary limit to partial "self"-explanation, a natural line does exist between those cases in which some part (or all) of E occurs inessentially in C and those cases in which some part of E occurs essentially in C.

With this in mind, let us consider instances of relevant explanations which are also partial "self"-explanations. These are cases that, as Kim (1963) suggested, an adequate definition of DN-explanation should rule out.

(5) (x) (Fx→Gx)
Fa v Ga
―――――
Ga

(5′) (x) (Fx & Ix)→Gx)
Fa v Ga
Ia
―――――
Ga

These two are not acceptable as instances of DN-explanation, since certain singular sentences of a particularly vicious sort, namely the explananda themselves, are essentially required in the logical deduction of the conclusion.

These two cases make it clear that the modified definition (II) by itself cannot solve all problem cases. This is not too surprising, as the relevancy requirement by itself cannot exclude the following:

(6) (x) (Fx→Fx)
Fa
―――――
Fa

Case (6) is, of course, dealt with by the second half of requirement (iii), namely "E does not follow from C alone." To accommodate cases such as (5) and (5′), the third requirement of definition (II) may be changed to the following:

> (iii′) E is derivable from T and C jointly and not from C alone, and C does not contain E essentially.

This revision excludes (5) and (5′), for a part of E, namely E itself, is contained essentially in each set of initial conditions listed for these cases.

The doubly amended version of definition (II) [here called definition (II′)], which contains clauses (iii′) and (iv′), limits DN-explanation to those cases which are both relevant and non-redundant. By "non-redundant" it is meant that no part of the logical derivation of the explanandum *requires* "self"-explanation. There is a powerful intuitive reason for this additional requirement: if a theory is not required for every part of the derivation, then the use of DN-explanation, *i.e.*, explanation by deduction from theories or laws of nature, is seriously compromised. Since requirement (iii′) excludes all cases in which some form of "self"-explanation is essential, this requirement seems clearly to be justified in Hempel's (1965) sense, "in terms of the rationale of scientific explanation."

SUMMARY

Two formal problems that arise for the covering-law or DN-model of explanation are discussed. Means by which these problems seemingly can be met are suggested. It is demonstrated that the solution is not *ad hoc*, but is motivated by sound intuitions.

REFERENCES

Hempel, C. G. 1965. *Aspects of scientific explanation and other essays in the philosophy of science.* New York, The Free Press: 504p.

_____. 1966. *The philosophy of natural science.* Englewood Cliffs, New Jersey, Prentice Hall, Incorporated: 116p.

_____, and P. Oppenheim. 1948. Studies in the logic of explanation. *Philosophy of Science,* 15:135-175.

Hilpinen, R. 1971. Knowledge and justification. *Ajatus,* 33: 7-39.

Kim, J. 1963. Discussion: on the logical conditions of deductive explanation. *Philosophy of Science,* 30:286-291.