tower? He vowed that shadow would cover the terrace where he first proclaimed his love, with every setting sun—that is why the tower had to be so high.'

I took this in but slowly. It is never easy to assimilate unexpected truths about people we think we know—and I have had occasion to notice this again and again.

'Why did he kill her?' I asked finally.

'Because, sir, she dallied with an English brigadier, an overnight guest in this house.' With these words she arose, collected her bodice and cap, and faded through the wall beside the doorway.

I left early the next morning, making my excuses as well as I could.

§4. *A Model for Explanation*

I shall now propose a new theory of explanation. An explanation is not the same as a proposition, or an argument, or list of propositions; it is an *answer*. (Analogously, a son is not the same as a man, even if all sons are men, and every man is a son.) An explanation is an answer to a why-question. So, a theory of explanation must be a theory of why-questions.

To develop this theory, whose elements can all be gleaned, more or less directly, from the preceding discussion, I must first say more about some topics in formal pragmatics (which deals with context-dependence) and in the logic of questions. Both have only recently become active areas in logical research, but there is general agreement on the basic aspects to which I limit the discussion.

§4.1 *Contexts and Propositions*[36]

Logicians have been constructing a series of models of our language, of increasing complexity and sophistication. The phenomena they aim to save are the surface grammar of our assertions and the inference patterns detectable in our arguments. (The distinction between logic and theoretical linguistics is becoming vague, though logicians' interests focus on special parts of our language, and require a less faithful fit to surface grammar, their interests remaining in any case highly theoretical.) Theoretical entities introduced by logicians in their models of language (also called 'formal languages') include domains of discourse ('universes'), possible words, accessibility ('relative possibility') relations, facts and propositions, truth-values, and, lately, contexts. As might be guessed, I take it to be part of empiricism to insist that the adequacy of these models

does not require all their elements to have counterparts in reality. They will be good if they fit those phenomena to be saved.

Elementary logic courses introduce one to the simplest models, the languages of sentential and quantificational logic which, being the simplest, are of course the most clearly inadequate. Most logic teachers being somewhat defensive about this, many logic students, and other philosophers, have come away with the impression that the over-simplifications make the subject useless. Others, impressed with such uses as elementary logic does have (in elucidating classical mathematics, for example), conclude that we shall not understand natural language until we have seen how it can be regimented so as to fit that simple model of horseshoes and truth tables.

In elementary logic, each sentence corresponds to exactly one proposition, and the truth-value of that sentence depends on whether the proposition in question is true in the actual world. This is also true of such extensions of elementary logic as free logic (in which not all terms need have an actual referent), and normal modal logic (in which non-truth functional connectives appear), and indeed of almost all the logics studied until quite recently.

But, of course, sentences in natural language are typically context-dependent; that is, which proposition a given sentence expresses will vary with the context and occasion of use. This point was made early on by Strawson, and examples are many:

> 'I am happy now' is true in context x exactly if the speaker in context x is happy at the time of context x,

where a context of use is an actual occasion, which happened at a definite time and place, and in which are identified the speaker (referent of 'I'), addressee (referent of 'you'), person discussed (referent of 'he'), and so on. That contexts so conceived are idealizations from real contexts is obvious, but the degree of idealization may be decreased in various ways, depending on one's purposes of study, at the cost of greater complexity in the model constructed.

What must the context specify? The answer depends on the sentence being analysed. If that sentence is

> Twenty years ago it was still possible to prevent the threatened population explosion in that country, but now it is too late

the model will contain a number of factors. First, there is a set of possible worlds, and a set of contexts, with a specification for each context of the world of which it is a part. Then there will be for each world a set of entities that exist in that world, and also various relations of relative possibility among these worlds. In addition there is time, and each context must have a time of occurrence. When we evaluate the above sentence we do so relative to a context and a world. Varying with the context will be the referents of 'that country' and 'now', and perhaps also the relative possibility relation used to interpret 'possible', since the speaker may have intended one of several senses of possibility.

This sort of interpretation of a sentence can be put in a simple general form. We first identify certain entities (mathematical constructs) called propositions, each of which has a truth-value in each possible world. Then we give the context as its main task the job of selecting, for each sentence, the proposition it expresses 'in that context'. Assume as a simplification that when a sentence contains no indexical terms (like 'I', 'that', 'here', etc.), then all contexts select the same proposition for it. This gives us an easy intuitive handle on what is going on. If *A* is a sentence in which no indexical terms occur, let us designate as $|A|$ the proposition which it expresses in every context. Then we can generally (though not necessarily always) identify the proposition expressed by any sentence in a given context as the proposition expressed by some indexical-free sentence. For example:

> In context *x*, 'Twenty years ago it was still possible to prevent the population explosion in that country' expresses the proposition 'In 1958, it is (tenseless) possible to prevent the population explosion in India'

To give another example, in the context of my present writing, 'I am here now' expresses the proposition that Bas van Fraassen is in Vancouver, in July 1978.

This approach has thrown light on some delicate conceptual issues in philosophy of language. Note for example that 'I am here' is a sentence which is true no matter what the facts are and no matter what the world is like, and no matter what context of usage we consider. Its truth is ascertainable *a priori*. But the proposition expressed, that van Fraassen is in Vancouver (or whatever else it is) is not at all a necessary one: I might not have been here. Hence, a clear distinction between *a priori* ascertainability and necessity appears.

The context will generally select the proposition expressed by a given sentence *A* via a selection of referents for the terms, extensions for the predicates, and functions for the functors (i.e. syncategorematic words like 'and' or 'most'). But intervening contextual variables may occur at any point in these selections. Among such variables there will be the assumptions taken for granted, theories accepted, world-pictures or paradigms adhered to, in that context. A simple example would be the range of conceivable worlds admitted as possible by the speaker; this variable plays a role in determining the truth-value of his modal statements in that context, relative to the 'pragmatic presuppositions'. For example, if the actual world is really the only possible world there is (which exists) then the truth-values of modal statements in that context but *tout court* will be very different from their truth-values relative to those pragmatic presuppositions—and only the latter will play a significant role in our understanding of what is being said or argued in that context.

Since such a central role is played by propositions, the family of propositions has to have a fairly complex structure. Here a simplifying hypothesis enters the fray: propositions can be uniquely identified through the worlds in which they are true. This simplifies the model considerably, for it allows us to identify a proposition with a set of possible worlds, namely, the set of worlds in which it is true. It allows the family of propositions to be a complex structure, admitting of interesting operations, while keeping the structure of each individual proposition very simple.

Such simplicity has a cost. Only if the phenomena are simple enough, will simple models fit them. And sometimes, to keep one part of a model simple, we have to complicate another part. In a number of areas in philosophical logic it has already been proposed to discard that simplifying hypothesis, and to give propositions more 'internal structure'. As will be seen below, problems in the logic of explanation provide further reasons for doing so.

§4.2 *Questions*

We must now look further into the general logic of questions. There are of course a number of approaches; I shall mainly follow that of Nuel Belnap, though without committing myself to the details of his theory.[37]

A theory of questions must needs be based on a theory of propositions, which I shall assume given. A *question* is an abstract entity;

it is expressed by an *interrogative* (a piece of language) in the same sense that a proposition is expressed by a declarative sentence. Almost anything can be an appropriate response to a question, in one situation or another; as 'Peccavi' was the reply telegraphed by a British commander in India to the question how the battle was going (he had been sent to attack the province of Sind).[38] But not every response is, properly speaking, an answer. Of course, there are degrees; and one response may be more or less of an answer than another. The first task of a theory of questions is to provide some typology of answers. As an example, consider the following question, and a series of responses:

Can you get to Victoria both by ferry and by plane?
(*a*) Yes.
(*b*) You can get to Victoria both by ferry and by plane.
(*c*) You can get to Victoria by ferry.
(*d*) You can get to Victoria both by ferry and by plane, but the ferry ride is not to be missed.
(*e*) You can certainly get to Victoria by ferry, and that is something not to be missed.

Here (*b*) is the 'purest' example of an answer: it gives enough information to answer the question completely, but no more. Hence it is called a *direct answer*. The word 'Yes' (*a*) is a *code* for this answer.

Responses (*c*) and (*d*) depart from that direct answer in opposite directions: (*c*) says properly less than (*b*)—it is implied by (*b*)—while (*d*), which implies (*b*), says more. Any proposition implied by a direct answer is called a *partial answer* and one which implies a direct answer is a *complete answer*. We must resist the temptation to say that therefore an answer, *tout court*, is any combination of a partial answer with further information, for in that case, every proposition would be an answer to any question. So let us leave (*e*) unclassified for now, while noting it is still 'more of an answer' than such responses as 'Gorilla!' (which is a response given to various questions in the film *Ich bin ein Elephant, Madam*, and hence, I suppose, still more of an answer than some). There may be some quantitative notion in the background (a measure of the extent to which a response really 'bears on' the question) or at least a much more complete typology (some more of it is given below), so it is probably better not to try and define the general term 'answer' too soon.

The basic notion so far is that of direct answer. In 1958, C. L. Hamblin introduced the thesis that a question is uniquely identifiable through its answers.[39] This can be regarded as a simplifying hypothesis of the sort we come across for propositions, for it would allow us to identify a question with the set of its direct answers. Note that this does not preclude a good deal of complexity in the determination of exactly what question is expressed by a given interrogative. Also, the hypothesis does not identify the question with the disjunction of its direct answers. If that were done, the clearly distinct questions

Is the cat on the mat?
direct answers: The cat is on the mat.
The cat is not on the mat.
Is the theory of relativity true?
direct answers: The theory of relativity is true.
The theory of relativity is not true.

would be the same (identified with the tautology) if the logic of propositions adopted were classical logic. Although this simplifying hypothesis is therefore not to be rejected immediately, and has in fact guided much of the research on questions, it is still advisable to remain somewhat tentative towards it.

Meanwhile we can still use the notion of direct answer to define some basic concepts. One question Q may be said to *contain* another, Q', if Q' is answered as soon as Q is—that is, every complete answer to Q is also a complete answer to Q'. A question is *empty* if all its direct answers are necessarily true, and *foolish* if none of them is even possibly true. A special case is the *dumb* question, which has no direct answers. Here are examples:

1. Did you wear the black hat yesterday or did you wear the white one?
2. Did you wear a hat which is both black and not black, or did you wear one which is both white and not white?
3. What are three distinct examples of primes among the following numbers: 3, 5?

Clearly 3 is dumb and 2 is foolish. If we correspondingly call a necessarily false statement foolish too, we obtain the theorem *Ask a foolish question and get a foolish answer*. This was first proved by Belnap, but attributed by him to an early Indian philosopher mentioned in Plutarch's *Lives* who had the additional distinction of being an early nudist. Note that a foolish question contains all questions, and an empty one is contained in all.

Example 1 is there partly to introduce the question form used in 2, but also partly to introduce the most important semantic concept after that of direct answer, namely presupposition. It is easy to see that the two direct answers to 1 ('I wore the black hat', 'I wore the white one') could both be false. If that were so, the respondent would presumably say 'Neither', which is an answer not yet captured by our typology. Following Belnap who clarified this subject completely, let us introduce the relevant concepts as follows:

> a *presupposition*[40] of question Q is any proposition which is implied by all direct answers to Q.
>
> a *correction* (or *corrective answer*) to Q is any denial of any presupposition of Q.
>
> the (*basic*) *presupposition* of Q is the proposition which is true if and only if some direct answer to Q is true.

In this last notion, I presuppose the simplifying hypothesis which identifies a proposition through the set of worlds in which it is true; if that hypothesis is rejected, a more complex definition needs to be given. For example 1, 'the' presupposition is clearly the proposition that the addressee wore either the black hat or the white one. Indeed, in any case in which the number of direct answers is finite, 'the' presupposition is the disjunction of those answers.

Let us return momentarily to the typology of answers. One important family is that of the partial answers (which includes direct and complete answers). A second important family is that of the corrective answer. But there are still more. Suppose the addressee of question 1 answers 'I did not wear the white one.' This is not even a partial answer, by the definition given: neither direct answer implies it, since she might have worn both hats yesterday, one in the afternoon and one in the evening, say. However, since the questioner is presupposing that she wore at least one of the two, the response is *to him* a complete answer. For the response plus the presupposition together entail the direct answer that she wore the black hat. Let us therefore add:

> a *relatively complete answer* to Q is any proposition which, together with the presupposition of Q, implies some direct answer to Q.

We can generalize this still further: a complete answer to Q, relative to theory T, is something which together with T, implies some direct answer to Q—and so forth. The important point is, I think, that we should regard the introduced typology of answers as open-ended, to be extended as needs be when specific sorts of questions are studied.

Finally, which question is expressed by a given interrogative? This is highly context-dependent, in part because all the usual indexical terms appear in interrogatives. If I say 'Which one do you want?' the context determines a range of objects over which my 'which one' ranges—for example, the set of apples in the basket on my arm. If we adopt the simplifying hypothesis discussed above, then the main task of the context is to delineate the set of direct answers. In the 'elementary questions' of Belnap's theory ('whether-questions' and 'which-questions') this set of direct answers is specified through two factors: a *set of alternatives* (called the *subject* of the question) and *request* for a selection among these alternatives and, possibly, for certain information about the selection made ('distinctness and completeness claims'). What those two factors are may not be made explicit in the words used to frame the interrogative, but the context has to determine them exactly if it is to yield an interpretation of those words as expressing a unique question.

§4.3 *A Theory of Why-questions*

There are several respects in which why-questions introduce genuinely new elements into the theory of questions.[41] Let us focus first on the determination of exactly what question is asked, that is, the contextual specification of factors needed to understand a why-interrogative. After that is done (a task which ends with the delineation of the set of direct answers) and as an independent enterprise, we must turn to the evaluation of those answers as good or better. This evaluation proceeds with reference to the part of science accepted as 'background theory' in that context.

As example, consider the question 'Why is this conductor warped?' The questioner implies that the conductor is warped, and is asking for a reason. Let us call the proposition that the conductor is warped the *topic* of the question (following Henry Leonard's terminology, 'topic of concern'). Next, this question has a *contrast-class*, as we saw, that is, a set of alternatives. I shall take this contrast-class, call it X, to be a class of propositions which includes the topic. For this particular interrogative, the contrast could be that it is *this* conductor rather than *that* one, or that this conductor has warped rather than retained its shape. If the question is 'Why does this material burn yellow' the contrast-class could be the set of propositions: this material burned (with a flame of) colour x.

Finally, there is the respect-in-which a reason is requested, which determines what shall count as a possible explanatory factor, the relation of *explanatory relevance*. In the first example, the request might be *for events 'leading up to' the warping*. That allows as relevant an account of human error, of switches being closed or moisture condensing in those switches, even spells cast by witches (since the evaluation of what is a good answer comes later). On the other hand, the events leading up to the warping might be well known, in which case the request is likely to be for the standing conditions that made it possible for those events to lead to this warping: the presence of a magnetic field of a certain strength, say. Finally, it might already be known, or considered immaterial, exactly how the warping is produced, and the question (possibly based on a misunderstanding) may be about exactly what function this warping fulfils in the operation of the power station. Compare 'Why does the blood circulate through the body?' answered (1) 'because the heart pumps the blood through the arteries' and (2) 'to bring oxygen to every part of the body tissue'.

In a given context, several questions agreeing in topic but differing in contrast-class, or conversely, may conceivably differ further in what counts as explanatorily relevant. Hence we cannot properly ask what is relevant to this topic, or what is relevant to this contrast-class. Instead we must say of a given proposition that it is or is not relevant (in this context) to the topic with respect to that contrast-class. For example, in the same context one might be curious about the circumstances that led Adam to eat the apple rather than the pear (Eve offered him an apple) and also about the motives that led him to eat it rather than refuse it. What is 'kept constant' or 'taken as given' (that he ate the fruit; that what he did, he did to the apple) which is to say, the contrast-class, is not to be dissociated entirely from the respect-in-which we want a reason.

Summing up then, the why-question Q expressed by an interrogative in a given context will be determined by three factors:

The *topic* P_k
The *contrast-class* $X=\{P_1, \ldots, P_k, \ldots\}$
The *relevance relation* R

and, in a preliminary way, we may identify the abstract why-question with the triple consisting of these three:

$$Q=\langle P_k, X, R\rangle$$

A proposition A is called *relevant to* Q exactly if A bears relation R to the couple $\langle P_k, X\rangle$.

We must now define what are the direct answers to this question. As a beginning let us inspect the form of words that will express such an answer:

(*) P_k *in contrast to* (the rest of) X *because* A

This sentence must express a proposition. What proposition it expresses, however, depends on the same context that selected Q as the proposition expressed by the corresponding interrogative ('Why P_k?'). So some of the same contextual factors, and specifically R, may appear in the determination of the proposition expressed by (*).

What is claimed in answer (*)? First of all, that P_k is true. Secondly, (*) claims that the other members of the contrast-class are not true. So much is surely conveyed already by the question—it does not make sense to ask why Peter rather than Paul has paresis if they both have it. Thirdly, (*) says that A is true. And finally, there is that word 'because': (*) claims that A is a *reason*.

This fourth point we have awaited with bated breath. Is this not where the inextricably modal or counterfactual element comes in? But not at all; in my opinion, the word 'because' here signifies only that A is relevant, in this context, to this question. Hence the claim is merely that A bears relation R to $\langle P_k, X\rangle$. For example, suppose you ask why I got up at seven o'clock this morning, and I say 'because I was woken up by the clatter the milkman made'. In that case I have interpreted your question as asking for a sort of reason that at least includes events-leading-up-to my getting out of bed, and my word 'because' indicates that the milkman's clatter was that sort of reason, that is, one of the events in what Salmon would call the causal process. Contrast this with the case in which I construe your request as being specifically for a motive. In that case I would have answered 'No reason, really. I could easily have stayed in bed, for I don't particularly want to do anything today. But the milkman's clatter had woken me up, and I just got up from force of habit I suppose.' In this case, I do not say 'because' for the milkman's clatter does not belong to the relevant range of events, as I understand your question.

It may be objected that 'because A' does not only indicate that A is *a* reason, but that it is *the* reason, or at least that it is a good reason. I think that this point can be accommodated in two ways. The first is that the relevance relation, which specifies what sort of thing is being requested as answer, may be construed quite strongly: 'give me a motive strong enough to account for murder', 'give me a statistically relevant preceding event not screened off by other events', 'give me a common cause', etc. In that case the claim that the proposition expressed by A falls in the relevant range, is already a claim that it provides a telling reason. But more likely, I think, the request need not be construed that strongly; the point is rather that anyone who answers a question is in some sense tacitly claiming to be giving a good answer. In either case, the determination of whether the answer is indeed good, or telling, or better than other answers that might have been given, must still be carried out, and I shall discuss that under the heading of 'evaluation'.

As a matter of regimentation I propose that we count (*) as a direct answer *only if* A is relevant.[42] In that case, we don't have to understand the claim that A is relevant as explicit part of the answer either, but may regard the word 'because' solely as a linguistic signal that the words uttered are intended to provide an answer to the why-question just asked. (There is, as always, the tacit claim of the respondent that what he is giving is a good, and hence a relevant answer—we just do not need to make this claim part of the answer.) The definition is then:

> B is a *direct answer* to question $Q = \langle P_k, X, R\rangle$ exactly if there is some proposition A such that A bears relation R to $\langle P_k, X\rangle$ and B is the proposition which is true exactly if (P_k; *and* for all $i \neq k$, *not* P_i; and A) is true

where, as before, $X = \{P_1, \ldots, P_k, \ldots\}$. Given this proposed definition of the direct answer, what does a why-question presuppose? Using Belnap's general definition we deduce:

> a why-question *presupposes* exactly that
>
> (a) its topic is true
>
> (b) in its contrast-class, only its topic is true
>
> (c) at least one of the propositions that bears its relevance relation to its topic and contrast-class, is also true.

However, as we shall see, if all three of these presuppositions are true, the question may still not have a *telling* answer.

Before turning to the evaluation of answers, however, we must consider one related topic: when does a why-question arise? In the general theory of questions, the following were equated: question Q arises, all the presuppositions of Q are true. The former means that Q is not to be rejected as mistaken, the latter that Q has some true answer.

In the case of why-questions, we evaluate answers in the light of accepted background theory (as well as background information) and it seems to me that this drives a wedge between the two concepts. Of course, sometimes we reject a why-question because we think that it has no true answer. But as long as we do not think that, the question does arise, and is not mistaken, regardless of what is true.

To make this precise, and to simplify further discussion, let us introduce two more special terms. In the above definition of 'direct answer', let us call proposition A the *core* of answer B (since the answer can be abbreviated to '*Because* A'), and let us call the proposition that (P_k *and for all* $i \neq k$, *not* P_i) the *central presupposition* of question Q. Finally, if proposition A is relevant to $\langle P_k, X\rangle$ let us also call it relevant to Q.

In the context in which the question is posed, there is a certain body K of accepted background theory and factual information. This is a factor in the context, since it depends on who the questioner and audience are. It is this background which determines whether or not the question arises; hence a question may arise (or conversely, be rightly rejected) in one context and not in another.

To begin, whether or not the question genuinely *arises*, depends on whether or not K implies the central presupposition. As long as the central presupposition is not part of what is assumed or agreed to in this context, the why-question does not arise at all.

Secondly, Q presupposes *in addition* that one of the propositions A, relevant to its topic and contrast-class, is true. Perhaps K does not imply that. In this case, the question will still arise, provided K does not imply that all those propositions are false.

So I propose that we use the phrase 'the question arises in this context' to mean exactly this: K implies the central presupposition, and K does not imply the denial of any presupposition. Notice that this is very different from 'all the presuppositions are true', and we may emphasize this difference by saying 'arises in context'. The reason we must draw this distinction is that K may not tell us which of the possible answers is true, but this *lacuna* in K clearly does not eliminate the question.

§4.4 *Evaluation of Answers*

The main problems of the philosophical theory of explanation are to account for legitimate rejections of explanation requests, and for the asymmetries of explanation. These problems are successfully solved, in my opinion, by the theory of why-questions as developed so far.

But that theory is not yet complete, since it does not tell us how answers are evaluated as telling, good, or better. I shall try to give an account of this too, and show along the way how much of the work by previous writers on explanation is best regarded as addressed to this very point. But I must emphasize, first, that this section is not meant to help in the solution of the traditional problems of explanation; and second, that I believe the theory of why-questions to be basically correct as developed so far, and have rather less confidence in what follows.

Let us suppose that we are in a context with background K of accepted theory plus information, and the question Q arises here. Let Q have topic B, and contrast-class $X=\{B, C, \ldots, N\}$. How good is the answer *Because A*?

There are at least three ways in which this answer is evaluated. The first concerns the evaluation of A itself, as acceptable or as likely to be true. The second concerns the extent to which *A favours* the topic B as against the other members of the contrast-class. (This is where Hempel's criterion of giving reasons to expect, and Salmon's criterion of statistical relevance may find application.) The third concerns the comparison of *Because A* with other possible answers to the same question; and this has three aspects. The first is whether A is more probable (in view of K); the second whether it favours the topic to a greater extent; and the third, whether it

is made wholly or partially irrelevant by other answers that could be given. (To this third aspect, Salmon's considerations about *screening off* apply.) Each of these three main ways of evaluation needs to be made more precise.

The first is of course the simplest: we rule out *Because A* altogether if K implies the denial of A; and otherwise ask what probability K bestows on A. Later we compare this with the probability which K bestows on the cores of other possible answers. We turn then to favouring.

If the question why B rather than $C, \ldots, N$ arises here, K must imply B and imply the falsity of $C, \ldots, N$. However, it is exactly the information that the topic is true, and the alternatives to it not true, which is irrelevant to how favourable the answer is to the topic. The evaluation uses only that part of the background information which constitutes the general theory about these phenomena, plus other 'auxiliary' facts which are known but which do not imply the fact to be explained. This point is germane to all the accounts of explanation we have seen, even if it is not always emphasized. For example, in Salmon's first account, A explains B only if the probability of B given A does not equal the probability of B *simpliciter*. However, if I know that A and that B (as is often the case when I say that B because A), then my *personal probability* (that is, the probability given all the information I have) of A equals that of B and that of B given A, namely 1. Hence the probability to be used in evaluating answers is not at all the probability given all my background information, but rather, the probability given some of the general theories I accept plus some selection from my data.[43] So the evaluation of the answer *Because A* to question Q proceeds with reference only to a certain part $K(Q)$ of K. How that part is selected is equally important to all the theories of explanation I have discussed. Neither the other authors nor I can say much about it. Therefore the selection of the part $K(Q)$ of K that is to be used in the further evaluation of A, must be a further contextual factor.[44]

If $K(Q)$ plus A implies B, and implies the falsity of $C, \ldots, N$, then A receives in this context the highest marks for favouring the topic B.

Supposing that A is not thus, we must award marks on the basis of how well A redistributes the probabilities on the contrast-class so as to favour B against its alternatives. Let us call the probability in the light of $K(Q)$ alone the *prior* probability (in this context) and the probability given $K(Q)$ plus A the *posterior* probability. Then A will do best here if the posterior probability of B equals 1. If A is not thus, it may still do well provided it shifts the mass of the probability function toward B; for example, if it raises the probability of B while lowering that of $C, \ldots, N$; or if it does not lower the probability of B while lowering that of some of its closest competitors.

I will not propose a precise function to measure the extent to which the posterior probability distribution favours B against its alternatives, as compared to the prior. Two factors matter: the minimum odds of B against $C, \ldots, N$, *and* the number of alternatives in $C, \ldots, N$ to which B bears these minimum odds. The first should increase, the second decrease. Such an increased favouring of the topic against its alternatives is quite compatible with a decrease in the probability of the topic. Imagining a curve which depicts the probability distribution, you can easily see how it could be changed quite dramatically so as to single out the topic—as the tree that stands out from the wood, so to say—even though the new advantage is only a relative one. Here is a schematic example:

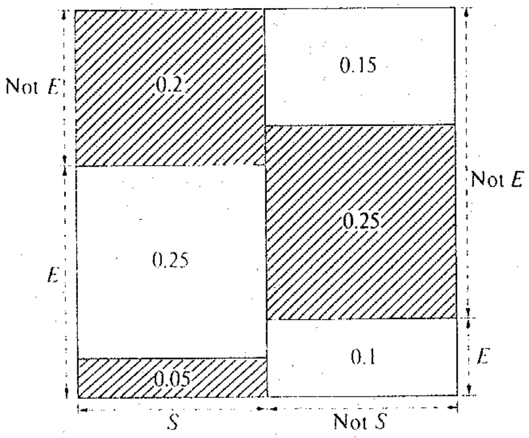
Why E_1 rather than $E_2, \ldots, E_{1000}$?
Because A.
$Prob\ (E_1) = \ldots = Prob\ (E_{10}) = 99/1000 = 0.099$
$Prob\ (E_{11}) = \ldots = Prob\ (E_{1000}) = 1/99{,}000 \doteq 0.00001$
$Prob\ (E_1/A) = 90/1000 = 0.090$
$Prob\ (E_2/A) = \ldots = Prob\ (E_{1000}/A) = 910/999{,}000 \simeq 0.001$

Before the answer, E_1 was a good candidate, but in no way distinguished from nine others; afterwards, it is head and shoulders above all its alternatives, but has itself lower probability than it had before.

I think this will remove some of the puzzlement felt in connection with Salmon's examples of explanations that lower the probability of what is explained. In Nancy Cartwright's example of the poison ivy ('Why is this plant alive?') the answer ('It was sprayed with defoliant') was statistically relevant, but did not redistribute the probabilities so as to favour the topic. The mere fact that the probability was lowered is, however, not enough to disqualify the answer as a telling one.

There is a further way in which A can provide information which favours the topic. This has to do with what is called Simpson's

Paradox; it is again Nancy Cartwright who has emphasized the importance of this for the theory of explanation (see n. 13 above). Here is an example she made up to illustrate it. Let H be 'Tom has heart disease'; S be 'Tom smokes'; and E, 'Tom does exercise'. Let us suppose the probabilities to be as follows:



Shaded areas represent the cases in which H is true, and numbers the probabilities. By the standard calculation, the conditional probabilities are

$Prob\ (H/S) = Prob\ (H) = \frac{1}{2}$
$Prob\ (H/S\&E) = \frac{1}{6}$
$Prob\ (H/E) = \frac{1}{8}$
$Prob\ (H/S\ \&\ \text{not}\ E) = 1$
$Prob\ (H/\ \text{not}\ E) = \frac{3}{4}$

In this example, the answer 'Because Tom smokes' does favour the topic that Tom has heart disease, in a straightforward (though derivative) sense. For as we would say it, the odds of heart disease increase with smoking regardless of whether he is an exerciser or a non-exerciser, and he must be one or the other.

Thus we should add to the account of what it is for A to favour B as against $C, \ldots, N$ that: if $Z = \{Z_1, \ldots, Z_n\}$ is a logical partition of explanatorily relevant alternatives, and A favours B as against $C, \ldots, N$ if any member of Z is added to our background information, then A does favour B as against $C, \ldots, N$.

We have now considered two sorts of evaluation: how probable is A itself? *and*, how much does A favour B as against $C, \ldots, N$? These are independent questions. In the second case, we know what aspects to consider, but do not have a precise formula that 'adds them all up'. Neither do we have a precise formula to weigh the importance of how likely the answer is to be true, against how favourable the information is which it provides. But I doubt the value of attempting to combine all these aspects into a single-valued measurement.

In any case, we are not finished. For there are relations among answers that go beyond the comparison of how well they do with respect to the criteria considered so far. A famous case, again related to Simpson's Paradox, goes as follows (also discussed in Cartwright's paper): at a certain university it was found that the admission rate for women was lower than that for men. Thus 'Janet is a woman' appears to tell for 'Janet was not admitted' as against 'Janet was admitted'. However, this was not a case of sexual bias. The admission rates for men and women for each department in the university were approximately the same. The appearance of bias was created because women tended to apply to departments with lower admission rates. Suppose Janet applied for admission in history; the statement 'Janet applied in history' *screens off* the statement 'Janet is a woman' from the topic 'Janet was not admitted' (in the Reichenbach–Salmon sense of 'screens off': P screens off A from B exactly if the probability of B given P and A is just the probability of B given P alone). It is clear then that the information that Janet applied in history (or whatever other department) is a much more telling answer than the earlier reply, in that it makes that reply irrelevant.

We must be careful in the application of this criterion. First, it is not important if some proposition P screens off A from B if P is not the core of an answer to the question. Thus if the why-question is a request for information about the mechanical processes leading up to the event, the answer is no worse if it is statistically screened off by other sorts of information. Consider 'Why is Peter dead?' answered by 'He received a heavy blow on the head' while we know already that Paul has just murdered Peter in some way. Secondly, a screened-off answer may be good but partial rather than irrelevant. (In the same example, we know that there must be some true proposition of the form 'Peter received a blow on the head with impact x', but that does not disqualify the answer, it only means that some more informative answer is possible.) Finally, in the case of a deterministic process in which state A_i, and no other state, is followed by state A_{i+1}, the best answers to the question 'Why is the system in state A_n at time t_n?' may all have the form 'Because the system was in state A_i at time t_i', but each such answer is screened off from the event described in the topic by some other, equally good answer. The most accurate conclusion is probably no more than that if one answer is screened off by another, and not conversely, then the latter is better in some respect.

When it comes to the evaluation of answers to why-questions, therefore, the account I am able to offer is neither as complete nor as precise as one might wish. Its shortcomings, however, are shared with the other philosophical theories of explanation I know (for I have drawn shamelessly on those other theories to marshal these criteria for answers). And the traditional main problems of the theory of explanation are solved not by seeing what these criteria are, but by the general theory that explanations are answers to why-questions, which are themselves contextually determined in certain ways.

§4.5 *Presupposition and Relevance Elaborated*

Consider the question 'Why does the hydrogen atom emit photons with frequencies in the general Balmer series (only)?' This question presupposes that the hydrogen atom emits photons with these frequencies. So how can I even ask that question unless I believe that theoretical presupposition to be true? Will my account of why-questions not automatically make scientific realists of us all?

But recall that we must distinguish carefully what a theory *says* from what we believe when we accept that theory (or rely on it to predict the weather or build a bridge, for that matter). The epistemic commitment involved in accepting a scientific theory, I have argued, is not belief that it is true but only the weaker belief that it is empirically adequate. In just the same way we must distinguish what the question says (i.e. *presupposes*) from what we believe when we ask that question. The example I gave above is a question which arises (as I have defined that term) in any context in which those hypotheses about hydrogen, and the atomic theory in question, are *accepted*. Now, when I ask the question, if I ask it seriously and in my own person, I imply that I believe that this question arises. But that means then only that my epistemic commitment indicated by, or involved in, the asking of this question, is exactly—no more and no less than—the epistemic commitment involved in my acceptance of these theories.

Of course, the discussants in this context, in which those theories are accepted, are conceptually immersed in the theoretical world-picture. They talk the language of the theory. The phenomenological distinction between objective or real, and not objective or unreal, is a distinction between what there is and what there is not which is drawn in that theoretical picture. Hence the questions asked are asked in the theoretical language—how could it be otherwise? But the epistemic commitment of the discussants is not to be read off from their language.

Relevance, perhaps the other main peculiarity of the why-question, raises another ticklish point, but for logical theory. Suppose, for instance, that I ask a question about a sodium sample, and my background theory includes present atomic physics. In that case the answer to the question may well be something like: because this material has such-and-such an atomic structure. Recalling this answer from one of the main examples I have used to illustrate the asymmetries of explanation, it will be noted that, *relative to* this background theory, my answer is a proposition necessarily equivalent to: because this material has such-and-such a characteristic spectrum. The reason is that the spectrum is unique—it identifies the material as having that atomic structure. But, here is the asymmetry, I could not well have answered the question by saying that this material has that characteristic spectrum.

These two propositions, one of them relevant and the other not, are equivalent relative to the theory. Hence they are true in exactly the same possible worlds allowed by the theory (less metaphysically put: true in exactly the same models of that theory). So now we have come to a place where there is a conflict with the simplifying hypothesis generally used in formal semantics, to the effect that propositions which are true in exactly the same possible worlds are identical. If one proposition is relevant and the other not, they cannot be identical.

We could avoid the conflict by saying that of course there are possible worlds which are not allowed by the background theory. This means that when we single out one proposition as relevant, in this context, and the other as not relevant and hence distinct from the first, we do so in part by thinking in terms of worlds (or models) regarded in this context as impossible.

I have no completely telling objection to this, but I am inclined to turn, in our semantics, to a different modelling of the language, and reject the simplifying hypothesis. Happily there are several sorts of models of language, not surprisingly ones that were constructed in response to other reflections on relevance, in which propositions can be individuated more finely. One particular sort of model, which provides a semantics for Anderson and Belnap's logic of tautological entailment, uses the notion of *fact*.[45] There one can say that

It is either raining or not raining
It is either snowing or not snowing

although true in exactly the same possible situations (namely, in all) are yet distinguishable through the consideration that today, for example, the first is *made true* by the fact that it is raining, and the second is made true by quite a different fact, namely, that it is not snowing. In another sort of modelling, developed by Alasdair Urquhart, this individuating function is played not by facts but by bodies of information.[46] And still further approaches, not necessarily tied to logics of the Anderson–Belnap stripe, are available.

In each case, the relevance relation among propositions will derive from a deeper relevance relation. If we use facts, for example, the relation *R* will derive from a request to the effect that the answer must provide a proposition which describes (is made true by) facts of a certain sort: for example, facts about atomic structure, or facts about this person's medical and physical history, or whatever.

§5. *Conclusion*

Let us take stock. Traditionally, theories are said to bear two sorts of relation to the observable phenomena: *description* and *explanation*. Description can be more or less accurate, more or less informative; as a minimum, the facts must 'be allowed by' the theory (fit some of its models), as a maximum the theory actually implies the facts in question. But in addition to a (more or less informative) description, the theory may provide an explanation. This is something 'over and above' description; for example, Boyle's law describes the relationship between the pressure, temperature, and volume of a contained gas, but does not explain it—kinetic theory explains it. The conclusion was drawn, correctly I think, that even if two theories are strictly empirically equivalent they may differ in that one can be used to answer a given request for explanation while the other cannot.

Many attempts were made to account for such 'explanatory power' purely in terms of those features and resources of a theory that make it informative (that is, allow it to give better descriptions). On Hempel's view. Boyle's law does explain these empirical facts about gases. but minimally. The kinetic theory is perhaps better *qua* explanation simply because it gives so much more information about the behaviour of gases. relates the three quantities in question to other observable quantities, has a beautiful simplicity, unifies our over-all picture of the world, and so on. The use of more sophisticated statistical relationships by Wesley Salmon and James Greeno (as well as by I. J. Good, whose theory of such concepts as weight of evidence, corroboration, explanatory power, and so on deserves more attention from philosophers), are all efforts along this line.[47] If they had succeeded, an empiricist could rest easy with the subject of explanation.

But these attempts ran into seemingly insuperable difficulties. The conviction grew that explanatory power is something quite irreducible, a special feature differing in kind from empirical adequacy and strength. An inspection of examples defeats any attempt to identify the ability to explain with any complex of those more familiar and down-to-earth virtues that are used to evaluate the theory *qua* description. Simultaneously it was argued that what science is really after is understanding, that this consists in being in a position to explain, hence what science is really after goes well beyond empirical adequacy and strength. Finally, since the theory's ability to explain provides a clear reason for accepting it, it was argued that explanatory power is evidence for the *truth* of the theory, special evidence that goes beyond any evidence we may have for the theory's empirical adequacy.

Around the turn of the century, Pierre Duhem had already tried to debunk this view of science by arguing that explanation is not an aim of science. In retrospect, he fostered that explanation–mysticism which he attacked. For he was at pains to grant that explanatory power does not consist in resources for description. He argued that only metaphysical theories explain, and that metaphysics is an enterprise foreign to science. But fifty years later, Quine having argued that there is no demarcation between science and philosophy, and the difficulties of the ametaphysical stance of the positivist-oriented philosophies having made a return to metaphysics tempting, one noticed that scientific activity does involve explanation, and Duhem's argument was deftly reversed.

Once you decide that explanation is something irreducible and special, the door is opened to elaboration by means of further concepts pertaining thereto, all equally irreducible and special. The premisses of an explanation have to include lawlike statements; a statement is lawlike exactly if it implies some non-trivial counterfactual conditional statement; but it can do so only by asserting relationships of necessity in nature. Not all classes correspond to genuine properties; properties and propensities figure in explanation. Not everyone has joined this return to essentialism or neo-Aristotelian realism, but some eminent realists have publicly explored or advocated it.

Even more moderate elaborations of the concept of explanation make mysterious distinctions. Not every explanation is a scientific explanation. Well then, that irreducible explanation-relationship comes in several distinct types, one of them being scientific. A scientific explanation has a special form, and adduces only special sorts of information to explain—information about causal connections and causal processes. Of course, a causal relationship is just what 'because' must denote; and since the *summum bonum* of science is explanation, science must be attempting even to describe something beyond the observable phenomena, namely causal relationships and processes.

These last two paragraphs describe the flights of fancy that become appropriate if explanation is a relationship *sui generis* between theory and fact. But there is no direct evidence for them at all, because if you ask a scientist to explain something to you, the information he gives you is not different in kind (and does not sound or look different) from the information he gives you when you ask for a description. Similarly in 'ordinary' explanations: the information I adduce to explain the rise in oil prices, is information I would have given you to a battery of requests for description of oil supplies, oil producers, and oil consumption. To call an explanation scientific, is to say nothing about its form or the sort of information adduced, but only that the explanation draws on science to get this information (at least to some extent) and, more importantly, that the criteria of evaluation of how good an explanation it is, are being applied using a scientific theory (in the manner I have tried to describe in Section 4 above).

The discussion of explanation went wrong at the very beginning when explanation was conceived of as a relationship like description: a relation between theory and fact. Really it is a three-term relation, between theory, fact, and context. No wonder that no single relation between theory and fact ever managed to fit more than a few examples! Being an explanation is essentially relative, for an explanation is an *answer*. (In just that sense, being a daughter is something relative: every woman is a daughter, and every daughter is a woman, yet being a daughter is not the same as being a woman.) Since an explanation is an answer, it is evaluated *vis-à-vis* a question, which is a request for information. But exactly what is requested, by means of the interrogative 'Why is it the case that *P*?', differs from context to context. In addition, the background theory plus data relative to which the question is evaluated, as arising or not arising, depends on the context. And even what part of that background information is to be used to evaluate how good the answer is, *qua* answer to that question, is a contextually determined factor. So to say that a given theory can be used to explain a certain fact, is always elliptic for: there is a proposition which is a telling answer, relative to this theory, to the request for information about certain facts (those counted as relevant for *this* question) that bears on a comparison between this fact which is the case, and certain (contextually specified) alternatives which are not the case.

So scientific explanation is not (pure) science but an application of science. It is a use of science to satisfy certain of our desires; and these desires are quite specific in a specific context, but they are always desires for descriptive information. (Recall: every daughter is a woman.) The exact content of the desire, and the evaluation of how well it is satisfied, varies from context to context. It is not a single desire, the same in all cases, for a very special sort of thing, but rather, in each case, a different desire for something of a quite familiar sort.

Hence there can be no question at all of explanatory power as such (just as it would be silly to speak of the 'control power' of a theory, although of course we rely on theories to gain control over nature and circumstances). Nor can there be any question of explanatory success as providing evidence for the truth of a theory that goes beyond any evidence we have for its providing an adequate

description of the phenomena. For in each case, a success of explanation is a success of adequate and informative description. And while it is true that we seek for explanation, the value of this search for science is that the search for explanation is *ipso facto* a search for empirically adequate, empirically strong theories.