the probability for a randomly selected member of the human population at large. Moreover, it can plausibly be argued that the evolutionary biologist, in explaining occurrences in that domain, also invokes statistically relevant facts. The issue of high probability vs. statistical relevance has thus been joined. It will prove to be a question of considerable importance in the further development of our story.

The Third Decade (1968–77) *Deepening Differences*

The third decade is bracketed by Hempel's last two publications on scientific explanation. It opened with his emendation of the requirement of maximal specificity (RMS*), which was designed to fix a couple of technical flaws in his account of I-S explanation (1968).[1] It ended with the publication (in German) of a substantial postscript to section 3 of "Aspects" which is devoted to statistical explanation (1977). Among other things, he retracted the high-probability requirement on I-S explanations.

Insofar as published material was concerned, the issue of high probability vs. statistical relevance remained quite dormant for about five years after I had raised it rather obscurely in (1965).[2] Richard C. Jeffrey (1969) argued elegantly that statistical explanations – with the exception of certain limiting cases – are not arguments, and that the degree of probability conferred upon an explanandum by an explanans is not a measure of the goodness of the explanation. My next publication on the topic was in 1970; it contained a theory of statistical explanation, based upon statistical relevance relations, that was spelled out in considerable detail.[3] An ingenious information-theoretic account of scientific explanation, in which statistical relevance relations play a key role, was published by James G. Greeno in 1970. An account of the statistical-relevance (S-R) model of scientific explanation – based on the three papers by Greeno, Jeffrey, and me – was published as a small book the following year (W. Salmon et al. 1971).

The introduction of the inductive-statistical model of scientific explanation constituted, I believe, a crucial turning point for the received view. Before that, I suspect, many philosophers felt (as I did) quite comfortable with the basic idea of D-N explanation, and confident that an equally satisfactory statistical conception would be forthcoming. As things turned out, the I-S model gave rise to a number of fundamental problems. Three avenues presented themselves as ways of coping with the difficulties. The first was to maintain the received view – to pursue the course already laid out by Hempel – seeking to defend the I-S model against objections, and to repair the faults as they were detected. The second was to attempt to construct an alternative account of statistical explanation of particular facts, such as the S-R model, in the hope of avoiding difficulties encountered by the received view. The third was to reject altogether the possibility of providing probabilistic or statistical explanations of particular facts, thereby maintaining strict deductivism with regard to scientific explanation. The deductivist claims that the box in the lower left-hand corner of table 1 is empty; all legitimate explanations are of the D-N variety (where we persist in regarding D-S explanations as a subset of D-N). These three avenues correspond roughly to three basic conceptions that seem to me, in retrospect, to have dominated the discussion of scientific explanation from the time of Aristotle to the present.[4] Their distinctness stands out clearly, I think, only when we try to give an explicit account of statistical explanation. Statistical explanation is like a Stern-Gerlach magnet that separates the incoming notion of scientific explanation into three divergent beams—the epistemic, modal, and ontic conceptions.

3.1 The Statistical-Relevance Model

If one takes seriously the notion that statistical relevance rather than high probability is the key concept in statistical explanation, one naturally asks what role inductive arguments have in the statistical explanation of particular facts. The simple and rather obvious answer is "None."[5] Approaching Hempel's I-S model from a different angle, in a philosophical gem (1969), Richard Jeffrey argued incisively to that very conclusion. He maintained that when a stochastic mechanism—e.g., tossing of coins or genetic determination of inherited characteristics—produces a variety of outcomes, some more probable and others less probable, we understand those with small probabilities exactly as well as we do those that are highly probable. Our understanding results from an understanding of that mechanism and a recognition of the fact that it is stochastic. Showing that the outcome is highly probable, and that it was to be expected, has nothing to do with the explanation. That the outcome was probable or improbable might be added—in the margin, so to speak—as an interesting gloss on the main text. This paper also provided inspiration for some philosophers who adopted an explicitly mechanistic conception of scientific explanation in the fourth decade.[6]

At about the same time, James G. Greeno (1970) developed an information-theoretic account of statistical explanation in terms of the amount of information transmitted by a statistical theory. Transmitted information is a statistical relevance concept, and among its other virtues, it is defined quantitatively. It can be used to evaluate the explanatory power of statistical hypotheses, but it does not yield quantitative evaluations of individual explanations. Within this approach there is no suggestion that statistical explanations should be identified with inductive arguments.

At around the same time I was trying to work out an alternative to the I-S model that would be based on statistical relevance rather than high probability. The first crucial step is to notice that, whereas high probability involves just one probability value, statistical relevance involves a comparison between two probability values. To construct an explanation based upon statistical relevance we need to compare a posterior probability with a prior probability.

Consider John Jones and his quick recovery from a strep infection. Initially, Jones is simply a person with a strep infection—this is the reference class within which he is included for purposes of assessing the *prior* probability of quick recovery. We may assume, I believe, that this probability is fairly small. This reference class is not, however, homogeneous. We can partition it into two subclasses, namely, those who are treated with penicillin and those who are not. Among those who are treated with penicillin the probability of quick recovery is quite high; among those not so treated the probability is much smaller. Since Jones belongs to the subclass treated with penicillin, the *posterior* probability of his quick recovery is much greater than the prior probability. The original reference class has been relevantly partitioned.

We begin by asking why this member of the class of people with strep infections experienced quick recovery. We answer that he belonged to the subclass who received penicillin, noting that the probability in the subclass is different from the probability in the original reference class. As Hempel points out, however, we are not quite done. The subclass of people who are treated with penicillin can be further relevantly partitioned into those whose infection is of the penicillin-resistant strain, and those whose infection is not. The probability of quick recovery for those who have the penicillin-resistant infection is much smaller than is that for those who have a non-penicillin-resistant infection. We may therefore answer the why-question by stating that Jones recovered quickly from his strep infection because his infection was of the non-penicillin-resistant strain and it was treated with penicillin.

Let us define some of the terms we are using. By a *partition* of a class F we mean a set of subclasses that are mutually exclusive and exhaustive within F—that is, every member of F belongs to one and only one member of the partition. Each of these subclasses is a *cell* of the partition of F. A partition is *relevant* with respect to some attribute G if the probability of G in each cell is different from its probability in each of the other cells. The possibility is left open that the probability of G in one of the cells is equal to the probability of G in the original reference class F.[7] A class F is *homogeneous* with respect to the attribute G if no relevant partition can be made in F. It is *epistemically homogeneous* if we do not know how to make a relevant partition—if, that is, our total body of knowledge does not contain information that would yield a relevant partition. The class is *objectively homogeneous* if it is impossible in principle, regardless of the state of our knowledge, to make a relevant partition. We may then define a *homogeneous relevant partition* as a relevant partition in which each cell is homogeneous. We may then distinguish in an obvious way between *epistemically homogeneous relevant partitions* and *objectively homogeneous relevant partitions*.

Two points should be carefully noted: (1) when an objectively homogeneous relevant partition of a reference class has been given, all relevant factors have been taken into account – i.e., all relevant partitions have been effected – and (2) a relevant partition admits *only* relevant factors, since no two cells in the partition have the same probability for the attribute G.

These concepts deserve illustration. Greeno (1970) offers the example of an American teenager, Albert, who is convicted of stealing a car (D = delinquent offense).[8] Albert is a boy and he lives in an urban environment (San Francisco). Take as the original reference class for purposes of explanation the class of American teenagers (T). If we subdivide it into American teenage boys (M) and American teenage girls (F), we will make a relevant partition, for the probability that a boy will steal a car is greater than the probability that a girl will do so. Moreover, if we partition the class of American teenagers into urban dwellers (U) and rural dwellers (R), we will make another relevant partition, for the delinquency rate in urban areas is higher than it is in rural areas. Taking both of these partitions into account, we have four subclasses of the original reference class – namely, urban male, urban female, rural male, and rural female – and we find that the probability of delinquency is different in these four cells of the combined partition. Symbolically, we can write

$$P(D|T.C_i) = p_i; \quad p_i \neq p_j \text{ if } i \neq j$$

where

$$U.M = C_1; \quad U.F = C_2; \quad R.M = C_3; \quad R.F = C_4.$$

This is a relevant partition, but it is obviously not homogeneous, for there are many other factors, such as religion and socio-economic status, that are relevant to delinquency.

Hempel's example of quick recovery (Q) from a strep infection provides a different illustration. In this example, the original reference class is the class of people who have streptococcus infections (S). Clearly, the partition of this class into those who receive treatment with penicillin (T) and those who do not ($\sim T$) is relevant. Also, a partition can be made into those who have penicillin-resistant infections (R) and those whose infections are not penicillin-resistant ($\sim R$). If, however, we combine these two partitions to form the cells

$$T.R = C_1; \quad T.\sim R = C_2; \quad \sim T.R = C_3; \quad \sim T.\sim R = C_4$$

the resulting partition is not a relevant partition, for the probabilities

$$P(Q|C_i) = p_i$$

are not all different. If the infection is penicillin resistant, it makes no difference (I assume) whether penicillin is administered or not, and (I further assume) the probability of quick recovery is the same for a person with a non-penicillin-resistant infection who does not receive penicillin as it is for anyone with a penicillin-resistant strain. Thus it appears that

$$p_1 = p_3 = p_4 \neq p_2,$$

so our relevant partition of S is

$$S.C_1 = S.\sim R.T; \quad S.C_2 = S.(R.T \vee R.\sim T \vee \sim R.\sim T).$$

This partition is not likely to be homogeneous, for – as Philip Kitcher pointed out – a further partition in terms of allergy to penicillin would be relevant.

The basic structure of an S-R explanation can now be given. Although there are many different ways to ask for a scientific explanation, we can reformulate the request into an explanation-seeking why-question that has the canonical form, "Why does this (member of the class) A have the attribute B?" When an explanation is requested, the request may come in the canonical form – for example, "Why did Jones, who is a member of the class of people who have strep infections, recover quickly?" Or, it may need translation into standard form – for example, "Why did Albert steal a car?" becomes "Why did this American teenager commit a delinquent act?" A is the reference class for the prior probability, and B is the attribute. When a translation of this sort is required, pragmatic considerations determine the choice of an appropriate reference class.

An S-R explanation consists of the prior probability, a homogeneous relevant partition with respect to the attribute in question, the posterior probabilities of the attribute in cells of the partition, and a statement of the location of the individual in question in a particular cell of the partition:

$P(B|A) = p$
$P(B|A.C_1) = p_1; \quad P(B|A.C_2) = p_2; \ . \ . \ .$
$A.C_1$, $A.C_2$, . . . constitute a homogeneous relevant partition of A
b is a member of $A.C_k$

Given our definition of a homogeneous relevant partition, we know that our explanation appeals to all of the factors relevant to the explanandum, and only to factors that are relevant.

These relevance considerations enable us to deal not only with such counterexamples to the I-S model as CE-8 (vitamin C and colds) and CE-9 (psychotherapy vs. spontaneous remission), but also with such counterexamples as CE-3 (barometer) that we encountered much earlier in connection with the D-N model. To do so, we need a relation known as *screening off*.

Consider the fact that a sharp drop in the reading on a barometer (B) is highly relevant to the occurrence of a storm in the vicinity (S). The probability of a storm, given the sharply falling barometric reading, is much higher than its probability without reference to the barometric reading, i.e.,

$$P(S|B) > P(S).$$

This is the reason that the barometer is a useful instrument for predicting storms. However, as we remarked above, the reading on the barometer does not explain the storm. The fact is that a certain set of atmospheric conditions, including a sharp decrease in atmospheric pressure, are responsible both for the storm and for the reading on the barometer. Indeed, the actual drop in atmospheric pressure renders the reading on the barometer irrelevant. The probability of a storm (S), given both the drop in atmospheric pressure (A) and the drop in the reading on the barometer (B), is equal to the probability of a storm, given only the drop in atmospheric pressure—i.e.,

$$P(S|A.B) = P(S|A).$$

Thus, although B is relevant to S if A is not taken into account, B is irrelevant to S in the presence of A. This is what it means to say that A screens B off from S. It should be noted that B does not screen A off from S. If the barometer gives a faulty reading because of some malfunction, or if it is placed in a vacuum chamber and thus caused to register a sharp drop, that will have no effect on the probability of a storm. Thus,

$$P(S|A.B) \neq P(S|B).$$

Indeed, destroying all of the barometers inside a thousand mile radius will not affect the probability of storms in the least.

According to the S-R model, no irrelevant factors should be included in an explanation. "To screen off" *means* "to render irrelevant." Accordingly, no factor that is screened off can be invoked legitimately in an S-R explanation. The barometer example is an instance of a large and important class of cases in which a common cause gives rise to two or more effects that are correlated with one another. The atmospheric conditions cause both the barometric reading and the storm. In such cases, we want to block any attempt to explain one of these common effects in terms of others. This can often be accomplished by appealing to the fact that the common cause screens any one of the correlated effects off from any of the others.

Hempel provides another example of just this type. He points out that Koplik spots—small white blisters on the inside surface of the cheek—are a very reliable sign of a measles infection, and they can be used to predict that other symptoms, such as a skin rash or a fever, will soon appear. All of these symptoms, including the Koplik spots, are among the effects of the measles infection. According to the S-R model, then, the Koplik spots have no part in explaining the other symptoms of the disease. Hempel (1965, 374–75) suggests, however, that there might have

been occasions, at a time when the nature of the underlying infection was totally unknown, on which it could have been appropriate to invoke them in order to explain the other symptoms.

In my earliest criticism of the I-S model, I suggested that positive relevance rather than high probability is the key explanatory relation. At that time I had not worked out the details of any alternative model. In developing the S-R model—leaning heavily on the work of Jeffrey and Greeno—I came to the conclusion that positive relevance is not required, but that negative relevance could also have explanatory import. The reasons can be given quite straightforwardly and simply. If one makes a relevant partition of a reference class, in some cells the posterior probability of the attribute will be higher than the prior probability, and in some the posterior probability will be lower. This point is obvious in the case of a partition with just two cells. Suppose that we have a large class of coin tosses, and that the probability of heads in that class is ½. Suppose, further, that this reference class can be partitioned into two subclasses, one consisting of tosses of a coin heavily biased toward heads (0.9), the other consisting of tosses of another coin equally biased toward tails. Among the tosses of the second coin some yield heads; the probability for this result in this cell of the partition is 0.1. However, according to an argument of Jeffrey mentioned briefly above, we understand the low-probability outcomes of any given stochastic process just as well as we understand the high-probability outcomes; we understand the heads resulting from tosses of the coin biased for tails just as well as we understand the tails, and by parity of reasoning, we understand the tails resulting from tosses of the coin biased toward heads just as well as we understand the heads. Thus, it appears, we understand all of the results in the original reference class equally well. Some have high probabilities, some have low. Some are found in cells of the relevant partition which reduce their probability, some in cells that increase their probability. Negatively relevant factors as well as positively relevant factors contribute to our understanding of the phenomena we are studying.

We are now in a position to offer a succinct contrast between the I-S and S-R models of statistical explanation.

> I-S model: An explanation of a particular fact is an *inductive argument* that confers upon the fact-to-be-explained a *high inductive probability*.
>
> S-R model: An explanation of a particular fact is an assemblage of facts *statistically relevant* to the fact-to-be-explained *regardless of the degree of probability* that results. (W. Salmon et al. 1971, 11)

Both models conform to the covering law conception of scientific explanation, for they both require laws in the explanantia. In particular, in an S-R explanation, the statements of values of prior and posterior probabilities qualify as statistical laws.

3.2 Problems with Maximal Specificity

Hempel's recognition of what he termed "the ambiguity of I-S explanation," which led to the requirement of maximal specificity and the doctrine of essential epistemic relativization of I-S explanation, gave rise to deep difficulties. J. Alberto Coffa was, to the best of my knowledge, the first to recognize their profundity. In a classic article, "Hempel's Ambiguity" (1974) he spelled them out with great care.[9]

Coffa begins by identifying the following schema

$$\text{(I)}\quad \begin{array}{l} (x)[Fx \supset Gx] \\ Fb \\ \hline Gb \end{array}$$

as the 'basic schema' for the simplest form of D-N explanation (1974, 142). Since the following schema

$$\text{(II)}\quad \begin{array}{l} P(G|F) = r \\ Fb \\ \hline\hline Gb \end{array}\ [r]$$

is the obvious inductive analog of schema I, we might be tempted to offer *the naïve model of inductive explanation* in which an I-S explanation is an argument of form II, where r is close to 1, the first premise is a statistical law, and the premises are true. But, as we have seen, given an instance of (II), we can often find another argument of the form

$$\text{(III)}\quad \begin{array}{l} P({\sim}G|H) = s \\ Hb \\ \hline\hline {\sim}Gb \end{array}\ [s]$$

where s is close to 1, the first premise is a statistical law, and the premises are true as well. Such pairs of arguments exemplify what Hempel called "the ambiguity of I-S explanation," and its existence utterly undermines the naïve model (Coffa 1974, 143–44).

The existence of such pairs of arguments is an example of what Hempel (1960) called "inductive inconsistency"—namely, two inductive arguments with mutually consistent (indeed, true!) premises that lend strong inductive support to two mutually inconsistent conclusions. Finding this situation intolerable in the domain of scientific explanation, Hempel seeks a ground for rejecting at least one of the two 'explanations.' Having briefly suggested in his first article on statistical explanation (1962) that the requirement of total evidence be invoked, he subsequently

formulated (and reformulated (1968)) the *requirement of maximal specificity* to do the job. But, Coffa points out, if the problem were really just the inductive inconsistency, a *much* simpler device would suffice. Since, as Hempel seems to agree, we explain only actual occurrences, not things that do not happen, it is easy to choose the correct explanation from the foregoing pair. It is the one with the true conclusion. Since Hempel did not adopt this obvious and easy solution, Coffa suggests, he must *not* have been exercised about *this* problem. When we look at the way he actually tried to solve the problem, we see that the *real* problem was not the problem of inconsistency, but rather the venerable *reference class problem*.

The problem of choosing the appropriate reference class arose traditionally when adherents of the frequency interpretation of probability, such as John Venn and Hans Reichenbach, attempted to apply probabilities to single cases. What is the probability, for example, that William Smith will still be alive 15 years from now? This is a question of considerable importance to his wife and children, to his employer, and to his insurance company. The problem is that he belongs to many different reference classes, and the frequencies of 15-year survival may differ greatly from one of these classes to another. Within the class of American males it is fairly high, among 40-year-old American males it is even higher, among grossly obese Americans it is not nearly as high, and among heavy cigarette-smoking individuals it is still different. If we are going to try to predict whether he will be alive in 15 years, or estimate his chances of surviving that long, we will want to take into account all available relevant evidence.

Starting with a very broad reference class, such as the class of all Americans, we should partition it in terms of such factors as sex, age, occupation, state of health, etc., until we have taken into account all known factors that are statistically relevant. By the same token, we do not want to partition the class in terms of factors that are known to be irrelevant to his survival, or in terms of factors whose relevance or lack thereof is unknown. We may say that the rule for assigning a probability to a single case is to refer it to the *broadest homogeneous reference class* available, where it is understood that the class is *epistemically* broad and homogeneous—i.e., we have not used any partitions that are not known to be relevant and we do not know how to make any further relevant partitions (see W. Salmon 1967, 90–96, and Salmon et al. 1971, 40–47). Since we are concerned here with prediction, the maxim is to use all available evidence; epistemic relativization is entirely appropriate.

The problem of choosing the correct reference class for explanatory purposes is quite different. When we want to explain some fact, such as John Jones's rapid recovery from his strep infection, we already know that his recovery has occurred—i.e., the explanandum is already part of our total evidence—so we must not appeal to that portion of our total evidence. The reference class problem becomes the problem of determining precisely to what part of our total evidence we are allowed to appeal in constructing an I-S explanation. Hempel's RMS and its successor RMS* were designed to accomplish this aim. Roughly speaking (referring back to the schema II), both versions of the requirement of maximal specificity stipulated that, *in a knowledge situation K*, every known relevant partition of F, except one in terms of G or in terms of properties *logically* related to G, be made. The idea of this requirement is to tell us what factors must be taken into account and what factors must not be taken into account in constructing I-S explanations. Hempel attempted to prove, for each version of the requirement, that it would eliminate the unwholesome ambiguity. RMS* was devised in response to a counterexample to RMS constructed by Richard Grandy (Hempel 1968).

The profound problem Coffa recognized arises, not from the observation that *many* explanations that fit schema II are subject to ambiguity, but from Hempel's assertion that *all* of them are. It is a striking fact that Hempel does not offer arguments to support this transition from many to all (Coffa 1974, 151). Yet, this is the crucial issue, for it is only from the assertion of *universal* ambiguity that the doctrine of *essential* epistemic relativity arises. So Coffa does two things. First, he attempts to reconstruct conjecturally the kind of considerations that may have led Hempel to his conclusion of essential epistemic relativity (151–55), and second, he maintains that it leads to disaster for the theory of I-S explanation (155–60).

Coffa argues that Hempel was looking for some condition to impose on the law-premise of schema II–"$P(G|F) = r$"–that would guarantee that the problem of ambiguity would not arise. RMS (later, RMS*) were the conditions he proposed. Each contains a reference to a knowledge situation, thus implying that the explanation in question would be epistemically relativized. It appears, then, that Hempel felt that no adequate substitute for RMS could exist that did not also contain reference to a knowledge situation. Consequently, according to Hempel, I-S explanation is *essentially epistemically relativized*.

Both Coffa and I agreed that epistemic relativization of probabilistic explanation could be avoided–though we took different tacks in trying to do so–but before looking at them we should consider the consequences of accepting epistemic relativity. In (1971) I thought that nothing of much importance hinged on the issue of epistemic relativity; Coffa's 1974 article changed my mind radically on this point (W. Salmon 1974).

Coffa begins his discussion of epistemic relativization of I-S explanation by analyzing the distinction between epistemic and non-epistemic concepts:

> It is an obvious fact that the meaning of some expressions or concepts can be given without referring to knowledge, whereas that of others cannot. Let me call the latter epistemic and the former non-epistemic. Examples of non-epistemic expressions are easy to find. 'Table,' 'chair,' 'electron,' according to many, 'truth,' would be typical instances. Examples of epistemic notions are also readily available. The best known instance may be that of the concept of confirmation. Although the syntactic form of expressions like 'hypothesis *h* is well-confirmed' may mislead us into believing that confirmation is a property of sentences, closer inspection reveals the fact that it is a relation between sentences and knowledge situations and that the concept of confirmation cannot be properly defined (that is, its meaning cannot be given) without a reference to sentences intended to describe a knowledge situation. . . .
>
> Having introduced the distinction between epistemic and non-epistemic concepts, we go on to notice that there is a further interesting distinction to be drawn within the class of epistemic notions based upon the kind of role knowledge plays in them. On the one hand there are those epistemic notions in which knowledge enters essentially as an argument in a confirmation function, or, equivalently, as an ingredient in a statement of rational belief. And then there is the obscure and largely unintelligible remainder.
>
> In the first group we find a significant example provided by Hempel's theory of deductive explanation. After introducing his non-epistemic notion of D-N explanation Hempel went on to say that he could define now the concept of a well-confirmed D-N explanation, a well-confirmed D-N explanation in a tacitly assumed knowledge situation *K* being, in effect, an argument which in knowledge situation *K* it is rational to believe is a D-N explanation, i.e., a true D-N explanation. In precisely the same fashion we could correctly and uninterestingly define the concepts of well-confirmed table, well-confirmed chair and well-confirmed electron, given that we started by having the concepts of table, chair and electron. *Since we can only have reason to believe meaningful sentences, a confirmational epistemic predicate is an articulation of independently meaningful components.*
>
> Of course we can understand what a well-confirmed chair is because we began by understanding what a chair is. If '*x* is a chair' had not had a meaning, we would not even have been able to make sense of the statement of rational belief made about it. Similarly, we can understand, if not appreciate, the notion of well-confirmed D-N explanation, because we were told first what kind of thing a D-N explanation is.*
>
> Now we are in a position to state Hempel's thesis of the epistemic relativity of inductive explanation. As a consequence of his analysis of the phenomenon of ambiguity, Hempel concludes that the concept of inductive explanation, unlike its deductive counterpart, is epistemic; and he goes on to add that it is not epistemic in the sense in which well-confirmed deductive explanations are.

*Prof. Gerald Massey has drawn my attention towards the apparent opacity of epistemically relativized predicates. He has pointed out that this raises serious doubts concerning the possibility of viewing them as expressing properties. (Coffa's note).

The concept of inductive explanation is a non-confirmational epistemic concept. . . .

As Hempel is careful to point out, this means that there is no concept that stands to his epistemically relativized notion of inductive explanation as the concept of true D-N explanation stands to that of well-confirmed D-N explanation. *According to the thesis of epistemic relativity there is no meaningful notion of true inductive explanation*. Hence, we couldn't possibly have reasons to believe that anything is a true inductive explanation. Thus, it would be sheer confusion to see inductive explanations relative to K in Hempel's sense as those inductive arguments which in knowledge situation K it is rational to believe are inductive explanations.

It is clear that, according to Hempel, *there is a remarkable and surprising disanalogy between deductive and inductive explanations*. When somebody asks us to give an account of deductive explanations, we can do so without referring to anybody's knowledge. If asked, for instance, what sort of thing would it be to explain deductively the present position of a planet, we would refer to descriptions of certain nomic and non-nomic facts but never to anybody else's knowledge. This is a desirable feature in a non-psychologistic account of explanation. Yet, according to Hempel, when we ask what an inductive explanation of the same event would look like, there is no way in which an appropriate answer can be given without talking about knowledge; presumably the knowledge available at the time of the explanation. Even more surprisingly, this reference to knowledge does not play the standard role that such references usually play, to wit, that of providing the epistemic platform for a judgment of rational belief. What role such reference plays is a question which deserves serious attention, since here we find the Achilles' heel of Hempel's whole construction. (1974, 148–50, emphasis added)

Having established the distinction between epistemic and non-epistemic concepts, and having pointed out, within the former category, the distinction between confirmational and non-confirmational concepts, Coffa now announces the result he intends to draw.

I have argued that from a certain problem Hempel felt forced to draw the conclusion that the notion of inductive explanation is epistemically relativized. . . . Now, I would like to explain why I find Hempel's conclusion worth avoiding. I will try to convince you that *to accept Hempel's thesis of epistemic relativity amounts to accepting the claim that there are no inductive explanations*, the concept of I-S explanation relative to K functioning as a placebo which can only calm the intellectual anxieties of the uncautious user. If I am right, anyone willing to hold that there are inductive explanations will have to begin by spotting a flaw in Hempel's argument. . . .

Maybe the best way in which I can briefly convey my feelings about the oddity implicit in Hempel's theory of inductive explanation is by noting, that in my view, Hempel's decision to develop a theory of I-S explanation relative to K after having argued that the notion of true inductive explanation makes no sense, seems comparable to that of a man who establishes conclusively that Hegel's philosophy is strict nonsense, and then proceeds to devote the rest of his life to produce the definitive edition of Hegel's writings. (1974, 155, emphasis added)

The strategy Coffa adopts to carry out his assault on Hempel's I-S model of explanation and, indeed, any essentially epistemically relativized model of explanation, is to pose a fundamental challenge for Hempel.

Now the question I would like to put to Hempel is the following. Take any I-S explanation relative to K, for some given K. . . . What is it about this inductive argument that makes it an explanation of its last formula? What reason could anyone have to say that it is an explanation of its conclusion?

It is not difficult to answer this question when we pose it, not for the inductive, but for the deductive case. If one asks, for example, what reason we have to believe that a causal deductive explanation explains its explanandum, the answer is that its premises identify certain features of the world that are nomically responsible for the occurrence of the explanandum event.

Could we say, as in the deductive case, that I-S explanations relative to K explain because their premises somehow identify features of the world that are nomically responsible for the explanandum event? Certainly not. This is what we vaguely conceived to be possible while tacitly espousing the naïve model, until Hempel shattered our illusions to pieces by focusing the reference class problem on the theory of explanation. Indeed, *if there is no characterization of true inductive explanation, then it must be because there are no such things that go on in the non-epistemic world of facts that can inductively explain the event*. For if there were such non-epistemic goings on, their characterization would be a characterization of true inductive explanation. Thus, the possibility of a notion of true explanation, inductive or otherwise, is not just a desirable but ultimately dispensable feature of a model of explanation: it is the *sine qua non* of its realistic, non-psychologistic inspiration. It is because certain features of the world can be deterministically responsible for others that we can describe a concept of true deductive explanation by simply describing the form of such features. If there are features of the world which can be non-deterministically responsible for others, then we should be able to define a model of true inductive explanation. And, conversely, if we could define a model of true inductive explanation, there could be features of the world non-deterministically responsible for others. The thesis of epistemic relativity implies that, for Hempel, there are no such features. What, then, is the interest of I-S explanations relative to K? Surely not, as we have seen above, that in knowledge situation *K* we have reason to believe that they are inductive explanations. Then what? We detect in Hempel's writings not even a hint as to what an answer to this question might be. (1974, 157-58, emphasis added)

I have quoted from Coffa at length to convey the subtlety and cogency of his argument, and to capture the style and clarity of his exposition. I think his conclusion is entirely correct.

Now that we have exhibited the consequences of the doctrine of essential epistemic relativity of I-S explanation, let us examine more closely the considerations that seem to have motivated that doctrine. Coffa has shown beyond doubt that it arises somehow from the reference class problem. Looking at the problem in the context of the frequency interpretation of probability, in which it originally arose, the problem seems to be that any reference class we choose will be inhomogeneous.

Coffa considers the example of Jones who is both a Texan and a philosopher. Given that most Texans are millionaires and that most philosophers are not millionaires, we have a naïve inductive explanation of Jones being a millionaire and of Jones not being a millionaire (1974, 144). Obviously, neither the class of Texans nor the class of philosophers is homogeneous with respect to being a millionaire, and it appears that there will always be ways of partitioning any given subclass that will yield different relative frequencies. Perhaps Jones was born at the moment a Chinese mandarin sneezed, and perhaps the class of Texas philosophers born at any such moment had a proportion of millionaires different from that in the class of Texan philosophers (1974, 154). Coffa offers a suggestion – he does not spell it out in detail – that the main difficulty can be circumvented by thinking of probabilities in terms of propensities rather than frequencies. I shall return to this idea below. In the meantime, let me follow the frequentist line a little further.

If we consider some less fanciful cases, such as Greeno's delinquency example or Hempel's example of the quick recovery from a streptococcus infection, a plausible line of argument emerges. In explaining why Albert stole a car, we saw that there were many relevant factors that could be summoned – sex, age, place of residence, socio-economic status of the family, religious background, etc. The list seems almost endless. Whenever we arrive at a narrower reference class as a result of another relevant partition, if that class still contains youths who do not commit delinquent acts, we are tempted to suppose that there is some further factor that relevantly distinguishes Albert from the non-delinquents. The process of relevant partitioning will end only when we have found a reference class in which all members are juvenile delinquents. At that point, it may be assumed, we will have a bona fide explanation of Albert's delinquency. It should be noted emphatically that *the resulting explanation is not inductive, it is deductive-nomological*. Obviously, whenever we reach a universal generalization of the form "(x)[Fx ⊃ Gx]" we have a reference class F that is trivially homogeneous with respect to the attribute G, for every F is a G, and so is every member of every subclass of F. For this reason, D-N explanations automatically satisfy RMS. A similar analysis could plausibly be offered in the case of Jones's quick recovery from his strep infection.

One reason for claiming that I-S explanations are essentially epistemically relativized is the supposition that the statistical generalization among the premises does not contain a homogeneous reference class – i.e., that the reference class can, in all cases, in principle, be relevantly partitioned, even if we do not know how to do it at the time. One obvious motivation for such a claim would be a commitment to determinism. According to the determinist, every event that happens – including Albert's theft of an auto and Jones's quick recovery from his strep infection – are completely determined by antecedent causes. Those causes define a reference class in which every member has the attribute in question. If we construct an I-S explanation, the statistical law that occurs as a premise must embody an inhomogeneous reference class – i.e., one that could be relevantly partitioned. The only reason for using such a reference class would be ignorance of the additional factors needed to effect the further partition.

If this were the situation, we could easily understand the epistemic relativity of I-S explanations. If determinism is true, all bona fide explanations are deductive. I-S explanations are not well-confirmed inductive explanations; they are simply incomplete D-N explanations. An I-S explanation is analogous to an enthymeme – a deductive argument with missing premises. By supplying missing premises we can make the argument more complete, but when all of the missing premises are furnished the argument is no longer an enthymeme. At that point it becomes a valid deductive argument. There are no valid enthymemes, for, by definition, they lack premises needed for validity.

Similarly, an I-S explanation is simply an incomplete D-N explanation as long as it embodies an inhomogeneous reference class. As more and more relevant partitions are made, it approaches more closely to a D-N explanation, but it ceases to be an I-S explanation when homogeneity is achieved because, according to the deterministic assumption, homogeneity obtains only when the statistical law has been transformed into a universal law. From the determinist's standpoint, then, there are no genuine inductive explanations; the ones we took to be inductive are simply incomplete deductive explanations. It is easy to see what epistemic relativity amounts to in the context of determinism. When I posed the situation in this way to Hempel – suggesting that he was implicitly committed to determinism (1974) – he informed me emphatically that he is not a determinist (private correspondence).

An indeterminist, it seems to me, is committed to the view that there are reference classes in bona fide statistical laws that are homogeneous – not merely epistemically homogeneous, but objectively homogeneous. One might suppose that the class of carbon-14 nuclei is homogeneous with respect to spontaneous radioactive decay within the next 5730 years (the half-life of C^{14})—that there is, in principle, no way to make a relevant partition. Unlike William Smith (with whose survival we were concerned), C^{14} atoms do not age, suffer ill health, indulge in hazardous occupations, etc. Their chances of suffering spontaneous decay are remarkably unaffected by their environments. If quantum theory is correct, there is no way to select a subclass of the class of C^{14} nuclei in which the probability of spontaneous decay within the next period of 5730 years is other than one-half. In other words, modern physics strongly suggests that there are nontrivial cases of objectively homogeneous reference classes.

If one accepts the notion that there may be objectively homogeneous reference classes, then one might construct an unrelativized model of I-S explanation along the following lines:

An argument of the form

$$\text{(II)}\quad \begin{array}{l} P(G|F) = r \\ Fb \\ \hline\hline Gb \end{array} \quad [r]$$

is an I-S explanation of Gb, provided that (1) the first premise is a statistical law, (2) both premises are true, (3) r is close to 1, and (4) F is an objectively homogeneous reference class with respect to G.

Since this characterization makes no reference to a knowledge situation, it can be taken as an explication of a *true I-S explanation*. An argument of form (II) is a *well-confirmed I-S explanation* with respect to a knowledge situation K if the information in K provides good grounds for believing that it satisfies the conditions following schema (II) above. An argument of form (II) that satisfies conditions (1)-(3), but not condition (4), is—at best— an *incomplete I-S explanation*. An argument of form (II) that qualifies as an incomplete I-S explanation could be considered an *optimal I-S explanation with respect to knowledge situation K* if the reference class F is epistemically homogeneous—that is, in knowledge situation K we do not know how to effect a relevant partition (although a relevant partition is possible in principle). In the foregoing explication, the role of RMS is taken over by condition (4).

As I understood his communication, Hempel implicitly rejected the foregoing strategy, not because of a commitment to determinism, but, rather, because of deep doubts about the intelligibility of the notion of objective homogeneity. On the one hand, these doubts were surely justified as long as no one had produced a reasonably clear explication of the concept. As I learned in the process of attempting to provide such an explication, it is no easy task. I shall return to this problem below. On the other hand, there were two strong reasons for thinking

that objective homogeneity is an intelligible concept. First, in the trivial case of universal generalizations (affirmative or negative), the concept is clearly applicable. Second, in a vast range of nontrivial cases, we can easily see that the negation of the concept is applicable. We can be quite certain that the class of American teen-agers is not homogeneous with respect to delinquency, and that the class of humans is not homogeneous with respect to survival for an additional 15 years (W. Salmon 1984, 53-55). It would be surprising to find that the negation of a meaningful concept were unintelligible. So one wonders whether a deterministic prejudice is not operating after all. As Peter Railton (1980) most aptly remarked, some people who do not hold the doctrine of determinism are nevertheless held by it.

It seems to me that proponents of either the I-S model or the S-R model must face the reference class problem. Since the S-R model appealed to objective homogeneity, it was clear that an explication of that concept was sorely needed, and I attempted to offer one (1977).[10] The I-S model (in Hempel's epistemically relativized version) also required a homogeneity condition, but, of course, an epistemic one. RMS was introduced for this purpose. It seems to me, however, that neither RMS nor RMS* was adequate even to Hempel's needs, and that examining its inadequacy provides a good start on the analysis of objective homogeneity. For this purpose, we will need to look at RMS in detail.

When Hempel first elaborated his I-S model of explanation, he realized that a crucial problem concerning relevance arises. He sought to deal with it in terms of RMS. The basic idea behind that requirement is that, subject to certain important restrictions, all available relevant knowledge should be brought to bear when I-S explanations are constructed. The problem concerns these restrictions.

The most obvious consideration, as we have seen, is that when we attempt to explain any given fact we already know that it is a fact—for example, when we try to explain John Jones's rapid recovery from his strep infection (Gb), we know that the rapid recovery has occurred. Thus, if we were to bring to bear all of our relevant knowledge, we would have to place him in the class of people who have strep infections (F), who are treated with penicillin (H), and who experience quick recoveries (G). Relative to that class, the probability of his rapid recovery is trivially equal to 1; indeed,

$$P(G|F.H.G) = 1$$

is a theorem of the mathematical calculus of probability. The fact that he belongs to that class clearly has no explanatory value. Let us recall the formulation of RMS. With respect to arguments that satisfy schema (II), let s be the conjunction of the premises, and let k be a statement that is equivalent to the content of our body of knowledge K. RMS says:

> If s.k implies that b belongs to a class F_1, and that F_1 is a subclass of F, then s.k must also imply a statement specifying the statistical probability of G in F_1, say
>
> $P(G \mid F_1) = r_1$
>
> Here, r_1 must equal r unless the probability statement just cited is simply a theorem of mathematical probability theory. (Hempel 1965, 400)

If the restriction stated in the unless-clause were not incorporated in RMS, there could be no I-S explanations.

The unless-clause is not, however, strong enough. Suppose, for instance, that John Jones is a prominent public figure whose illnesses are newsworthy. On the evening of his recovery the local television news program reports that fact. Thus, he belongs to the class of individuals who have strep infections (F) treated by penicillin (H) whose quick recoveries are reported by reliable news media (K). Let us assume, at least for the sake of the example, that the probability that a quick recovery took place (G), given that it is so reported (K), is virtually 1; hence, it is greater than the probability of the quick recovery, given only the administration of penicillin. Since it is not a theorem of the probability calculus that $P(G \mid F.H.K) = 1$, Hempel's unless-clause does not block the use of that relation in connection with RMS. Thus, RMS would disqualify Hempel's original example, even though the report of Jones's recovery is irrelevant to the explanation of the recovery. Although the news report may be the source of our knowledge that Jones has recovered quickly, it plays no part in telling us why the quick recovery occurred. Though it may answer an *evidence-seeking* why-question, it contributes nothing toward an answer to the *explanation-seeking* why-question. Thus, the unless-clause needs to be strengthened.

In a probability expression of the form "$P(Y \mid X)$" X is known as the *reference class* and Y as the *attribute class*. The problem RMS is intended to address is that of choosing a statistical law with a suitable—i.e., *maximally specific*—reference class for an I-S explanation of any particular fact. A reference class F is maximally specific with respect to any attribute class G if our available knowledge does not provide any way of making a relevant partition of it—i.e., to pick out a subset of F in which the probability of G is different than it is in F. But we cannot allow complete freedom in the choice of properties in terms of which to make such partitions. Certainly the restriction formulated in Hempel's unless-clause is required, but as our example has just shown, additional restrictions are also needed. Even if a relevant partition of a reference class can be made—but only in terms of information not available until after the occurrence of the fact-to-be-explained—the reference class should not on that account be ruled out as not maximally specific.

A further restriction is needed, as can easily be seen in terms of another example. Suppose we want to know why, in a genetic experiment, a particular plant has a red blossom. In answer, we are told that, in this species, the color of the blossom is determined by a single gene and that red is a recessive characteristic. In addition, both parent plants had red blossoms. Therefore, barring a fairly improbable mutation, the offspring would have red blossoms. This is a legitimate I-S explanation, for given the statistical laws and antecedent conditions, the explanandum is highly probable, but not certain.

This explanation could, however, be challenged on the basis of RMS. Given certain well-known facts about human vision, we know that red is at the opposite end of the visible spectrum from violet. This is *not* a matter of definition.[11] Therefore, it could be said, we know that the blossom is on a plant both of whose parents had red blossoms, and that the blossom on this plant has a color that lies at the opposite end of the visible spectrum from violet. Given this information (which serves to rule out the occurrence of a mutation affecting the color), the probability that the color of the blossom on this plant is red is unity. Therefore, the reference class in the statistical law in the original explanation was not maximally specific and the putative explanation fails to satisfy RMS.

The patent fault in this sort of use of RMS lies in the fact that knowledge that the color of the blossom is at the opposite end of the visible spectrum from violet is tantamount to knowledge that the color is red. In this example—as I intend it to be construed—the only basis of our knowledge that the color is at the opposite end of the visible spectrum from violet is our observation that the blossom is red. We have no way of knowing independently that the color is at the opposite end from violet. Whenever the blossom has a red color, it has a color at the opposite end of the spectrum from violet. Thus, to cite the fact that the color is at the opposite end from violet as an explanatory fact is to explain the red color on the basis of itself.

The best way to block this sort of difficulty is, it seems to me, to insist that the antecedent conditions in an I-S explanation must be temporally antecedent to the explanandum. Hempel's RMS might, therefore, be amended to read as follows:

> If s.k implies that b belongs to a class F_1, and that F_1 is a subclass of F, then s.k must also imply a statement specifying the statistical probability of G in F_1, say
>
> $P(G \mid F_1) = r_1$
>
> Here, r_1 must equal r unless the probability statement just cited is simply a theorem of mathematical probability theory, *or unless b's membership in F_1 cannot be known before its membership in G has been ascertained.*[12]

This version of RMS is obviously relativized to the state of knowledge of the explainer, but Hempel's versions all had that characteristic as well. I am inclined to think that this formulation adequately expresses the intent of Hempel's requirement.[13] It should be recalled, however, that Hempel explicitly rejected a similar temporal constraint on the 'antecedent' conditions in D-N explanations of particular facts.

To make good on the homogeneity requirements for the S-R model of scientific explanation, as remarked above, I attempted to explicate a fully objective concept of homogeneity (1977).[14] The general idea is that, if determinism is true, no explanations of particular facts will be irreducibly statistical; every statistical explanation can, in principle, be supplemented with additional information so as to be transformed into a D-N explanation. If indeterminism is true, some explanations will be irreducibly statistical—that is, they will be full-blooded explanations whose statistical character results not merely from limitations of our knowledge, but rather from the fact that there are no additional factors that would make it possible in principle to beef them up, thereby transforming them into D-N explanations. Whether determinism or indeterminism is true is a question of fact; I believe that the evidence points toward indeterminism, but that is not crucial for this discussion. Whichever is true, we should try to understand clearly the notion of an irreducibly statistical explanation. Speaking roughly and intuitively, given an irreducibly statistical explanation, it must be impossible in principle to identify anything that happens before the event-to-be-explained that would physically necessitate its occurrence.

These general ideas can be stated more precisely in terms of reference classes. Given an explanation that fits form (II), it will be irreducibly statistical if the term F that occurs in the statistical law premise "$P(G|F) = r$" designates a class that cannot be relevantly partitioned, even in principle, with respect to the attribute G. Suppose, that is, that we want to explain why b has the attribute G (at some particular time), and that we have appealed to the fact that b belongs to the class F to do so. If the explanation is to be irreducibly statistical, then prior to the fact that b has the attribute G, b cannot have any attribute in addition to F that has any bearing upon the fact that b has attribute G.

To try to make these ideas more precise, I introduced the concept of a *selection by an associated sequence*. It is, perhaps, best explained in terms of concrete examples. Consider, for instance, all of the basketball games played by the State College team. This class of events can be taken to constitute a reference class F; we are concerned, say, with the attribute G (games won by this team). The members of F (the games) can be taken in chronological order and designated x_1, x_2, . . . [15] Suppose just before each game there is a solo rendition of "The Fight Song" over the public-address system; call this class of events A, and let its members be taken in chronological order and designated y_1, y_2, ... , where y_i is the rendition of the song immediately preceding game x_i. A is an *associated sequence* for the reference class F, because there is a one-one correspondence (signified by the subscripts) between the members of F and those of A, and each y_i

in A precedes the corresponding x_i in F. Suppose that sometimes the song is sung by a male vocalist and sometimes by a female. We might wonder whether the sex of the vocalist has any bearing upon the chances of winning. So, designating the renditions by male vocalists B, we make a selection S of members of F on the basis of whether the corresponding member of A belongs to B. S consists of exactly those games that are preceded by a male's rendition of "The Fight Song." S is a *selection by an associated sequence*. Formally, x_i belongs to S iff y_i belongs to B. Someone seriously interested in the team's record might wonder whether the sex of the singer has any bearing on the team's success. If $P(G|F.S)$ differs from $P(G|F)$, the sex of the vocalist is relevant to winning. If so, the probability of winning is not invariant with respect to this particular selection by an associated sequence. Given this situation, we must acknowledge that the reference class F is not homogeneous with respect to winning, for we have just discovered a way to partition it relevantly.

Consider Greeno's example of Albert's theft of a car. The reference class F is the class of American teen-agers; let its members be arranged in some sequential order, x_1, x_2, ... The attribute G is commission of a car theft. In this case, we should note, the reference class consists not of events but of enduring objects. Nevertheless, the attribute is an event involving a member of the reference class. (The negation of the attribute would be to reach the age of 20 without having stolen a car) We might wonder whether coming from a 'broken home' is relevant to car theft by a teenager. To construct an associated sequence we can again take the class of American teen-agers, but, so to say, at earlier times in their lives. To make the case particularly clear, we might specify that the breakup of the home had to occur before the youth in question turned 13. What is crucial is that the selective attribute (coming from a broken home) occur earlier than the attribute of interest (theft of a car). Examples of this sort are most easily handled by reformulating them to make both the original reference class and the associated sequence consist of events (the breakup of a home, the theft of a car, becoming 20 without having stolen a car, etc.).

As another example, let F be a sequence of tosses of a given coin and let G be the coin landing heads up. We wonder whether the outcome of the preceding toss has a bearing on the outcome of this toss. We form the associated sequence A consisting of tosses of the same coin, where each toss x_i is associated with its immediate predecessor $y_i = x_{i-1}$. B is the attribute of landing heads up. The selection S thus picks out the tosses immediately following a toss resulting in heads. If this is a standard coin being tossed in the standard way, then this selection S will not furnish a relevant partition in F, for $P(G|F) = P(G|F.S)$. In this example, as in the former ones, it is crucial that the event used to make the selection occur before the event consisting of the occurrence or non-occurrence of the attribute of interest.

The general idea is that a reference class F is objectively homogeneous with respect to a given attribute G iff the probability of G in F is invariant with respect to all selections by associated sequences.[16] To rule out the counterexamples offered above in connection with Hempel's RMS, we have to ensure that the selection be made on the basis of events prior to those we are trying to explain. Thus, the case of the red blossoms, we do not even have an associated sequence, for the class F under consideration (flowers) constitutes the sequence each member of which has or lacks 'both' the color red and the color at the opposite end of the visible spectrum from violet, and each member has or lacks 'both' colors at the same time. In the case of Jones's quick recovery from his strep infection, we set up another sequence—TV news reports—but it does not qualify as an associated sequence because its members come after the corresponding members of the original reference class of strep infections. To qualify as an associated sequence, the members must precede the corresponding members of the original reference class. Consequently, although the report of Jones's recovery is statistically relevant to his quick recovery, it does not signify any lack of homogeneity in the original reference class.

The most difficult aspect of the explication of objective homogeneity is to impose suitable restrictions on the properties B that may be used to effect the selection S. Consider the probability that a person who contracts pneumonia will die as a result. There are several types of pneumonia—bacterial, viral, and fungal—and I presume that the probability of death of a victim varies from one type to another. Moreover, an individual may have more than one type at the same time; according to a recent report the combination of a bacterial and a viral infection is particularly lethal. Now, if we start with the general reference class of cases of pneumonia, we should partition in terms of the various types of infection, and, no doubt, in terms of various other factors such as the age and general physical condition of the patient, and the sort of treatment (e.g., penicillin) given. But we must not partition in terms of fatal vs. nonfatal infections. The reason is that very definition of the term "fatal" depends not just on the occurrence of an infection, but also on the occurrence or non-occurrence of a subsequent event, namely, death. Once you see how to do it in one or two examples, it is easy to cook up predicates whose applicability to a given event depends upon something that occurs (or fails to occur) at some later time.

The suggestion I offered for the restriction to be imposed is that B must determine an *objectively codefined class.* Very roughly, such a class is one consisting of events whose membership in the class can, in principle, be determined at the time the event occurs. More precisely, given that y_i is the member of an associated sequence corresponding to x_i, it must be possible in principle to set up a physical detecting apparatus (connected to a computer if necessary) that can deliver the verdict of whether y_i has the property B before it receives any information about the occurrence of x_i or any other event that happens later than x_i. My hope is that the resulting characterization of objective homogeneity is unrelativized either to particular knowledge situations or to particular languages (see W. Salmon 1984, chap. 3, for details). Whether the explication as given is satisfactory or not, it does not appear impossible in principle to define an adequate concept of objective homogeneity.[17]

The utility of the notion of objective homogeneity is not confined, of course, to the S-R model of explanation. If this concept is used in conjunction with Hempel's I-S model, the problem of essential epistemic relativization, as analyzed by Coffa, evaporates, for we can then say what constitutes a true I-S explanation. The fact that almost all of our actual I-S explanations are incomplete would not pose any fundamental conceptual problem. At the end of the third decade, Hempel (1977) acknowledged that there might be some nontrivial cases of objectively homogeneous reference classes. By implication he thereby admitted that the concept of objective homogeneity is not unintelligible.

3.3 Coffa's Dispositional Theory of Inductive Explanation

After articulating his devastating critique of the doctrine of essential epistemic relativization, Coffa offered a different solution to the problem of ambiguity of inductive explanation. According to Coffa, a great deal of the trouble Hempel encountered was a result of his implicit identification of the problem of ambiguity with the reference class problem. As we have seen, the reference class problem arises out of the frequency interpretation of probability when we try to apply probabilities to single cases. The situation could have been significantly improved, Coffa suggests, if Hempel had stuck with the propensity interpretation, which can be construed as a single case physical probability concept. In his doctoral dissertation Coffa (1973, chap. IV) argues that an appeal to the propensity interpretation of probability enables us to develop a theory of inductive explanation that is a straightforward generalization of deductive-nomological explanation, and that avoids both epistemic relativization and the reference class problem. This ingenious approach has, unfortunately, received no attention, for it was never extracted from the dissertation for publication elsewhere.

Coffa begins with a critical examination of the D-N model. Consider a deductive-nomological explanation of the simplest form:

(1) $(x)[Fx \supset Gx]$
Fa
$\overline{}$
Ga

for example—one cited by Carnap and Hempel—in answer to the question, "Why did this iron rod expand?" the explanation that it was heated and that "whenever a body is heated, it expands." This latter statement is, of course, the law required in the explanans – in this case, the law of thermal expansion.[18] One might suppose that this law should read

$$\Delta L = k \cdot L_o \cdot \Delta T$$

where L_o is the original length of the rod, k the coefficient of thermal expansion of the substance in question, and ΔT is the amount by which its temperature was raised. The foregoing explanation could presumably be construed as

(2) $\Delta L = k \cdot L_o \cdot \Delta T$
Rod r increased its temperature by ΔT.

Rod r increased its length by $k \cdot L_o \cdot \Delta T$.

This explanation is not acceptable as it stands, Coffa claims, for the statement of the law of thermal expansion is defective; a correct formulation requires what he calls an "extremal clause." It is not true, in general, that bodies expand when heated. The iron rod will not expand when heated if a sufficient compressing force is present. The correct statement is "*In the absence of other changes relevant to length*, an increase in temperature ΔT is physically sufficient for an increase in length $k \cdot L_o \cdot \Delta T$" (1973, 210). The italicized phrase is the extremal clause (EC). Only when the extremal clause is included do we have a bona fide law, and that is what is needed as the first premise of our explanation. Explanation (2) must be amended as follows:

(3) If EC then $\Delta L = k \cdot L_o \cdot \Delta T$
EC
Rod r increased its temperature by ΔT.

Rod r increased its length by $k \cdot L_o \cdot \Delta T$.

Because the first premise is a law of nature, change of temperature is *nomically relevant* to change of length. EC says that there is no other factor in the situation that is nomically relevant to the change of length. Assuming that EC is true, and that the initial condition stated in the third premise is true, Coffa claims, we now have a bona fide D-N explanation of the increase in length of the rod. Coffa's major thesis about D-N explanation is that *every law involves an extremal clause*, and consequently, *every D-N explanation must contain, as one of its premises, an assertion that the extremal clause holds on the occasion in question.*

An obvious objection can be raised at this point. If the extremal clause has to be incorporated into the law-statement, does that not make the law vacuous? Hempel and Oppenheim, we should recall, explicitly demanded that the law-statement in a D-N explanation have empirical content. The question is, does the extremal clause in the first premise rob the law of thermal expansion of its empirical content?

Coffa meets this objection squarely, citing an argument that had been given by Hempel to the effect that any law containing an extremal (or ceteris paribus) clause can always be protected from refutation – in all cases in which the law seems to fail – by maintaining that, because the extremal clause is not fulfilled, the law is actually correct. Coffa points out that any law in conditional form, even without the extremal clause, can be protected against refutation – in all cases in which it *seems* to fail – by maintaining that its antecedent was unfulfilled. Thus, any law having a conditional structure can be rendered vacuous by the same technique. He then observes that the pertinent question is not whether laws – with or without extremal clauses – *can* be treated in this way, but whether they *must* be so treated.

The answer, he argues, is negative. Consider, for example, Newton's second law of motion. It states that any body of mass m, upon which the net force is F, experiences an acceleration F/m. When this law makes reference to the *net force* acting on a given body, it is implicitly making a statement about all forces, known or unknown, detected or undetected. Nevertheless, we use Newton's laws of motion nonvacuously for predictive and explanatory purposes. In other words, we sometimes explicitly identify a set of forces acting upon a body and assert (corrigibly, of course) that there are no other forces. This latter assertion is an extremal clause.

Suppose Coffa is right in his claim that all D-N explanations involve extremal clauses. How does this fact apply to nondeductive explanations? It applies quite directly. We can think of universal laws as describing physical dispositions. Massive bodies have a disposition to accelerate when forces are applied; metal wires have a disposition to carry electrical current when a potential difference exists between the two ends. In addition to universal dispositions, the world also seems to contain probabilistic dispositions of varying strengths. A fair coin flipped by an unbiased device has a disposition of strength ½ to land heads up; a tritium atom has a disposition of strength ¾ to decay within a period of 25 years.

If we grant that there are both universal and probabilistic dispositions, then it is natural to think of both deductive and inductive explanation in terms of the following schema (1973, 273):

(4) $P(Fa|Ga) = r$
Ga
EC(a,F,G)

Fa

where "EC(a,F,G)" is an extremal clause stating that nothing nomically relevant to a's having property F, other than the fact that a has the property G, is present in this explanatory situation. In the case of inductive explanation, the extremal clause does the job of Hempel's requirement of maximal specificity and of my requirement of objective homogeneity. It makes inductive explanation just as objective and epistemically unrelativized as is deductive explanation. Before accepting Coffa's model I would, of course, insist upon a requirement banning inclusion of irrelevant factors, but that requirement is easily imposed.

Coffa's theory of scientific explanation adheres strongly to the ontic conception. Proponents of this conception can speak in either of two ways about the relationship between explanations and the world. First, one can say that explanations exist in the world. The explanation of some fact is whatever produced it or brought it about. The explanans consists of certain particular facts and lawful relationships. The explanandum is also some fact. This manner of speaking will sound strange to philosophers who have been strongly influenced by the thesis that explanations are arguments, or by the deeply linguistic approaches that regard explanations as speech acts. In nonphilosophical contexts, however, it seems entirely appropriate to say such things as that the gravitational attraction of the moon explains the tides, or the drop in temperature explains the bursting of the pipes. The gravitational attraction and the drop in temperature are out there in the physical world; they are neither linguistic entities (sentences) nor abstract entities (propositions). Second, the advocate of the ontic conception can say that an explanation is something – consisting of sentences or propositions – that reports such facts. It seems to me that either way of putting the ontic conception is acceptable; one can properly say either that the explanandum-fact is explained by the explanans-facts or that the explanans-statements explain the explanandum-statement. Coffa frequently adopts the first of these two manners of speaking; nevertheless, he does identify explanations as arguments, and that would seem to commit him, on those occasions, to the second.

The fact that Coffa often identifies the explanans with what produced or brought about the explanandum strongly suggests that his model is a causal model. He also adopts a covering law conception. The premise "$P(Fa|Ga) = r$" is a law-statement, so there must be a nomic connection between G and F, whether the law is universal ($r = 1$) or statistical ($r < 1$). The law-premise contains a symbol standing for probability, and, as I have mentioned, Coffa construes the probabilities as propensities. Coffa maintains that explanations should appeal only to nomically relevant factors – excluding those that are merely statistically relevant. He suggests, for example, that there may be no nomic relationship whatsoever between the sneeze of a Chinese mandarin at the time of one's birth and one's being a millionaire.

It would be natural, I should think, to restrict the laws admissible in such explanations to causal laws – where we admit probabilistic as well as deterministic causes. Propensities, I would suggest, are best understood as some sort of probabilistic causes. However, Coffa does not impose that restriction. According to Hempel's covering law approach, we can explain why a gas has a certain temperature by specifying the volume, pressure, and number of moles. Such an explanation may tell us nothing at all about how the temperature was brought about; consequently, for Coffa, I believe, it should not count as an explanation.

Several comments must be made regarding Coffa's proposed theory of inductive explanation.[19] In the first place, I have great sympathy with the idea that only nomic relevance should have a place in statistical explanation. In articulating the S-R model, I always maintained that it is a covering law model, all of the probability relations appearing in any such explanation being statistical laws. It is, of course, a difficult matter to distinguish statistical laws from true accidental statistical generalizations – at least as difficult as making the corresponding distinction in the case of universal generalizations, which is, as we have seen, an unsolved problem. But switching from statistical generalizations to propensities does not help in making this crucial distinction.

Second, the propensity interpretation does not escape the problem of the single case; indeed, it faces a problem that is the precise counterpart of the reference class problem. Basic to the single case propensity interpretation is the concept of a *chance set-up*. A chance set-up is a device that produces a set of alternative outcomes, each with a determinate probability. When a fair die is rolled from a dice cup in the standard manner, for example, there are six possible outcomes, and this chance set-up has a propensity of 1/6 to produce each of them on a given roll. A chance set-up can be operated repeatedly to produce a sequence of results. However, when we describe the chance set-up, we have to include relevant features and omit various irrelevant ones to determine what constitutes an additional trial on the same chance set-up. In the example of rolling the die, it is important that the die be physically symmetrical and that a standard cup be used, but it is irrelevant whether the device is operated during the day or during the night, or whether the roll is by a left-handed or right-handed player.

Consider a famous example – due to Laplace. Suppose that a coin is about to be flipped. It is known to be asymmetrical; it has a bias either for heads or for tails, but we do not know which. Laplace claimed, on the basis of his notorious *principle of indifference*, that, because of our ignorance, the probability of heads on this toss is ½. A modern propensity theorist would maintain that this chance set-up has a propensity other than ½; it is either greater than ½ or less than ½, but we do not know what it is. I would insist that either of the foregoing claims might be correct, depending upon how we specify the chance set-up. Suppose this coin has been produced by a machine that operates in the following way. Asymmetrical disks are fed into the machine in a random fashion, and the machine stamps the insignia on these disks. Half of the coins get heads on the favored side and half get tails on the favored side. A coin has been chosen randomly from the pile of coins produced by this machine, and it is about to be flipped. If we describe the operation of the chance set-up as picking a coin and flipping it in the standard manner, we do not know whether the propensity is equal to ½ or unequal to ½. If the continued operation of the chance set-up is to flip the same coin again in the same manner, then the propensity is different from ½. If the continued operation of the chance set-up is to pick another coin from the output of the machine and flip it just once, then the propensity of heads is ½. Even for the propensity interpretation, it is necessary to deal with a problem not essentially different from the reference class problem. I suggest, then, that switching to the propensity interpretation of probability does nothing to ease Hempel's problem of ambiguity.[20]

Third, as Paul Humphreys pointed out to me in conversation, the "propensity interpretation of probability" is not an admissible interpretation of the probability calculus at all (see Humphreys 1985). In fact, Coffa had already hinted at this point in his dissertation (1973, 272–73, note 32).[21] Bayes's theorem is one of the theorems of the mathematical calculus, and it provides a way of calculating 'inverse probabilities' from 'direct probabilities.' Imagine a factory that produces corkscrews. It has two machines, one old and one new, each of which makes a certain number per day. The output of each machine contains a certain percentage of defective corkscrews. Without undue strain, we can speak of the relative propensities of the two machines to produce corkscrews (each produces a certain proportion of the entire output of the factory), and of their propensities to produce defective corkscrews. If numerical values are given, we can calculate the propensity of this factory to produce defective corkscrews. So far, so good. Now, suppose an inspector picks one corkscrew from the day's output and finds it defective. Using Bayes's theorem we can calculate the *probability* that the defective corkscrew was produced by the new machine, but it would hardly be reasonable to speak of the *propensity* of that corkscrew to have been produced by the new machine. Propensities make sense as direct probabilities (because we think of them as probabilistic causes, I suspect), but not as inverse probabilities (because the causal direction is wrong).

Fourth, the concept of probabilistic causality is fraught with difficulties. When Coffa completed his dissertation in 1973 there were three serious theories of probabilistic causality in the literature—one by Hans Reichenbach (1956, chap. IV), another by I. J. Good (1960–61), and the third by Patrick Suppes (1970). These attempts at explication are far from simple. It is not just a matter of picking out a suitable subset of probabilities and identifying them as probabilistic causes. I think that each of the theories suffers from fundamental defects. In (1980) I published a critical study of all three and offered some suggestions of my own. We shall return to this problem in the fourth decade.

Finally, I am not completely convinced that laws of nature always involve extremal clauses, but I shall not attempt to argue that issue here. It is interesting to note that in 1982 a workshop on The Limitations of Deductivism was held at the University of Pittsburgh. Both Hempel and Coffa were participants. Hempel presented a paper in which he claimed that the use of scientific laws and theories, in conjunction with initial and boundary conditions, to make predictions cannot be construed as a strictly deductive affair. The basic reason is the need for proviso—qualifications that strongly resemble Coffa's extremal clauses. The proceedings of this workshop are published in (Grünbaum and Salmon 1988). Whether or not extremal clauses are always involved when universal laws are used, they certainly can be invoked in connection with explanations that employ statistical laws. It is evident that, in such cases, establishing the extremal clause is exactly the same as establishing the objective homogeneity of a reference class. Coffa's appeal to propensities and extremal clauses does not escape the immense difficulties associated with objective homogeneity.

At the midpoint of the third decade Coffa offered profound critiques of Hempel's I-S model and my S-R model. As an alternative he produced an extremely suggestive model of causal explanation that included both deterministic and indeterministic causes. The basic unity of his treatment of deductive and inductive explanation is an appealing aspect of his approach. It certainly contains strong anticipations of developments that occurred in the fourth decade. Nevertheless, enormous problems remained. Chief among them, I believe, is the need for a detailed treatment of the causal concepts to which we must appeal if we are to have any satisfactory causal model of scientific explanation.

Another of the many partisans of propensities in the third decade is James H. Fetzer. Along with Coffa, he deserves mention because of the central place he accords that concept in the theory of scientific explanation. Beginning in 1971, he published a series of papers dealing with the so-called propensity interpretation of probability and its bearing on problems of scientific explanation (Fetzer 1971, 1974, 1974a, 1975, 1976, 1977). However, because the mature version of his work on these issues is contained in his 1971 book, *Scientific Knowledge*, we shall deal with his views in the fourth decade. The fourth decade also includes his collaborative work with Donald Nute on probabilistic causality (Fetzer and Nute 1979).

3.4 Explanation and Evidence

One of the first philosophers to attack the Hempel-Oppenheim thesis of the symmetry of explanation and prediction was Nicholas Rescher (1958). At the very beginning of the second decade, he argued that, because of certain fundamental physical asymmetries in nature, many retrodictive inferences can be made with great certainty, whereas predictions are often much more hazardous. Because of a strong commitment to the inferential conception of scientific explanation he maintained that explanations are arguments. Because of a commitment to a rather robust high-probability requirement he claimed that there is a basic difference between explanatory arguments and predictive arguments—namely, that in explanatory arguments the premises must support the conclusion with certainty or near certainty, whereas in predictive arguments we often have to settle for much lower degrees of probability. For purposes of prediction we need only know that a given occurrence is more probable than not—or, in many cases, merely more probable than any other member of a reasonable set of alternatives. Rescher does not exclude the possibility that there are predictions—such as an astronomical prediction of a solar eclipse—that can be made with high degrees of certainty. His claim is simply that many predictive inferences do not have this character. He stresses his view that this kind of asymmetry is a fact of nature rather than a truth of logic.

Given Hempel and Oppenheim's failure to impose temporal constraints on scientific explanations, they leave open the possibility that the so-called *antecedent conditions* appearing in the explanans may actually occur after the explanandum. Recall our discussion of standard counterexample CE-1 (the eclipse) in §2.3 above. Nevertheless, it seems clear that Hempel-Oppenheim, in enunciating their symmetry thesis, did *not* mean to assert that prediction and retrodiction are equally reliable. They meant rather to say that an argument from *temporally prior* antecedent conditions and laws to a subsequent fact may serve in some contexts as a prediction and in others as an explanation. This is surely the way other philosophers, such as Scheffler, construed the symmetry thesis, and Hempel continued to construe it in that way in his "Aspects" essay. Similarly, I take it, as long as he leaves open the possibility of explanation in terms of subsequent conditions, he must say that there could be cases in which an argument from subsequent 'antecedent' conditions and laws would serve as a retrodiction in some circumstances and as an explanation in others. If explanations may sometimes be given in terms of subsequent conditions, then the symmetry condition would need to be formulated in terms not of prediction in the narrow temporal sense, but, rather, in terms of inference from the observed to the unobserved, whether that inference be predictive or retrodictive.

In an article published in the middle of the second decade, Rescher (1963) further elaborates his views on the relationships among prediction, retrodiction, and explanation. To expound his position, he introduces the notion of a *discrete state system*—i.e., a physical system that may assume different states out of a finite or infinite collection of discrete states. He shows how to construct systems in which both prediction and retrodiction are possible, those in which prediction is possible but retrodiction is impossible, those in which prediction is impossible but retrodiction is possible, and those in which both are impossible. The fact that systems of all of these types can be physically realized—by digital computers, for example—undermines the symmetry thesis as he construes it.[22]

Rescher's 1963 article constitutes one main kernel of his book *Scientific Explanation* (1970), which was published near the beginning of the third decade.[23] Another main kernel comes from an article on the nature of evidence (1958). Because Rescher is strongly committed to the main tenet of the received view—that explanations are arguments—he construes the explanatory relation as an evidentiary relation. Given this view, he continues to maintain that explanations in terms

of subsequent conditions are just as legitimate as explanations in terms of temporally antecedent conditions, provided a strong argument can be supplied.

In his 1953 book, we recall, Braithwaite raised certain issues about teleological explanation; he asked, in particular, whether explanations in terms of fulfillment of functions or realization of goals can be considered legitimate. He decided not to preclude them even though such explanations appeal to subsequent conditions. Rescher also raises "the much agitated issue of 'teleology' vs 'mechanism' in the theory of explanation," but he acknowledges that "[f]or purposes of the present discussion, the concept of 'teleology' (and its opposite) will be construed in a somewhat artificial chronologized sense" (Rescher 1970, 66). A teleological explanation in this sense is simply one in which the "antecedent conditions" (in the Hempel-Oppenheim meaning of the term) occur later than the event to be explained. Such explanations need not have anything to do with purposes, ends, or functions. Rescher's entire concern is to argue that putative explanations should not be rejected *solely* on the basis of the temporal relation between the explanatory facts and the fact-to-be-explained. In this regard, he supports Braithwaite's claim, but he does not try to give a theory of explanations that appeal to purposes, ends, or functions. He does not address the difficulties—raised by Hempel—regarding the logical structure of functional explanations.

One of Rescher's central concerns revolves around the nature of evidence. His treatment of this topic is considerably more sophisticated than Hempel's. According to the received view, as usually construed, the premises of a D-N explanation constitute conclusive evidence for its explanandum, and the premises of an I-S explanation provide strong evidence for its explanandum. The latter claim amounts to Hempel's high-probability requirement, and Rescher rejects it. He recognizes the need for some sort of relevance requirement in connection with the concept of evidence, and he embodies it in his theory of evidence. In developing this theory he introduces the notions of evidential presumption, supporting evidence, and confirming evidence (1970, 76–77). Briefly, presumptive evidence renders a given hypothesis more probable than its negation; this is a very weak high-probability concept (1970, 78–79). Supporting evidence is a relevance concept; supporting evidence renders the hypothesis more probable in the presence of that evidence than it was in its absence (1970, 80–83). The concept of confirming evidence incorporates these two notions:

> The statement p is *confirming evidence* for q if (1) p is a presumptive factor for q, and (2) p is supporting evidence for q. Thus confirming evidence must at once render its hypothesis more likely than before *and* more likely than not. This two-pronged concept of confirming evidence perhaps most closely approximates to the idea represented in the common usage of "evidence." (1970, 84)

It is remarkable that—to the best of my knowledge—no one (including Rescher and me) noticed any connection between what he was doing in his evidential approach to probabilistic explanation and what those of us who were concerned with statistical relevance were up to. It may have been due, in part at least, to the fact that his investigation was embedded in a theory of evidence, which is naturally taken to be part of confirmation theory, while the S-R model was developed within the theory of scientific explanation. Such compartmentalized thinking seems, in retrospect, quite shocking.

Rescher reflects another aspect of the received view by maintaining the covering law conception of scientific explanation. As he says, "Scientific explanations are invariably subsumption arguments that cite facts to establish an explanandum as a special case within the scope of lawful generalizations" (1970, 9). This leads him to undertake an extended analysis of the concept of lawfulness (1970, 97–121). He takes it as a point of departure—a point acknowledged by most authors who deal with this issue—that lawful generalizations have counterfactual and modal import. Whereas accidental generalizations state only what happens under actual circumstances, laws state what would happen under circumstances that are merely possible, i.e., under counterfactual circumstances. He then argues that empirical evidence can support only the general claim about actual cases, and, consequently, the assertion of lawfulness regarding a generalization goes beyond the empirical evidence. From this consideration he does not conclude that we have super-empirical *knowledge* of lawfulness; instead, he argues, we *impute* lawfulness on the basis of epistemic considerations. Lawfulness is not something that we discover about generalizations; it is something we supply (1970, 107). Lawfulness is mind-dependent (1970, 113–14; see also Rescher 1969). It reflects the use to which we want to put a generalization in organizing and systematizing our knowledge of the world (1970, 107). Skyrms (1980) and van Fraassen (1980), exactly a decade later, offer similarly pragmatic accounts of lawfulness.

In his discussion of the role of laws in scientific explanation, and the modal import of laws, Rescher asserts,

> A recourse to laws is indispensable for scientific explanation, because this feature of nomic necessity makes it possible for scientific explanations to achieve their task of showing not just *what* is the case, but *why* it is the case. This is achieved by deploying laws to narrow the range of possible alternatives so as to show that the fact to be explained "had to" be as it is, in an appropriate sense of this term. (1970, 13–14)

Inasmuch as this passage comes at the conclusion of a discussion of the role of universal *and probabilistic* laws in explanation, it appears that Rescher wants to attribute some kind of modal import to probabilistic laws. He does not say what it is. As we shall see, D. H. Mellor, writing shortly before the close of the third decade, also advocates a modal interpretation of probabilistic explanation (1976).

I shall explain in greater detail below why I find the notion of *degrees of necessity* quite mystifying.

In spite of his strong adherence to the thesis that explanations are subsumptive arguments, Rescher does suggest, from time to time, that scientific explanation involves an exhibition of the mechanisms by which nature works. He makes this point, for example, in connection with the explanatory power of Darwin's theory (1970, 14–15). This discussion suggests that subsumption under nomic regularities is not really all there is to scientific explanation after all; we need to appeal to the underlying mechanisms as well.

In his discussion of discrete state systems, it will be recalled, Rescher provides a variety of types in which prediction and/or retrodiction, and hence, on his view, explanation is impossible. He makes the profound observation that, in the most fundamental sense, we may nevertheless understand completely what goes on in these systems—we know how they work. The issue arises in the context of statistical explanation:

> The consideration of stochastic systems forces us to the realization that scientific understanding can be present despite an impotence to explain (predict, etc.) *even in principle* certain particular occurrences. As regards explanation, this leads us to recognize the fundamentality of description, being able to deploy laws so as to describe the modes of functioning of natural systems. . . .
>
> The root task of science ought thus to be thought of as a fundamentally *descriptive* one: the search for the laws that delineate the functioning of natural processes. . . . It is thus our grasp of natural laws, be they universal or probabilistic—not our capacity to explain, predict, and so on—that appears to be the basic thing in scientific understanding. For it is undeniable that a knowledge of the pertinent laws goes a long way with endowing us with all that we can possibly ask for in the way of an understanding of "the way in which things work"—even in those cases where the specific desiderata of explanation, prediction, etc., are, in the very nature of things, beyond our reach. (1970, 133–34)

Although it is not Rescher's intention to deny the possibility of probabilistic explanation, the foregoing passage does suggest one line of argument that has been adopted by such current deductivists as John Watkins (1984) and Kitcher (Kitcher & Salmon 1989). They maintain that we can provide deductive explanations of lawful statistical regularities by appealing to fundamental statistical laws, but that we cannot give deductive explanations of particular occurrences. Nevertheless, they can maintain, that does not matter, for our knowledge of the lawful statistical regularities provides understanding of how the world works. We have no need, they argue, for any nondeductive type of explanation.

As I shall explain in discussing the opening of the fourth decade, there are at least three main conceptions of scientific explanation: the *epistemic*, the *modal*, and the *ontic*. Although they have been present since the time of Aristotle, they had not been clearly distinguished. After the advent of models of probabilistic or statistical explanation it became crucial to distinguish them. In the foregoing discussion of Rescher's account we have noticed all three.

3.5 Explanations of Laws

Early in our story we called attention to Hempel and Oppenheim's "notorious footnote 33," in which they pointed to a basic difficulty in connection with explanations of laws. The problem is to distinguish the cases in which we genuinely explain a lawful regularity by deductive subsumption under broader regularities from such nonexplanatory moves as deducing a law from a conjunction of itself with another totally unrelated law. Because they saw no solution to this problem, they refrained from even attempting to provide a D-N account of explanations of laws. To the best of my knowledge this problem was not seriously attacked until more than a quarter of a century later—when Michael Friedman (1974) finally took it on. The delay is especially surprising given the fact that, as Friedman notes, explanations of regularities are more usual than explanations of particular facts in the physical sciences.[24]

The basic theme of Friedman's paper is that science explains various phenomena (where phenomena are general regularities) through unification. Citing the kinetic theory of gases as an example, he says,

> This theory explains phenomena involving the behavior of gases, such as the fact that gases approximately obey the Boyle-Charles law, by reference to the behavior of molecules of which gases are composed. For example, we can deduce that any collection of molecules of the sort that gases are, which obeys the laws of mechanics will also approximately obey the Boyle-Charles law. How does this make us understand the behavior of gases? I submit that if this were all the kinetic theory did we would have added nothing to our understanding. We would have simply replaced one brute fact with another. But this is not all the kinetic theory does—it also permits us to derive other phenomena involving the behavior of gases, such as the fact that they obey Graham's law of diffusion and (within certain limits) that they have the specific heat capacities that they do have, from the laws of mechanics. The kinetic theory effects a significant *unification* in what we have to accept. Where we once had three independent brute facts—that gases approximately obey the Boyle-Charles law, that they obey Graham's law, and that they have the specific heat capacities they do have—we now have only one—that molecules obey the laws of mechanics. Furthermore, the kinetic theory also allows us to integrate the behavior of gases with other phenomena, such as the motions of the planets and of falling bodies near the earth. This is because the laws of mechanics also permit us to derive both the fact that planets obey Kepler's laws and the fact that falling bodies obey Galileo's laws. From the fact that *all* bodies obey the laws of mechanics it follows that the planets behave as they do, falling bodies behave as they do, and gases behave as they do. Once again, we have reduced a multiplicity of unexplained, independent phenomena to one. I claim that this is the crucial property of scientific theories we are looking for; this is the essence of scientific explanation—science increases our understanding of the world by reducing the total number of independent phenomena that we have to accept as ultimate or given. A world with fewer independent phenomena is, other things equal, more comprehensible than one with more. (1974, 14–15)

It is evident that, if Friedman is successfully to carry out his program, it will be crucial to provide a method for counting independently acceptable phenomena. By "phenomena" in this context we shall understand general uniformities; they can be represented by lawlike sentences. Since *acceptance* is a central notion, we presume that "at any given time there is a set *K* of *accepted* lawlike sentences, a set of laws accepted by the scientific community. Furthermore, . . . the set *K* is deductively closed . . . *K* contains all lawlike consequences of members of *K*" (1974, 15). Recalling the Hempel-Oppenheim footnote, we immediately ask whether we can reduce the number of phenomena by, for example, deducing the Boyle-Charles law and Graham's law from the conjunction of these two laws, thereby reducing the number from two to one. Clearly, we do not want this kind of move to count as a reduction. Friedman suggests that an appeal to the notion of *independently acceptable lawlike sentences* will enable us to deal with that problem.

Unfortunately, Friedman does not furnish any precise characterization of the of what it means for one sentence to be acceptable independently of another.

> Presumably, it means something like: there are sufficient grounds for accepting one which are not also sufficient grounds for accepting the other. If this is correct, the notion of independent acceptability satisfies the following conditions:
>
> (1) If S $\vdash$ Q, then S is not acceptable independently of Q.
> (2) If S is acceptable independently of P and Q $\vdash$ P, then S is acceptable independently of Q.
>
> (assuming that sufficient grounds for accepting S are also sufficient for accepting any consequence of S). (1974, 16–17)

The rationale for condition (1) is evident. By virtue of the entailment, grounds for accepting S simply *are* grounds for accepting Q. However, since S is, in many cases, a stronger statement than Q, grounds for accepting Q may well be insufficient grounds for accepting S.

The rationale for condition (2) is not quite as easy to see. Perhaps it can be made clearer by looking at a trivial reformulation:

(2′) Given $Q \vdash P$, if S is not acceptable independently of Q then S is not acceptable independently of P.

Suppose that S were not acceptable independently of Q. Then grounds for accepting S would be grounds for accepting Q. Because Q entails P, however, grounds for accepting Q would ipso facto be grounds for accepting P. Hence, grounds for accepting S would be grounds for P, and S would not be acceptable independently of P. To look at the condition a bit more formally, let "$E \models X$" mean "E constitutes sufficient grounds for accepting X." Then, given $Q \vdash P$, we have

$$E \models Q \rightarrow E \models P$$
$$E \models S \rightarrow E \models Q$$

from which it follows that

$$E \models S \rightarrow E \models P.$$

This obviously establishes Friedman's second condition in the form (2′).

An indispensable requirement for the implementation of this program is to develop a method for counting phenomena (as represented as lawlike sentences). To get on with the job, Friedman introduces the concept of a *partition* of a sentence S. A partition of S is a set Z of sentences such that Z is logically equivalent to S and each S′ in Z is acceptable independently of S. Note that the members of the partition do not have to be mutually exclusive. Using this definition, he defines a *K-atomic sentence S* as one that has no partition (in K). This means that there is no set of sentences $\{S_1, S_2\}$ in K such that S_1 and S_2 are acceptable independently of S and S_1 & S_2 is logically equivalent to S (1974, 17). If, for example, we took the conjunction of the Boyle-Charles law (S_1) and Graham's law (S_2), it would not be K-atomic, for $\{S_1, S_2\}$ is a partition of it. The question is whether there can be any K-atomic sentences.

Consider any sentence in K of the form

$$(x)[Fx \supset Gx] \quad (1)$$

which is equivalent to

$$(x)[Fx \,.\, (Hx \vee \sim Hx) \supset Gx] \quad (2)$$

and to

$$\{(x)[Fx \,.\, Hx \supset Gx], (x)[Fx \,.\, \sim Hx \supset Gx]\} \quad (3)$$

If (3) is a partition of (1), then (1) is not K-atomic. But (3) is a partition of (1) unless at least one member of (3) is not a member of K or at least one member of (3) is not acceptable independently of (1). Because (1) entails each member of (3), each member of (3) must belong to K unless it fails to be lawlike.

In private correspondence I posed the foregoing problem to Philip Kitcher (who has written incisively about Friedman's proposal (1976)) as well as to Friedman. Kitcher's response went something like this: If "H" is a reasonable predicate (signifying something like a natural kind), then "~H" is apt not to be a reasonable predicate and the second member of (3) may fail to be lawlike.

I would be inclined to reply to Kitcher as follows: One could argue (along the lines of Hempel and Oppenheim as discussed above) that if the first member of (3) is lawlike, so will be the second. If the first member has unlimited scope so has the second; if the first makes no reference to any particular entity, neither does the second; if the first contains only purely qualitative predicates so does the second; etc.[25] Leaving all that aside, however, it is plausible to suppose that a similar partition could be based upon an exclusive and exhaustive disjunction H_1, H_2, . . . , such that, for each i,

$$(x)[Fx \,.\, H_ix \supset Gx] \quad (4)$$

is a lawlike sentence. Since each sentence of the form (4) is a logical consequence of (1), each is a member of K. In that case, (4) provides a partition of (1).[26]

Here is Friedman's reply to my original problem:

> The problem you raise . . . is just the *kind* of problem I had in mind in requiring that a partition of a sentence S consist of sentences that are acceptable independently of S. In your case, it would seem in general that our ground for accepting the two conjuncts is just that they follow from 'All F are G.' If, on the other hand, we have grounds for accepting the two conjuncts independently—by testing for the two conjuncts directly, say—then it would seem that 'All F are G' is in no way an explanation of the two, but just a summary of what we already know.

I am not satisfied that Friedman's answer is adequate. Consider the generalization (in K), "All humans are mortal." Is it K-atomic? K contains "All men (i.e., human males) are mortal" and "All women (i.e., human females) are mortal." Suppose, as is plausible, that ancient Greek men were male chauvinist pigs. They noticed that each man seemed to die, so by simple induction they concluded that all men are mortal. At this point they could accept "All men are mortal" independently of "All humans are mortal." Some male then made the further generalization that all humans are mortal and it was accepted by the scientific community. Now, I take it, "All men are mortal" is still acceptable independently of "All humans are mortal," but "All women are mortal" is not, for the only basis for asserting this latter statement is that it follows from "All humans are mortal." Thus, {"All men are mortal"; "All women are mortal"} does not constitute a partition of "All humans are mortal," since the second member is not acceptable independently of it. Perhaps this general sentence, "All humans are mortal," is K-atomic.[27]

Suppose that, without changing the membership of the class K in any way, one or more of these MCP Greeks starts noticing that women also die sooner or later, thus accumulating inductive evidence sufficient to accept "All women are mortal" on that basis. Now it is possible to accept that sentence independently of "All humans are mortal," with the result that this latter sentence is not K-atomic, since it now has a partition.

Judging from remarks earlier in his essay, Friedman (1974, 14) – as well as Hempel and Oppenheim – is eager to develop an objective account of scientific explanation. In view of this consideration Friedman would, I should think, regard the order in which the different sorts of evidence for a sentence are actually acquired as irrelevant to the K-atomicity of that sentence. The foregoing example seems to violate this kind of objectivity.

Consider a more serious (quasi-historical) example. Newton's law of universal gravitation states that between any pair of material objects there exists a certain sort of attractive force. Is this a K-atomic sentence? By the time his *Principia* was published, there was considerable evidence to support this law in the domain of pairs of material bodies in which each member of the pair is a large astronomical body (earth, moon, planet, sun, etc.) and in the domain in which one member of the pair is a large astronomical body (earth) and the other a much smaller object (apple, stone, etc.). However, the only evidence for the claim that it also holds in the domain of pairs of which both members are small is the fact that it follows from the general law. Hence, the foregoing tripartite division of the set of all pairs of material bodies {both large, one large and one small, both small} could not have been used at the end of the seventeenth century to provide a K-partition of Newton's law of universal gravitation. However, after the Cavendish torsion balance experiment was performed in the following century there was independent evidence for Newton's law in the third domain of pairs. We could apparently then use as a K-partition: {"Given any two extremely massive material bodies, there exists an attractive force between them . . . "; "Given any two bodies one of which is extremely massive and the other is not, there exists a force . . . "; "Given any two material bodies, neither of which is extremely massive, there exists a force . . . "}. The result is that Newton's law is demonstrably not K-atomic.[28]

If a sentence that was K-atomic (with respect to a given class K of lawlike sentences) becomes non-K-atomic (with respect to the same class), that fact has a profound bearing upon its explanatory import. As we shall see, Friedman offered two definitions of scientific explanation. As we shall also see, Kitcher proved that, under Friedman's first definition of explanation, only K-atomic sentences explain.[29] Consequently, we find a strong presumption that the performance of the Cavendish experiment (along with others, perhaps) robs Newton's law of gravitation of its explanatory import. My personal intuition is that the explanatory power of Newton's law was enhanced as a result of such experiments.

Be that as it may, I must return to my original question, namely, on Friedman's characterization, can there be any K-atomic sentences. I have suggested that among simple generalizations of the form "All As are Bs" there cannot be any. I have not tried to extend the argument to lawlike generalizations of all sorts, but I should think that, if there are any K-atomic sentences, some of them would have this simple form. *If there are no K-atomic sentences, Friedman's program of characterizing scientific explanation in terms of the reduction of the number of independently accepted laws cannot get off the ground.*

It strikes me that Friedman's resolution of the problem is excessively pragmatic for his purposes.[30] If *there exists* evidence sufficient for the acceptance of a generalization having form (1) above, then *there exists* evidence for the independent acceptance of both of the generalizations in the partition (3) above. Which pieces of evidence are found first and which lawlike generalizations are formulated first should be irrelevant if our aim is to formulate an *objective* theory of scientific explanation.

While the difficulties in characterizing K-atomic statements seem serious, I do not mean to suggest that they are insuperable. Let us therefore try to see what Friedman has accomplished if the concept of K-atomic sentence turns out to be viable. The answer, I believe, is that he has solved the problem originally stated in the famous Hempel-Oppenheim footnote. Without going into all of the technical details of his explication, I shall attempt to sketch the development.

Given the notion of a K-atomic sentence, the *K-partition* of a set of sentences Y can be defined as a set Z of K-atomic sentences such that $Y \leftrightarrow Z$. In general there may be more than one K-partition of a given set Y. The *K-cardinality* of a class Y can be taken as the number of members of the smallest K-partition of Y (i.e., the greatest lower bound of the cardinality of K-partitions of Y). A sentence S *reduces* a class Y if the K-cardinality of the union of {S} with Y is smaller than the K-cardinality of Y. Friedman then proceeds to define explanation:

> How can we define *explanation* in terms of these ideas? If S is a candidate for explaining some S′ in K, we want to know whether S permits a reduction in the number of independent sentences. I think that the relevant set we want S to reduce is the set of *independently acceptable* consequences of S ($con_K(S)$). For instance, Newton's laws are a good candidate for explaining Boyle's law, say, because Newton's laws reduce the set of their independently acceptable consequences – the set containing Boyle's law, Graham's law, etc. On the other hand, the conjunction of Boyle's law and Graham's law is not a good candidate, since it does not reduce the set of its independently acceptable consequences. This suggests the following definition of explanation between laws:
>
> (D1) S_1 explains S_2 iff $S_2 \in con_K(S_1)$ and S_1 reduces $con_K(S_1)$..
>
> (1974, 17)

Friedman is not satisfied with this definition, for he finds it too restrictive.

> Actually this definition seems to me to be too strong; for if S_1 explains S_2 and S_3 is some independently acceptable law, then S_1 & S_3 will not explain S_2—since S_1 & S_3 will not reduce $\text{con}_K(S_1 \& S_3)$. This seems undesirable—why should the conjunction of a completely irrelevant law to a good explanation destroy its explanatory power? So I will weaken (D1) to
>
> (D1′) S_1 explains S_2 iff there exists a partition Z of S_1 and an $S_i \in Z$ such that $S_2 \in \text{con}_K(S_i)$ and S_i reduces $\text{con}_K(S_i)$.
>
> (1974, 17)

As Kitcher points out (1976, 211), (D1′) is not just a liberalization of (D1). Having demonstrated that, according to (D1), *only* K-atomic sentences can explain, he points to the obvious fact that, according to (D1′), *no* K-atomic sentences can explain. (D1) and (D1′) are simply different sorts of definition. Kitcher suggests that Friedman's intent might best be captured by a definition whose definiendum consisted of the disjunction of the right-hand sides of (D1) and (D1′). My response to this situation is different. Having already canvassed the many difficulties that arise if we allow explanations to contain irrelevancies, I believe (D1′) should simply be rejected.

If we continue to assume that the problem of K-atomic sentences can be solved or circumvented, then Friedman's theory provides a way for the received view to fill the upper right-hand sector of Table 1—the D-N explanation of universal laws. It is not clear whether Friedman's approach could be extended to handle the lower right-hand sector—what Hempel called D-S explanation. Since I have maintained above that D-S explanation should be regarded as a species of D-N explanation, the hope would be that Friedman's approach could handle this sector as well. As the end of the third decade approached, the received view had theories of D-N explanation of particular facts, I-S explanation of particular facts, D-N explanation of universal generalizations, and hopes for a theory of D-N explanation of statistical generalizations. Although each of these models suffered from severe difficulties, the received view was at least approaching a complete articulation.

Assuming that Friedman's account of D-N explanation of laws could handle the problem of K-atomic sentences, it appears to be just what Hempel and Oppenheim were looking for in answer to the problem stated in their "notorious footnote 33." However, as Kitcher showed in his critique of Friedman's paper, that theory of explanation of laws is not what we should have wanted, for it is subject to some severe shortcomings.

Kitcher claims that Friedman's definition (D1) is vulnerable to two sorts of counterexamples.

> Counterexamples of the first type occur when we have independently acceptable laws which (intuitively) belong to the same theory and which can be put together in genuine explanations. The explanantia that result are not K-atomic and hence fail to meet the necessary condition derived from Friedman's theory.
>
> Consider, for example, the usual derivation of the law of adiabatic expansion of an ideal gas, given in books on classical thermodynamics. The explanans here is the conjunction of the Boyle-Charles law and the first law of thermodynamics. These laws are acceptable on the basis of quite independent tests, so their conjunction is not K-atomic. However, the derivation of the law of adiabatic expansion from the conjunction is, intuitively, a genuine explanation. (1976, 209–10)

Kitcher's second type of counterexample concerns the explanation of complex phenomena where, typically, many different laws or theories are brought to bear. He cites, as one instance, the explanation of why lightning flashes are followed by thunderclaps. "The explanation utilizes laws of electricity, thermodynamics, and acoustics, which are independently acceptable."[31]

Kitcher goes on to show—convincingly, in my opinion—that an appeal to Friedman's other definition (D1′) does not help in dealing with the foregoing sorts of counterexamples. He concludes, therefore, that Friedman's account of explanation by unification is unsatisfactory. Nevertheless, Kitcher is firmly committed to the general thesis (with which I am strongly inclined to agree) that unification is a fundamental goal of explanation—that unification yields genuine scientific understanding. In the fifth decade Kitcher articulates his version of the unification theory in Kitcher and Salmon (1989).

3.6 Are Explanations Arguments?

Early in the third decade Jeffrey (1969) argued persuasively that, in many cases at least, statistical explanations are not arguments. His article as well as his thesis was incorporated into the publication in which the S-R model was first set forth in detail (W. Salmon et al. 1971). By the end of the third decade (1977c) I had labeled the claim that scientific explanations are arguments a *third dogma of empiricism* and urged its wholesale rejection.[32] Since the doctrine that scientific explanations are arguments—deductive or inductive—is absolutely central to the received view, my intent was to undermine the orthodox position as deeply as possible. In my zeal to rebut the claim that all explanations are arguments, I argued that no explanations are arguments. This view now seems too extreme; as it seems to me now, some are and some are not.[33] The challenge took the form of the following three questions I hoped would prove embarrassing to devotees of the received view. What I failed to notice until recently is that, while these questions *are* acutely embarrassing with regard to explanations of particular facts, they are innocuous for explanations of laws.

QUESTION 1. Why are irrelevancies harmless to arguments but fatal to explanations?

In deductive logic, irrelevant premises are pointless, but they have no effect whatever on the validity of the argument. Even in the Anderson-Belnap relevance logic, p & q ⊢ p is a valid schema. Consider the following variation on a time-honored argument:

All men are mortal.
Socrates is a man.
Xantippe is a woman.

Socrates is mortal.

It is strange, somewhat inelegant, and possibly, mildly amusing, but it is obviously valid. In contrast, as we have seen, the appearance of an irrelevancy in a D-N explanation can be disastrous.

> The rooster who explains the rising of the sun on the basis of his regular crowing is guilty of more than a minor logical inelegancy. So also is the person who explains the dissolving of a piece of sugar by citing the fact that the liquid in which it dissolved is *holy* water. So also is the man who explains his failure to become pregnant by noting that he has faithfully consumed birth control pills. (1977c, 150)

The situation is no different when we consider inductive relations. There is a *requirement of total evidence* that requires inductive inferences to contain all relevant premises, but it does not ensure against the inclusion of irrelevant premises. When a detective attempting to establish the identity of a murderer comes across evidence that may or may not be relevant—i.e., about whose relevance he or she cannot as yet be sure—it behooves him or her to include it. If it is irrelevant it can do no harm; an irrelevant premise simply does not change the probability of the conclusion. If it turns out to be relevant it may be very helpful.

Where inductive explanation is concerned, again the situation is entirely different. If psychotherapy is irrelevant to the remission of neurotic symptoms, it should not be invoked to explain a patient's psychological improvement. If massive doses of vitamin C are irrelevant to the rapidity of recovery from a common cold, such medication has no legitimate place in the explanation of a quick recovery.

> Inference, whether inductive or deductive, demands a requirement of total evidence—a requirement that *all* relevant evidence be mentioned in the premises. This requirement, which has substantive importance for inductive inferences, is automatically satisfied for deductive inferences. Explanation, in contrast, seems to demand a further requirement—namely, that *only* considerations relevant to the explanandum be contained in the explanans. This, it seems to me, constitutes a deep difference between explanations and arguments. (1977c, 151)

The second query comes in two forms that are so closely related as to constitute one question:

QUESTION 2: Can events whose probabilities are low be explained?
QUESTION 2′: Is genuine scientific explanation possible if indeterminism is true?

The close relationship between these two questions rests upon a principle of symmetry that we have discussed above. Suppose, in a genuinely indeterministic situation, there are two possible outcomes, one highly probable, the other quite improbable. Jeffrey argued, and I fully agree, that in such circumstances, when we understand that both are results of the same stochastic process, we understand the improbable outcome (when it occurs) just as well as we understand the probable outcome (when it occurs). But if probabilistic explanations are inductive arguments, an undesirable asymmetry arises. The explanans confers upon the probable explanandum a high inductive probability, so the explanation is a strong inductive argument. In the case of the improbable outcome, the explanans provides a strong inductive argument for the *non-occurrence* of the event to be explained. The fact that some events have low probabilities seems to cast doubt upon all statistical explanations of individual events if explanations are viewed as being, in essence, arguments.

QUESTION 3: Why should requirements of temporal asymmetry be imposed upon explanations (while arguments are not subject to the same constraints)?

Among the earliest counterexamples brought against the D-N model of Hempel and Oppenheim was (CE-1), the case of the eclipse. We noted that the occurrence of a given eclipse could be deductively inferred, on the basis of the same laws, either from conditions prior to the eclipse or from conditions subsequent to the eclipse. Only one of these inferences—that from prior conditions—might possibly count as an explanation. A similar point can be made in terms of (CE-2), the flagpole example. Given the angle of elevation of the sun in the sky, the length of the shadow can be inferred from the height of the flagpole, or the height of the flagpole can be inferred from the length of the shadow. The height of the flagpole explains the length of the shadow because the interaction between the sunlight and the flagpole occurs before the interaction between the sunlight and the ground.[34] The length of the shadow does not explain the height of the flagpole because the temporal relation is wrong.

As Rescher had already pointed out, it would be quite mistaken to suppose that inference is in general temporally symmetrical. In a wide variety of cases it is possible to make quite reliable inferences from subsequent facts to earlier occurrences. It is often difficult to predict whether precipitation will occur at a particular place and time; it is easy to look at the newspaper to determine whether rain or snow occurred there yesterday. Dendrochronologists can infer with considerable precision and reliability from the examination of tree rings the relative annual rainfall in certain areas for thousands of years in the past. No one can predict the relative annual rainfall with any reliability for even a decade into the future. The situation is obvious. We can have records—human or natural—of things that have happened in the past. We do not have records of the future, or any comparable resource for making predictions.

We see, then, that explanation involves a temporal asymmetry, and that inference in many cases also involves a temporal asymmetry. What is striking is that the two asymmetries are counterdirected. This would be very strange indeed if explanations are, in essence, arguments. The reason for the contrast is easy to fathom. We explain effects in terms of their causes, but in many cases we infer causes from their effects. Recognizing the crucial role of causal considerations in relation to the third question, I declared that the time had come to put "cause" back into "because" (1977c, 160).

Proponents of the received view had various possible answers to these questions. In response to the first question they could simply impose some sort of requirement that would block irrelevancies. No one (to the best of my knowledge) ever claimed that *all* logically correct arguments with true premises are explanations. As we saw, Hempel and Oppenheim worked hard to characterize the kinds of valid deductive arguments that can qualify as explanations. Advocates of the received view could simply admit that a further requirement, excluding irrelevancies, should be added to the D-N model.[35] Similarly, there are various ways of dealing with irrelevancies in the theory of I-S explanations. A simple way, as I remarked above, would be to amend the requirement of maximal specificity to make it into the requirement of *the maximal class of maximal specificity*.[36]

Several sorts of response to the second question are available to champions of the received view. In the first place, one could simply deny the symmetry principle offered by Jeffrey et al., insisting that we can explain highly probable occurrences, but not those with smaller probabilities. D. H. Mellor (1976) has defended this position. I must reiterate my strong intuition that this response perpetuates a highly arbitrary feature of Hempel's I-S model.

In the second place, one could deny that there are such things as probabilistic explanations of particular facts, insisting that all legitimate explanations are of the D-N variety (perhaps including the D-S type). For reasons that appear to be somewhat different from one another, Wolfgang Stegmüller (1973) and G. H. von Wright (1971) take this tack. I worry about whether this position reflects an anachronistic hankering after determinism, but I shall discuss this thesis in detail in connection with developments occurring in the fourth decade, especially §4.9.

It may turn out, in the end, to be the most promising way to avoid a host of problems associated with statistical explanation.

In the third place, Hempel (private correspondence) suggested the possibility of getting around the problem posed in the second question by reconstruing the concept of an inductive argument. During the late 1940s Carnap developed a detailed and rigorous system of inductive logic that was published in his monumental work *Logical Foundations of Probability* (1950, 2nd ed., 1962). Hempel was strongly influenced by Carnap's theory. One of the peculiarities of Carnap's system is that it has no place for *rules of acceptance* in inductive logic. In deductive logic we think of an argument as a group of statements consisting of a set of premises and a conclusion. If the argument is valid, then, if we accept the premises as true (or well-founded) we should be willing to accept the conclusion as well. It is natural to think of inductive arguments in an analogous way. Under this conception, an inductive argument would be a set of statements consisting of premises and a conclusion. If the argument has a correct inductive form, then, if we accept the premises as true (or well-founded) *and if they comprise all available evidence relevant to the conclusion* we should be willing to accept the conclusion as well. The schemas Hempel has offered in his various discussions of D-N and I-S explanation strongly suggest just such an analogy. In Carnap's inductive logic no such analogy holds. His denial of rules of acceptance for inductive logic means that *there are no inductive arguments* in this straightforward sense. Inductive logic furnishes only degree of confirmation statements of the form "c(h,e) = r" where the hypothesis h may be any statement, evidence e any consistent statement, and r any real number between 0 and 1 inclusive. Inductive logic tells us how to compute r. If r has been calculated correctly, it is the degree of confirmation of hypothesis h on evidence e. If e comprises all available evidence relevant to h, then r may be interpreted as a fair betting quotient; that is, r:1 - r constitute fair odds for a bet on the truth of h. But no matter how close r is to 1, we are not permitted to extract the hypothesis h from the degree of confirmation statement and assert it separately as a statement we accept. There are no arguments of the form

$$\begin{array}{l} c(h,e) = r \\ e \\ \text{e contains all relevant evidence} \\ \hline\hline h \end{array} \quad [r]$$

in Carnap's inductive logic.[37]

If one adopts this sort of inductive logic, it is not altogether clear how there could be any such thing as I-S explanation. For one thing, we could not literally construe I-S explanations as inductive arguments. For another, we would have to abandon the covering law conception, for we could never have any *accepted* law statements to include in the explanans of *any* explanation, D-N or I-S. But one benefit would accrue. There would no longer be any need to impose the high-probability requirement. In Carnap's inductive logic high probabilities do not enjoy special virtues. If inductive logic yields only betting quotients, determining fair odds for wagering, what counts is having a *correct* value, not a *high* value. It is not necessarily unreasonable to bet on improbable outcomes, provided the odds are right. Thus, by adopting this special Carnapian construal of "inductive argument" Hempel can drop the high-probability requirement and avoid the asymmetry between explanations of high-probability and low-probability occurrences. But the price, with respect to the received view of scientific explanation, is rather large (see W. Salmon 1977a).

Question 3, which ties in directly with the role of causality in scientific explanation, raises the most profound issues, I think. The notion that scientific explanation and causality are intimately connected is not new; it goes back at least to Aristotle. As we have seen, some of the earliest objections to Hempel's approach, voiced by Scriven, criticized the received view for its virtual neglect of causality. A little later, when the S-R model was being developed, it seemed clear to me that any adequate theory of scientific explanation must accord to causality a central role. The fact that we were dealing with statistical explanation, in contexts that might in some cases be indeterministic, did not preclude causality. After reading the nearly complete manuscript of Reichenbach's *The Direction of Time* (1956) in the summer of 1952 I was convinced that probabilistic causality is a viable notion. I hoped that such probabilistic concepts as his screening off and conjunctive forks would be helpful in explicating probabilistic causality, but none of it was really worked out at that time (W. Salmon et al. 1971, 76, 81).

Two major factors forced me to focus more carefully upon causality. The first was a brief article by Hugh Lehman (1972) that nicely pointed out the limitations, with respect to scientific explanation, of statistical relevance relations alone. Showing how distinct causal factors could give rise to identical statistical relevance relations, his argument strongly suggested the necessity of appealing to causal mechanisms.

The second factor involved theoretical explanation. Everyone recognizes the obvious fact that among our most impressive scientific explanations are many that appeal to theories that apparently make reference to such unobservable entities as molecules, atoms, and subatomic particles. There is no obvious way in which statistical relevance relations by themselves can account for this feature of scientific explanation. Greeno made a stab at doing something along that line, but his attempt just did not work (see Greeno 1971 and Salmon 1971). I decided to have a try at analyzing theoretical explanation in a paper (whose ambitious title was chosen before the paper was written) for the Conference on Explanation at the University of Bristol in 1973 (W. Salmon 1975). In attempting to cope with that task I found myself deeply involved with causal connections and causal mechanisms. The paper fell far short of its intended goal. But it reinforced in my mind the need to come to grips in a serious way with causal concepts. Theoretical explanation had to wait for the fourth decade (W. Salmon 1978).

3.7 The Challenge of Causality

With causality demanding attention, it was impossible to ignore the issues Hume had raised regarding the nature of causal relations. If we are to succeed in putting "cause" back into "because," we need to understand causal connections. In a great many cases, in everyday life and in science, we take one event to be a cause of another even though they are not spatio-temporally contiguous. In such cases we expect to be able to find some physical connection between them. We flick a switch on the wall by the door and a light on the ceiling at the center of the room lights up. They are connected by an electrical circuit, and a current flows when the switch is closed. Ammonia is spilled in the laundry room and its odor is detected in the hallway. Ammonia molecules diffuse through the air from one place to another. As long as we steer clear of quantum mechanical phenomena, the notion of continuous physical connections between causes and effects seems assured. Hume, in his characterization of cause-effect relations, mentioned priority, contiguity, and constant conjunction. He failed to find any additional feature of the situation that constitutes a causal connection. If the cause and effect are not contiguous, the natural move is to interpolate intermediate causes so as to form a causal chain with contiguous links. But every time a new link is found and inserted, the Humean question arises all over again: what is the *connection* between these intermediate causes and their effects? It seems that the Humean question never gets answered.

Oddly enough, the special theory of relativity offers a useful clue. According to that theory, light travels in a vacuum at a constant speed c, and no signal can be transmitted at any greater speed. Material particles can be accelerated to speeds approaching c, but they can never attain that speed or any greater speed.[38] It is often said, roughly, that nothing can travel faster than light, but that statement must be treated with great care. What does the "thing" in "nothing" refer to? Certainly it covers such entities as trains, baseballs, electrons, spaceships, photons, radio waves, and sound waves. But what about shadows? Are they *things* in the relevant sense of the term? The answer is negative; shadows can travel at arbitrarily high speeds. Their speed is not limited by the speed of light. Are there other examples? Yes. Consider the spot of light that is cast by a moving flashlight on the wall of a dark room. Like the shadow, this spot can travel at an arbitrarily high speed. There are many other examples. They all share the characteristic of being incapable of transmitting messages. If signals could be sent faster than light Einstein's famous *principle of relativity of simultaneity* would be undermined.

The best way to explain the difference illustrated by the examples in the preceding paragraph is, I think, to make a distinction between genuine *causal processes* and *pseudo-processes*. I shall use the unqualified term "process" in a general way to cover both the causal and pseudo varieties, but I shall not attempt to define it; examples will have to suffice. The general idea is quite simple. An event is something that happens in a fairly restricted region of spacetime. The context determines how large or small that region may be. A process is something that, in the context, has greater temporal duration than an event. From the standpoint of cosmology a supernova explosion–such as occurred in our cosmic neighborhood in the last year of the fourth decade–could be considered an event; the travel of a photon or neutrino from the explosion to earth (requiring thousands of years) would be a process. In ordinary affairs a chance meeting with a friend in a supermarket would normally be considered an event; the entire shopping trip might qualify as a process. In microphysics a collision of a photon with an electron would constitute an event; an electron orbiting an atomic nucleus would qualify as a process. A process, whether causal or pseudo, will exhibit some sort of uniformity or continuity. Something that, in one context, would be considered a single process (such as running a mile) would often be considered a complex combination of many processes from another standpoint (e.g., that of a physiologist).[39]

We must now try to distinguish between genuine causal processes and pseudo-processes. Material particles in motion (or even at rest[40]), radio waves, and photons are examples of genuine causal processes. All of them can be used to send signals; they can transmit information. Consider an ordinary piece of paper. You can write a message on it and send it through the mail to another person at another location. Or consider radio waves. They can be used to send information from a radio station to a receiver in someone's home, or from a command center to a space vehicle traveling to distant planets. Shadows and moving spots of light cast on walls are *not* causal processes; they are incapable of transmitting information. By means of another example I shall try to show why.

A lighthouse on a promontory of land sends a white beam of light that can be seen by ships at a great distance. This beam can be modified or marked. By fitting the light in the tower with a red lens, the beam becomes red. By inserting the red glass in the beam *at one place*, we can change the beam of light from white to red *from that point on*. It does not matter where in the beam the red filter is placed; wherever it is inserted it changes the beam from white to red from that point on. A genuine *causal process* is one that can transmit a mark; if the process is modified at one stage the modification persists beyond that point *without any additional intervention*.[41]

The lighthouse beacon rotates, so that it can be seen by ships in all directions. On a cloudy night, the rotating beacon casts a moving white spot on the clouds. This spot moves in a regular way; it is some sort of process, but it is *not* a causal process. If somehow the spot on the cloud is changed to red at one place–by someone in a balloon holding a piece of red cellophane right at the surface of the cloud, or by encountering a red balloon at the cloud's surface–it will suffer a modification or mark at that place, but the modification or mark will not persist beyond that place without further interventions. A 'process' that can be marked at one place, but without having any such modification persist beyond the point at which the mark is made, cannot *transmit* marks.[42] Such 'processes' are *pseudo-processes*.[43]

The moving spot of light is like the shadow in that it can travel at an arbitrarily high speed. This is easy to see. If the beacon rotates quite rapidly, its spot of light will traverse the clouds at high speed. Imagine now that the beacon continues to rotate at the same rate, but that the clouds are moved farther away. The farther away the clouds, the greater the linear velocity of the spot will be–since it has to traverse a circle of greater circumference in the same amount of time. A dramatic example is furnished by the pulsar in the Crab nebula, which rotates 30 times per second, sending out a continuous beam of electromagnetic radiation. It is about 6500 light years away. Its 'spot' of radiation sweeps by us at approximately 4×10^{13} times the speed of light.

There are many familiar examples of pseudo-processes. The scanning pattern on a cathode ray tube–a television receiver, a computer terminal, or an oscilloscope–furnishes one. The screen is made of a substance that scintillates when electrons strike it. An electron gun shoots electrons at it. As a result of the impinging electrons a spot of light moves rapidly back and forth across the screen. Another is the action as viewed at the cinema. Genuine causal processes–such as a horse running across a plain–are depicted, but what is seen on the screen is a pseudo-process created by light from a projector passing through a film and falling upon the screen. The light traveling from the projector to the screen is a whole sequence of causal processes–we might say that the light passing through each separate frame is a causal process. The motion of the horse on the screen is a pseudo-process. By momentarily shining a red light on the image of the horse on the screen it is possible to put a red spot on the horse, but if the red light does not continue to impinge on the image of the horse, the image of the horse will not continue to exhibit the red spot.[44]

The distinction between causal processes and pseudo-processes is important to our discussion of causality because processes that are *capable of transmitting marks* are processes that can also transmit information. Such processes transmit energy; they also transmit *causal influence*. They provide the *causal connections* among events that happen at different times and places in the universe.

Transmission is the key concept here. We can say, intuitively, that a process has some sort of structure–for instance, a light beam contains light of certain frequencies. A filter in its path changes that structure by removing light of some of these frequencies. Other modifications can be made by means of polarizers or mirrors. If the mark persists beyond the point in the process at which it is imposed, the structure has been modified in a lasting way *without further interventions*. If the change in structure is transmitted, the structure itself is being transmitted. Processes that can transmit marks *actually do transmit their own structure*.

Still, we should remind ourselves, the concept of transmission is a causal concept, and it must confront the Humean question. The terminology of mark transmission strongly suggests that the earlier parts of the marked process have the power to produce or reproduce the marked structure. In what does this power consist? How does the mark *get from* an earlier place in the process to a later one? Putting the question in this way suggests a strong analogy. More than 2500 years ago Zeno of Elea posed the famous paradox of the flying arrow. How does the arrow *get from* point A to point B in its trajectory? Zeno seems to have suggested that, since the arrow simply occupies the intervening positions, it must always be at rest. If it is always at rest it can never move; motion implies a contradiction.

Early in the twentieth century Bertrand Russell offered what is, I believe, a completely satisfactory resolution of the arrow paradox in terms of the so-called *at-at theory of motion*. Motion, Russell observed, is nothing more than a functional relationship between points of space and moments of time. To move *is* simply to occupy different positions in space at different moments of time. Considering any single point of space and single instant of time, an object simply is at that point of space at that instant of time. There is no distinction between being in motion and being at rest as long as we consider only that one point and that one instant.[45] To get from point A to point B consists merely of being *at* the intervening points of space *at* the corresponding moments of time. There is no further question of how the moving arrow gets from one point to another.

There was, it seemed to me, a similar answer to the question about mark transmission. A mark that is imposed at point A in a process is *transmitted* to point B in that same process if, *without additional interventions*, the mark is present at each intervening stage in the process. The difference between a process transmitting a mark and not transmitting that mark is that in the latter case the mark is present at the later stages in the process only if additional interventions occur reimposing that mark (W. Salmon 1977).

For a process to qualify as causal, it is not necessary that it actually be transmitting a mark; it is sufficient that it be capable of transmitting marks. Moreover, a causal process may not transmit every mark that is imposed upon it; a light wave that encounters another light wave will be modified at the point of intersection, but the modification will not persist beyond the point of intersection. That does not matter. The fact that it will transmit some kinds of marks is sufficient to qualify a light wave as a causal process.

With the *at-at theory of mark transmission* and the distinction between causal processes and pseudo-processes we have, it seemed to me, a satisfactory answer to Hume's basic question about the nature of causal connections. Although quite a bit more is required to develop a full-blown explication of causal explanation, this appeared to me to be the essential key (for further details see W. Salmon 1984, chap. 5).

3.8 Teleological and Functional Explanation

As we saw at the outset, a deep concern about teleology was a major impetus to the philosophical study of scientific explanation that has developed in the twentieth century. In roughly the first decade of our chronicle several of the principal early contributors to that literature—Braithwaite, Hempel, Nagel, and Scheffler—explicitly addressed the problem of teleological and/or functional explanation. This concern was instigated in large part by vitalism in biology.

Until the beginning of the third decade philosophy of biology was not pursued very actively by philosophers of science, but, as this decade opened, the situation began to change. Morton Beckner's classic, *The Biological Way of Thought*, appeared in 1968, and during the early 1970s the field began to expand dramatically with the work of such people as David Hull, Michael Ruse, and William Wimsatt. It has continued to flourish and grow right down to the present.[46] The problem of teleological/functional explanation is, of course, central to biology. Toward the end of the third decade, Larry Wright (1976) provided an account that strikes me as fundamentally correct, and it applies to many other areas in addition to biology. Eleven years later, at the very end of the fourth decade, John Bigelow and Robert Pargetter (1987) offer a theory of functional explanations that may be an improvement over Wright's, but, as we shall see, it is not *fundamentally* such a very different account.

Wright takes as his point of departure the view that teleological explanations occur frequently in science, and he maintains—contra Hempel and Beckner—that many of them are sound. He regards teleological explanations as causal in a straightforward sense—in a sense that does *not* require causes that come after their effects. At the same time, nevertheless, he considers them to be future-oriented; teleology by definition involves goal-seeking behavior. To capture both of these insights, he offers an account in terms of what he calls a *consequence-etiology*.

The basic idea of a consequence-etiology is as ingenious as it is simple. A particular bit of behavior B occurs because B has been *causally efficacious in the past* in achieving a goal G. A cat, hunting for prey, is clearly engaged in goal-directed behavior. It stalks in a typically catlike way because such stalking has resulted in the procurement of food. It is not caused by the future catching of this particular mouse, for (among other problems) in this instance he may not succeed in catching his prey. But such behavior has worked often enough to have conferred an evolutionary advantage on the members of the species. Similarly, a human being searching for water in an arid region may look for cottonwood trees, since past experience shows that they grow only when there is water in the vicinity. Such intentional behavior relies on a conviction that this action B is a suitable means of achieving G. The behavior of a homing torpedo also occurs because in the past such actions B have resulted in hitting the target (goal G). Roughly speaking, the causal efficacy of B in bringing about G in the past is, itself, an indispensable part of the cause of the occurrence of B on this occasion. It is a consequence-etiology because the consequences of doing B are a crucial part of the etiology of the doing of B.

In the case of deliberate action on the part of a human agent, it may be that direct experience of the consequence G of action B is what makes the agent do B on this occasion. Perhaps he or she has spent a great deal of time in arid regions and has found that locating cottonwood trees has been a successful strategy in finding water. Perhaps the agent has not had any such direct experience but has heard or read about this strategy. Perhaps this person has neither had direct experience, nor learned of it through reports, but, rather, inferred that it would work. Whatever the actual situation, the agent does B because he or she has reason to believe, before performing that act on this particular occasion, that B is an appropriate method for getting G.

A homing torpedo is an artifact created by a human designer to behave in a certain way. If it is being used in warfare there is a reasonable presumption (we hope) that this type of device was tested and found to perform in the desired way. The fact that similar behavior by similar objects in the past has resulted in the reaching of the goal etiologically causes the present behavior of this particular torpedo.

Examples from evolutionary biology are, I think, the clearest illustrations of Wright's model. Certain kinds of behavior, such as the stalking by a cat, become typical behavior for a species of animals because it confers an advantage with respect to the goal of survival or of reproduction. As Wright explicitly notes, the behavior may have arisen in either of two ways—deliberate creation by a supernatural agency or as a result of natural selection. The analysis of the concept of *teleology* does not tell us which of the two possible etiologies is the actual one. It only ensures that there is an etiology; empirical science must furnish the answer to which it is.

According to Wright's analysis, human conscious intentional behavior, the behavior of human artifacts that have been designed to do some particular job, and behavior that has resulted from natural selection all qualify as teleological. He offers the following schematic statement (T):

> S does B for the sake of G iff:
>
> (i) B tends to bring about G.
> (ii) B occurs because (i.e., is brought about by the fact that) it tends to bring about G. (1976, 39)

The term "tends" signifies the fact that B need not *always* succeed in bringing about G. Indeed, it may be that B never brings about G, as long as it has a disposition to do so under suitable conditions. Wright offers this gloss:

> teleological behavior is behavior with a consequence-etiology: and behavior with a consequence-etiology is behavior that occurs because it brings about, is the type of thing that brings about, tends to bring about, is required to bring about, or is in some other way appropriate for bringing about some specific goal. (1976, 38–39)

Although they are obviously closely related, Wright does not identify teleological explanations with functional ascriptions or explanations. The basic reason is that only behavior is teleological—goal-directed. But in many cases something fulfills a function just by "being there." A piece of newspaper, stuffed under a door, fulfills the function of preventing a draft just because of its location. A vinyl cover lying on a playing field fulfills the function of keeping the ground dry. The function of an entity is distinguished from all sorts of other things that might result from its being where it is by the fact that its location has a consequence-etiology. The vinyl cover is where it is because it keeps the field dry. It may also catch pools of water in which children play, but that is an accidental result. Wright offers the following schematic formulation (F):

> The function of X is Z iff:
>
> (i) Z is a consequence (result) of X's being there,
> (ii) X is there because it does (results in) Z. (1976, 81)

He comments,

> The ascription of a function simply *is* the answer to a 'Why?' question, and one with etiological force. . . . So, not only do functional explanations provide consequence-etiologies, just like explanations in terms of goals, but the simple attribution of a function ipso facto *provides* that explanation (ascription-explanation), just as does the simple attribution of a goal to behavior. This displays the enormous parallel that obtains between goals and functions, and possibly accounts for the tendency in the philosophical literature to run them together. (1976, 81)

Functional ascriptions are, of course, frequently made in physiology and in evolutionary biology. The function of the long neck of the giraffe is to enable it to reach food other animals cannot. The function of the stripes on the tiger is to provide camouflage. The function of the heart is to pump blood. The function of chlorophyll in green plants is to enable them to produce starch by photosynthesis. In the course of evolution certain attributes arise in the first instance as a result of a mutation. In such a first occurrence the attribute does not have a function, for it has no consequence-etiology at that stage. The function is established only when the recurrence or persistence of the attribute occurs because of a consequence of its previous presence.

Wright asserts explicitly that his account of teleological explanation and functional ascription is not inconsistent with a completely mechanistic explanation of goal-directed behavior (1976, 57–72). The fact that we can give a completely mechanistic account of the motion of a homing torpedo does not render it nonteleological. The fact that we can give a completely mechanical account of the operation of a governor on a steam engine does not deprive it of the function of regulating the engine's speed. He maintains, moreover—contra Charles Taylor (1964) and others—that, even if it should turn out to be possible to provide a completely physico-chemical account of the behavior of plants and animals (including the conscious behavior of human animals), that would not eliminate its teleological character. Teleological and functional explanations are on as firm ground as any explanations in any science.

The fourth-decade proposal of Bigelow and Pargetter (1987) is a "propensity theory" that they claim to be more forward-directed, temporally, than Wright's consequence-etiology. It relies heavily upon the tendency or disposition of a given item to do whatever is its function. As we have seen (schema F above), Wright refers explicitly to what the item *does*, not to what it tends to do, though on the same page, immediately after the schema, he allows for cases in which an item has a function even if it never successfully performs it. What Bigelow and Pargetter fail to note is that Wright's analysis of function supplements his analysis of teleological behavior (schema T above); like the majority of philosophers to whom Wright referred, they seem to run teleology and function together. Whether they are correct in so doing is an issue I shall not argue. The point is that Wright's analysis of teleological behavior is as much a propensity theory, and is just as forward-looking, as is the Bigelow-Pargetter theory of functions. And Wright has emphasized, as we have seen, the extremely close connection between teleology and function. I am inclined to think that, on Wright's account, every case of a function presupposes some bit of teleological behavior, but he might well disagree.[47] In any case, as Bigelow and Pargetter seem in some places to concede, their theory is not very different from Wright's. The main difference may be a disagreement over whether an item has a function the first time it occurs, or whether, as Wright maintains, it can properly be said to have a function only subsequently. If this is the *main difference*, it is a matter of fine-tuning, or as Larry used to say regarding engines of racing cars, "demon tweeking."

While I agree wholeheartedly with Wright's complete causal grounding of teleological and functional explanations, I do have one major philosophical disagreement with him. Following in the footsteps of his teacher, Michael Scriven, he declines to offer any characterization of causality itself; instead, he maintains that we have perfectly objective ways of recognizing causal relations, and that is enough. In my view, this is simply evasion of a fundamental philosophical problem. As we shall see in discussing the fourth decade, I have devoted considerable effort to the explication of causality, and I do not believe the unanalyzed notion is nearly as unproblematic as Scriven and Wright have claimed.

In 1977—the final year of the third decade—in his John Dewey Lectures, Nagel again addressed the issues of teleological and functional explanation (1977). In the course of his discussion he offers criticisms of Hempel (Nagel 1977, 305–9) and Wright (ibid., 296–301)[48] as well as others.[49] His main response to Hempel consists in a challenge to the notion of functional equivalents. According to Nagel, if one specifies with sufficient precision the nature of the organism or system within which an item has a function, it is in many cases—including Hempel's example of circulation of blood as the function of the heart in a normal human being—necessary for the fulfillment of the function. Nagel is not moved by other possible devices, such as Jarvik artificial hearts, inasmuch as they do not circulate blood in a *normal* humans. He is equally unmoved by fictitious possibilities that are not realized in nature. He therefore concludes that the normal functioning n in Hempel's schema entails the presence of item i. Nagel does acknowledge the fact that in many actual cases more than one item fulfill the same function; for example, a normal human has two ears for hearing, either one of which will do the job fairly adequately. In cases of this sort, the fulfilling of the function entails the existence of a nonempty set which may contain more than one member.

The conclusion Nagel draws from this argument is that functional explanations fit the deductive pattern. From the fact that the organism is in a normal state we can deduce the existence of the item i (or set of such items). If one accepts Nagel's argument about functional equivalents he overcomes the main problem that led Hempel to conclude that so-called 'functional explanations' are not real explanations at all.[50] Nagel considers that result desirable, for he maintains that many sciences, including biology in particular, do provide legitimate teleological or functional explanations. In accepting this conclusion, however, he embraces the major difficulty voiced by Braithwaite, Scheffler, and others. He allows that various kinds of facts can be explained by appeal to subsequent conditions:

> What then is accomplished by such explanations? They make explicit *one* effect of an item i in system S, as well as that the item must be present in S on the assumption that the item does have that effect. In short, explanations of function ascriptions make evident one role some item plays in a given system. But if this is what such explanations accomplish, would it not be intellectually more profitable, so it might be asked, to discontinue investigations of the *effects* of various items, and replace them by inquiries into the *causal* (or antecedent) conditions for the occurrence of those items? The appropriate answer, it seems to me, is that inquiries into effects or consequences are just as legitimate as inquiries into causes or antecedent conditions; that biologists as