and the organism to be kept alive by it in the extraordinary condition of lunar gravity.

There remains then only the problem of living with lots of 'mini-theories' in practice, as we actually do. Physiologists need not make relativistic corrections in their mechanical calculations, and can treat almost all processes deterministically (and some stochastically which physics implies to be near deterministic). Philosophy of science could do with a more accurate picture of this situation—it is the actual situation of the working scientist and may well harbour problems obscured by our preoccupation with global theories.[6] But there seems to me no doubt that the aim of empirical adequacy already requires the successive unification of 'mini-theories' into larger ones, and that the process of unification is mainly one of correction and not of conjunction.

§4. *Pragmatic Virtues and Explanation*

§4.1 *The Other Virtues*

When a theory is advocated, it is praised for many features other than empirical adequacy and strength: it is said to be mathematically elegant, simple, of great scope, complete in certain respects: *also* of wonderful use in unifying our account of hitherto disparate phenomena, and most of all, explanatory. Judgements of simplicity and explanatory power are the intuitive and natural vehicle for expressing our epistemic appraisal.[7] What can an empiricist make of these other virtues which go so clearly beyond the ones he considers pre-eminent?

There are specifically human concerns, a function of our interests and pleasures, which make some theories more valuable or appealing to us than others. Values of this sort, however, provide reasons for using a theory, or contemplating it, whether or not we think it true, and cannot rationally guide our epistemic attitudes and decisions. For example, if it matters more to us to have one sort of question answered rather than another, that is no reason to think that a theory which answers more of the first sort of questions is more likely to be true (not even with the proviso 'everything else being equal'). It is merely a reason to prefer that theory in another respect.

Nevertheless, in the analysis of the appraisal of scientific theories, it would be a mistake to overlook the ways in which that appraisal is coloured by contextual factors. These factors are brought to the

situation by the scientist from his own personal, social, and cultural situation. It is a mistake to think that the terms in which a scientific theory is appraised are purely hygienic, and have nothing to do with any other sort of appraisal, or with the persons and circumstances involved.

Theory acceptance has a pragmatic dimension. While the only belief involved in acceptance, as I see it, is the belief that the theory is empirically adequate, *more than belief is involved.* To accept a theory is to make a commitment, a commitment to the further confrontation of new phenomena within the framework of that theory, a commitment to a research programme, and a wager that all relevant phenomena can be accounted for without giving up that theory. That is why someone who has accepted a certain theory, will henceforth answer questions *ex cathedra*, or at least feel called upon to do so. Commitments are not true or false; they are vindicated or not vindicated in the course of human history.

Briefly, then, the answer is that the other virtues claimed for a theory are *pragmatic* virtues. In so far as they go beyond consistency, empirical adequacy, and empirical strength, they do not concern the relation between the theory and the world, but rather the use and usefulness of the theory; they provide reasons to prefer the theory independently of questions of truth.

Of course, this answer raises immediately the further question: why is this a *rational* procedure to follow in the appraisal of theories, in the deliberation that leads us to follow one approach rather than another in scientific research, or to commit ourselves epistemically, by accepting one theory rather than another?

I shall broach this question in the specific instance: why is it rational to pursue explanation? To answer this question fully we need an account of what explanation is—and I shall devote the next chapter to that. But beforehand, it is possible to sketch the answer which that account is meant to substantiate. It is this: the *epistemic* merits a theory may have or must have to figure in good explanations are not *sui generis*; they are just the merits it had in being empirically adequate, of significant empirical strength, and so forth. This does not mean that something is automatically a good explanation if it has those other merits; what more it needs is the pragmatic aspect of explanation. But in the pursuit of explanation we pursue *a fortiori* those more basic merits, which is what makes the pursuit of explanation of value to the scientific enterprise as such.

To praise a theory for its great explanatory power, is therefore to attribute to it *in part* the merits needed to serve the aim of science. It is not tantamount to attributing to it *special* features which make it more likely to be true, or empirically adequate. But it might be arguable that, for purely pragmatic (that is, person- and context-related) reasons, the pursuit of explanatory power is the best means to serve the central aims of science.

§4.2 *The Incursion of Pragmatics*

To spell out these contentions to the extent we can before we have an account of what explanation is, I must refer first of all to the terminology originally introduced by Charles Morris.[8] His basic concern was language, but we can transpose his concepts from words and statements to theories. In the study of language he saw three main levels: *syntax*, *semantics*, and *pragmatics*. The syntactic properties of an expression are determined only by its relations to other expressions, considered independently of meaning or interpretation. An example would be 'has six letters' which can be predicated of 'Cicero'. Semantic properties concern the relation of the expression to the world; an example is

1. 'Cicero' denotes Cicero.

Finally, pragmatics concerns the relation of the language to the users of that language; as in

2. Cicero preferred to be called 'Cicero' rather than 'Tully'.

In some sense, semantics is only an abstraction from pragmatics. It would not make sense to say 'I know that this man was named "Cicero" by his parents, and everyone always calls him that—but is his name really "Cicero"?' Yet we can study properties construed by abstracting from usage and its possible variations; this is merely one instance of scientific model building, in this case in the study of language.

But in certain cases, no abstraction is possible without losing the very thing we wish to study. How does the word 'I' differ from the word 'Cicero'? Exactly in the fact that the denotation of 'I' depends on who is using it—for every user uses it to refer to him or herself. Thus the semantic study of language can only go so far—then it must give way to a less thorough abstraction (that is, a less shallow level of analysis) and we find that we are doing pragmatics proper.

In the case of a statement, *truth* is the most important semantic property. A statement is true exactly if the actual world accords with this statement. But if some of the words, or grammatical devices, in that statement have a context-dependent semantic role, truth *simpliciter* does not make sense, and we must move again to pragmatics:

3. 'Cicero is dead' is true if and only if Cicero is dead.
4. In any context or occasion of language use, 'I am happy' is true if and only if the person who says it on that occasion, is happy at the time of saying it.

Syntactic properties of and relations among statements include those studied in traditional logic, for 'is a logical truth', 'is not self-contradictory', 'is deducible from' are there all syntactically definable for large, useful fragments of our language.

Turning now to theories, we find there also a threefold division of properties and relations. First there are the purely internal or logical ones, such as axiomatizability, consistency, and various sorts of completeness. Attempts have been made to locate simplicity on this level, but these, as all other attempts so far to explain precisely what people could possibly mean when they call a theory simple or simpler, have failed.

Simplicity is quite an instructive case. It is obviously a criterion in theory choice, or at least a term in theory appraisal. For that reason, some writings on the subject of induction suggest that simple theories are more likely to be true. But it is surely absurd to think that the world is more likely to be simple than complicated (unless one has certain metaphysical or theological views not usually accepted as legitimate factors in scientific inference). The point is that the virtue, or patchwork of virtues, indicated by the term is a factor in theory appraisal, but does not indicate *special* features that make a theory more likely to be true (or empirically adequate).

Semantic properties and relations are those which concern the theory's relation to the world, or more specifically, the facts about which it is a theory. Here the two main properties are truth and empirical adequacy. Hence this is the area where both realism and constructive empiricism locate a central aim of science.

Are there also philosophically significant pragmatic theoretical properties? The working language of science is no doubt context-dependent, but surely that is a practical point only? Scientific

theories can be stated in context-independent language, in what Quine calls 'eternal sentences'. So we do not need to stray into pragmatics, it would seem, to interpret science.

This may be true of those products of scientific activity which we call theories. It is not true of other parts of that activity, according to my view, and specifically I hold that

(a) the language of theory appraisal, and specifically the term 'explains' is radically context-dependent;
(b) the language of the use of theories to explain phenomena, is radically context-dependent.

These are two distinct points, for it is one thing to assert that Newton's theory explains the tides, and another to explain the tides by means of Newton's theory. For example, in doing the second you may never use the word 'explain'.

The pragmatics of language is also the place where we must locate such concepts as immersion in the language, or world-picture, of science. The basic factors in the linguistic situation, pragmatically conceived, are the speaker or user, the syntactic entity (sentence or set of sentences) uttered or displayed, the audience, and the factual circumstances. Any factor which relates to the speaker or audience is a pragmatic factor; and if it furthermore pertains specifically to that particular linguistic situation, a contextual factor. For example, the uttered word 'Cicero' may be discussed in isolation or in relation to the bearer of that name while still remaining on the level of abstraction properly called semantics. But the fact that it rather than 'Tully' was uttered, is a contextual factor. The fact that the speaker used it, on this occasion, to refer to his cat rather than to the senator, is also a contextual factor; that the speaker is a person in the habit of using the word that way, a pragmatic factor which may also play a role in this situation.

One such pragmatic or contextual factor may be a tacit agreement between speaker and audience (or unilateral commitment on the part of either) to be guided in his inferences by something more than bare logic. There may be a standing linguistic commitment, such as the educated layman's not to call anything table salt unless it is mainly sodium chloride, or the commitment, now falling into disuse, not to call anything cream unless it was produced by a cow. Such commitments may be more or less permanent or temporary. I know enough about astrology and about psychoanalysis to enter a

conversation with an *aficionado* of either, in which that theory is what guides the use of terms and the allowed inferences. More commonly, the discussion of a movie, say, Renoir's *Day in the Country*, may go that way: 'Do you think he really seduced her? No, in that milieu a kiss was an earthshaking event.' A certain suspension of disbelief, a momentary commitment to the world depicted by the theory, play, painting, or novel, determines *in that linguistic situation* what it is correct to say, and the correct way to say it. Robert Stalnaker has given the name *pragmatic presuppositions* to propositions which play this role of guiding assumptions.

The total immersion in the scientific world-picture, which is proper to situations in which science is pursued or used, is a case in point. I shall again return to this topic at the end of Chapter 6 when discussing the use of modal language in science.

§4.3 *Pursuit of Explanation*

Very strong claims are sometimes made for the centrality of explanation among the aims of science. In some cases, indeed, the demand for explanation is held up as overriding and not subject to qualification, as unlimited. Such an extreme ideal of explanatory completeness we found in the arguments for scientific realism examined earlier on. But even more moderate philosophers, less easily accused of metaphysical leanings, make far-reaching claims. Thus, Ernest Nagel:

> It is the desire for explanations which are at once systematic and controllable by factual evidence that generates science; and it is the organization and classification of knowledge on the basis of explanatory principles that is the distinctive goal of the sciences.[9]

None of this entails realism, and the first part is, I think, undoubtedly true. The second part might still conflict with the view that empirical adequacy is the pre-eminent virtue, depending on how 'explanatory' is understood. But if we look to how Nagel understands explanation, we find that he holds to an account that is rather like Hempel's (to be examined in the next chapter). Let us call whatever Nagel understands to be explanation, *N-explanation*. Then Nagel here reported his conviction that the distinctive goal of the sciences is *N*-explanation. This may well be true even if the search for explanation is totally explicable as being of value to science because it serves the aim of giving us empirically adequate and strong theories. That this is not a far-fetched construal, seems to me clear from the following passage

on the same page, where he gives as main example for his contention that a few principles formulated by Newton

> suffice to show that propositions concerning the moon's motion, the behavior of the tides, the paths of projectiles, and the rise of liquids in thin tubes *are intimately related*, and that all these propositions *can be rigorously deduced* from those principles conjoined with various special assumptions of fact.

This certainly does not contradict the idea that the name of the game is saving the phenomena, even while there is a strong flavour of that distinctive satisfaction the human mind finds in encompassing an elegant, tightly and coherently constructed theory in order to win in that game.

There is a totally false issue that tends to be brought up in this connection—and was, for instance, by Paul Feyerabend.[10] Let us suppose for a moment that what more there is to explanation is merely a function of human interests. Then scientists need not worry unduly about achieving explanation over and above empirical adequacy. They can stop when they believe that they have *that*. However, in the history of science it is clear that scientists would have been ill advised to be so sanguine. The search for a dynamics compatible with Copernicus's new astronomical scheme, the search for the details of the atomic structure that would explain discrete spectra, the pursuit of the kinetic theory even when phenomenological thermodynamics seemed entirely adequate—there are many examples in which the search for explanation paid off handsomely. So only Realism is a philosophy that stimulates scientific inquiry; anti-realism hampers it.

Paid off handsomely, how? Paid off in new theories we have more reason to believe empirically adequate. But in that case even the anti-realist, when asked questions about *methodology* will *ex cathedra* counsel the search for explanation! We might even suggest a loyalty oath for scientists, if realism is so efficacious. In any case, the criticism is based on a very naïve view of scientific certainty; there always have been reasons to doubt the empirical adequacy of extant theories and these were reasons operative in the cited examples of 'searches for explanation'.

I call this a false issue, for the interpretation of science, and the correct view of its methodology, are two separate topics. But I have sketched *en passant* my answer to the question about methodology: the search for explanation is valued in science because it consists *for the most part* in the search for theories which are simpler, more

unified, and more likely to be empirically adequate. This is not because explanatory power is a separate quality *sui generis* which, mysteriously, makes those other qualities more likely, but because having a good explanation *consists* for the most part in having a theory with those other qualities.

To see whether explanation really is pre-eminent among the theoretical virtues sought in science, we should gauge how it is regarded in competition with other virtues.

First, there are rock-bottom criteria of minimal acceptability: consistency, internally and with the facts. Cases are known of mathematically inconsistent theories (Dirac introduced a function at one point, which was very useful, but later shown to be an impossible one) but that is a defect that must be repaired. You cannot advocate a theory as correct and inconsistent. Inconsistency with the observed facts is similarly minimal; if the theory conflicts with any previously acceptable data, we must either change the theory or deny that those data are correct.

Explanation is not a rock-bottom minimal virtue of this sort. If explanation of the facts were required in the way consistency with the facts is, then every theory would have to explain every fact in its domain. Newton would have had to add an explanation of gravity to his celestial mechanics before presenting it at all. But instead he says:

> Hitherto we have explained the phaenomena of the heavens and of our sea, by the power of Gravity, but have not yet assign'd a cause of this power ... hitherto I have not been able to discover the cause of those properties of gravity from phaenomena, and I frame no hypotheses. For whatever is not deduc'd from the phaenomena, is to be called an hypothesis; and hypotheses, whether metaphysical or physical, whether of occult qualities or mechanical, have no place in experimental philosophy ... And to us it is enough, that gravity does really exist, and act according to the laws which we have explained, and abundantly serves to account for all the motions of the celestial bodies, and of our sea.[11]

His reasons are perhaps stronger than many would accept, but it remains that he can decline to explain whereas he couldn't very well decline to be consistent. Another example is that Newton did decline, clearly, to satisfy a criterion of Kepler for the adequacy of any theory of the heavens: that it should explain why there are exactly six planets.[12]

The question is then whether explanatory power would be preferred over other virtues when there is a conflict. This can surely

not be so when the other virtue is the one the non-realist regards as the highest—empirical adequacy. For to forgo empirical adequacy is to allow that there may arise inconsistencies with observed facts. That possibility we cannot allow while advocating a theory as correct. Indeed, empirical adequacy is a precondition: we don't say that we have an explanation unless we have an *acceptable* theory which explains.

Thirdly, we may ask whether explanation is a pre-eminent virtue in the sense of being required when it can be had. This would mean that if several theories were empirically equivalent, the one which explains most would have to be accepted. Against this idea count all the examples of scientists refusing to enlarge their theories in ways that do not yield different (or further) empirical consequences. One example is already given by the passage I quoted from Newton, but another instructive example is yet another feature of the discussion of hidden variables for quantum mechanics.

According to quantum theory there are correlations in the behaviour of particles which have interacted in the past, but are now physically separated. No causal mechanism is given to explain these correlations, which were dramatized in a famous paper by Einstein, Podolski, and Rosen. Various experiments have borne out the existence of these correlations, which may be found for instance if an atom 'cascading down' from an excited state, emits two photons. When polarization filters are set up for these photons to pass through, it is as if each photon 'knows' whether the other photon passed the other filter.

Certain hidden variable theories have been proposed which would explain such correlations (so-called 'hidden variable theories of the second kind').[13] These do not predict exactly the same correlations—this is what makes these theories interesting to physics. So far, experiments appear to support quantum theories against those rivals. But the one response which is conspicuous by its absence is that an explanation of the correlations *must be found* which fits in exactly with quantum theory and does not affect its empirical content at all. Such metaphysical extensions of the theory (if indeed possible) would be philosophical playthings only. There are only two camps to the debate as far as physics is concerned: either this non-locality makes quantum theory pre-eminently suited to the representation of the world (and we need to re-school our imaginations), or else quantum theory must be replaced by an empirically significant *rival*.

In none of the three senses examined is explanation an overriding virtue. Philosophy fathered the sciences, and philosophy aims pre-eminently, and perhaps only, to remove wonder, as Aristotle said; but these children have left the parental home.

5

The Pragmatics of Explanation[1]

> If cause were non-existent everything would have been produced by everything and at random. Horses, for instance, might be born, perchance, of flies, and elephants of ants: and there would have been severe rains and snow in Egyptian Thebes, while the southern districts would have had no rain, unless there had been a cause which makes the southern parts stormy, the eastern dry.
>
> Sextus Empiricus, *Outlines of Pyrrhonism* III, V, 1

A THEORY is said to have explanatory power if it allows us to explain; and this is a virtue. It is a pragmatic virtue, albeit a complex one that includes other virtues as its own preconditions. After some preliminaries in Section 1, I shall give a frankly selective history of philosophical attempts to explain explanation. Then I shall offer a model of this aspect of scientific activity in terms of why-questions, their presuppositions, and their context-dependence. This will account for the puzzling features (especially asymmetries and rejections) that have been found in the phenomenon of explanation, while remaining compatible with empiricism.

§1. *The Language of Explanation*

One view of scientific explanation is encapsulated in this argument: science aims to find explanations, but nothing is an explanation unless it is true (explanation requires true premisses); so science aims to find true theories about what the world is like. Hence scientific realism is correct. Attention to other uses of the term 'explanation' will show that this argument trades on an ambiguity.

§1.1 *Truth and Grammar*

It is necessary first of all to distinguish between the locutions 'we have an explanation' and 'this theory explains'. The former can be paraphrased 'we have a theory that explains'—but then 'have' needs

to be understood in a special way. It does not mean, in this case, 'have on the books', or 'have formulated', but carries the conversational implicature that the theory tacitly referred to is acceptable. That is, you are not warranted in saying 'I have an explanation' unless you are warranted in the assertion 'I have a theory *which is acceptable* and which explains'. The important point is that the mere statement 'theory T explains fact E' does not carry any such implication: not that the theory is true, not that it is empirically adequate, and not that it is acceptable.

There are many examples, taken from actual usage, which show that truth is not presupposed by the assertion that a theory explains something. Lavoisier said of the phlogiston hypothesis that it is too vague and consequently 's'adapte à toutes les explications dans lesquelles on veut le faire entrer'.[2] Darwin explicitly allows explanations by false theories when he says 'It can hardly be supposed that a false theory would explain, in so satisfactory a manner as does the theory of natural selection, the several large classes of facts above specified.'[3] Gilbert Harman, we recall, has argued similarly: that a theory explains certain phenomena is part of the evidence that leads us to accept it. But that means that the explanation-relation is visible before we believe that the theory is true. Finally, we criticize theories selectively: a discussion of celestial mechanics around the turn of the century could surely contain the assertion that Newton's theory does explain many planetary phenomena. Yet it was also agreed that the advance in the perihelion of Mercury seems to be inconsistent with the theory, suggesting therefore that the theory is not empirically adequate—and hence, is false—without this agreement undermining the previous assertion. Examples can be multiplied: Newton's theory explained the tides, Huygens's theory explained the diffraction of light, Rutherford's theory of the atom explained the scattering of alpha particles, Bohr's theory explained the hydrogen spectrum, Lorentz's theory explained clock retardation. We are quite willing to say all this, although we will add that, for each of these theories, phenomena were discovered which they could not only not explain, but could not even accommodate in the minimal fashion required for empirical adequacy.

Hence, to say that a theory explains some fact or other, is to assert a relationship between this theory and that fact, which is independent of the question whether the real world, as a whole, fits that theory.

Let us relieve the tedium of terminological discussion for a moment and return to the argument displayed at the beginning. In view of the distinctions shown, we can try to revise it as follows: science tries to place us in a position in which we have explanations, and are warranted in saying that we do have. But to have such warrant, we must first be able to assert with equal warrant that the theories we use to provide premisses in our explanations are true. Hence science tries to place us in a position where we have theories which we are entitled to believe to be true.

The conclusion may be harmless of course if 'entitled' means here only that one can't be convicted of irrationality on the basis of such a belief. That is compatible with the idea that we have warrant to believe a theory only because, and in so far as, we have warrant to believe that it is empirically adequate. In that case it is left open that one is at least as rational in believing merely that the theory is empirically adequate.

But even if the conclusion were construed in this harmless way, the second premiss will have to be disputed, for it entails that someone who merely accepts the theory as empirically adequate, is not in a position to explain. In this second premiss, the conviction is perhaps expressed that having an explanation is not to be equated with having an acceptable theory that explains, but with having a true theory that explains.

That conviction runs afoul of the examples I gave. I say that Newton could explain the tides, that he had an explanation of the tides, that he did explain the tides. In the same breath I can add that this theory is, after all, not correct. Hence I would be inconsistent if by the former I meant that Newton had a true theory which explained the tides—for if it was true then, it is true now. If what I meant was that it was true *then* to say that Newton had an acceptable theory which explains the tides, that would be correct.

A realist can of course give his own version: to have an explanation means to have 'on the books' a theory which explains and which one is entitled to believe to be true. If he does so, he will agree that to have an explanation does not require a true theory, while maintaining his contention that science aims to place us in a position to give *true* explanations. That would bring us back, I suppose, to our initial disagreement, the detour through explanation having brought no benefits. If you can only be entitled to assert that the theory is true because, and in so far as, you are entitled to assert that it is

empirically adequate, then the distinction drawn makes no practical difference. There would of course be a difference between *believe* (to-be-true) and *accept* (believe-to-be-empirically-adequate) but no real difference between be-entitled-to-believe and be-entitled-to-accept. A realist might well dispute this by saying that if the theory explains facts then that gives you an *extra* good reason (over and above any evidence that it is empirically adequate) to believe that the theory is true. But I shall argue that this is quite impossible, since explanation is not a special additional feature that can give you good reasons for belief in addition to evidence that the theory fits the observable phenomena. For 'what more there is to' explanation is something quite pragmatic, related to the concerns of the user of the theory and not something new about the correspondence between theory and fact.

So I conclude that (a) the assertion that theory T explains, or provides an explanation for, fact E does not presuppose or imply that T is true or even empirically adequate, and (b) the assertion that we have an explanation is most simply construed as meaning that we have 'on the books' an acceptable theory which explains. I shall henceforth adopt this construal.

To round off the discussion of the terminology, let us clarify what sorts of terms can be the grammatical subjects, or grammatical objects, of the verb 'to explain'. Usage is not regimented: when we say 'There is the explanation!', we may be pointing to a fact, or to a theory, or to a thing. In addition, it is often possible to point to more than one thing which can be called 'the explanation'. And, finally, whereas one person may say that Newton's theory of gravitation explained the tides, another may say that Newton used that theory to explain the tides. (I suppose no one would say that the hammer drove the nail through the wood; only that the carpenter did so, using the hammer. But today people do sometimes say that the computer calculated the value of a function, or solved the equations, which is perhaps similar to saying that the theory explained the tides.)

This bewildering variety of modes of speech is common to scientists as well as philosophers and laymen. In Huygens and Young the typical phrasing seemed to be that phenomena may be explained *by means of* principles, laws, and hypotheses, or *according* to a view.[4] On the other hand, Fresnel writes to Arago in 1815 'tous ces phénomènes... sont réunis et expliqués par la même théorie des vibrations',

and Lavoisier says that the oxygen hypothesis he proposes *explains* the phenomena of combustion.[5] Darwin also speaks in the latter idiom: 'In scientific investigations it is permitted to invent any hypothesis, and if it explains various large and independent classes of facts it rises to the rank of a well-grounded theory'; though elsewhere he says that the facts of geographical distribution are *explicable on* the theory of migration.[6]

In other cases yet, the theory assumed is left tacit, and we just say that one fact explains another. For example: the fact that water is a chemical compound of oxygen and hydrogen explains why oxygen and hydrogen appear when an electric current is passed through (impure) water.

To put some order into this terminology, and in keeping with previous conclusions, we can regiment the language as follows. The word 'explain' can have its basic role in expressions of form 'fact E explains fact F relative to theory T'. The other expressions can then be parsed as: 'T explains F' is equivalent to: 'there are facts which explain F relative to T'; 'T was used to explain F' equivalent to 'it was shown that there are facts which explain F relative to T'; and so forth. Instead of 'relative to T' we can sometimes also say 'in T'; for example, 'the gravitational pull of the moon explains the ebb and flow of the tides in Newton's theory'.

After this, my concern will no longer be with the derivative type of assertion that we *have* an explanation. After this point, the topic of concern will be that basic relation of explanation, which may be said to hold between facts relative to a theory, quite independently of whether the theory is true or false, believed, accepted, or totally rejected.

§1.2 *Some Examples*

Philosophical discussion is typically anchored to its subject through just a few traditional examples. The moment you see 'Pegasus', 'the king of France', or 'the good Samaritan', in a philosophical paper, you know exactly to what problem area it belongs. In the philosophical discussion of explanation, we also return constantly to a few main examples: paresis, the red shift, the flagpole. To combat the increasing sense of unreality this brings, it may be as well to rehearse, briefly, some workaday examples of scientific explanation.

(1) Two kilograms of copper at 60° C are placed in three kilograms of water at 20° C. After a while, water and copper reach the same

temperature, namely 22.5° C, and then cool down together to the temperature of the surrounding atmosphere.

There are a number of facts here for which we may request an explanation. Let us just ask why the equilibrium temperature reached is 22.5° C.

Well, the specific heats of water and copper are 1 and 0.1, respectively. Hence if the final temperature is T, the copper loses $0.1 \times 2 \times (60 - T)$ units of heat and the water gains $1 \times 3 \times (T - 20)$. At this point we appeal to the principle of the Conservation of Energy, and conclude that the total amount of heat neither increased nor diminished. Hence,

$$0.1 \times 2 \times (60 - T) = 1 \times 3 \times (T - 20)$$

from which $T = 22.5$ can easily be deduced.

(2) A short circuit in a power station results in a momentary current of 10^6 amps. A conductor, horizontally placed, 2 metres in length and 0.5 kg in mass, is warped at that time.

Let us ask why the conductor was warped. Well, the earth's magnetic field at this point is not negligible; its vertical component is approximately $5/10^5$ tesla. The theory of electro-magnetism allows us to calculate the force exerted on the conductor at the time in question:

$$(5/10^5) \times 2 \times 10^6 = 100 \text{ newtons}$$

which is directed at right angles to the conductor in the horizontal plane. The second law of Newton's mechanics entails in turn that, at that moment, the conductor has an acceleration of

$$100 \div 0.5 = 200 \text{ m/sec}^2$$

which is approximately twenty times the downward acceleration attributable to gravity (9.8 m/sec^2)—which allows us to compare in concrete terms the effect of the short circuit on the fixed conductor, and the normal effect of its weight.

(3) In a purely numerological way, Balmer, Lyman, and Paschen constructed formulae fitting frequency series to be found in the hydrogen spectrum, of the general form:

$$f_m^n = R\left(\frac{1}{m^2} - \frac{1}{n^2}\right)$$

where Balmer's law had $m = 2$, Lyman's had $m = 1$, and Paschen's $m = 3$; both m and n range over natural numbers.

Bohr's theory of the atom explains this general form. In this theory, the electron in a hydrogen atom moves in a stable orbit, characterized by an angular momentum which is an integral multiple of $h/2\pi$. The associated energy levels take the form

$$E_n = -E_o/n^2$$

where E_o is called the ground state energy.

When the atom is excited (as when the sample is heated), the electron jumps into a higher energy state. It then spontaneously drops down again, emitting a photon with energy equal to the energy lost by that electron in its drop. So if the drop is from level E_n to level E_m, the photon's energy is

$$\begin{aligned} E = E_n - E_m &= (-E_o/n^2) - (-E_o/m^2) \\ &= E_o/m^2 - E_o/n^2 \end{aligned}$$

The frequency is related to the energy by the equation

$$E = hf$$

so the frequencies exhibited by the emitted photons are

$$f_m^n = \frac{E}{h} = \frac{E_o}{h}\left(\frac{1}{m^2} - \frac{1}{n^2}\right)$$

which is exactly of the general form found above, with E_o/h being the constant R.

The reader may increase this stock of examples by consulting elementary texts and the *Science Digest*. It should be clear at any rate that scientific theories are used in explanation, and that how well a theory is to be regarded, depends at least in part on how much it can be used to explain.

§2. *A Biased History*

Current discussion of explanation draws on three decades of debate, which began with Hempel and Oppenheim's 'Studies in the Logic of Explanation' (1948).[7] The literature is now voluminous, so that a retrospective must of necessity be biased. I shall bias my account in such a way that it illustrates my diagnoses of the difficulties and points suggestively to the solution I shall offer below.

§2.1 *Hempel: Grounds for Belief*

Hempel has probably written more papers about scientific explanation than anyone; but because they are well known I shall focus on

the short summary which he gave of his views in 1966.[8] There he lists two criteria for what is an explanation:

explanatory relevance: 'the explanatory information adduced affords good grounds for believing that the phenomenon did, or does, indeed occur.'

testability: 'the statements constituting a scientific explanation must be capable of empirical test.'

In each explanation, the information adduced has two components, one ('the laws') information supplied by a theory, and the other ('the initial or boundary conditions') being auxiliary factual information. The relationship of providing good grounds is explicated separately for statistical and non-statistical theories. In the latter, the information *implies* the fact that is explained; in the former, it *bestows high probability* on that fact.

As Hempel himself points out, the first criterion does not provide either sufficient or necessary conditions for explanation. This was established through a series of examples given by various writers (but especially Michael Scriven and Sylvain Bromberger) and which have passed into the philosophical folklore.

First, giving good grounds for belief does not always amount to explanation. This is most strikingly apparent in examples of the asymmetry of explanation. In such cases, two propositions are strictly equivalent (relative to the accepted background theory), and the one can be adduced to explain why the other is the case, but not conversely. Aristotle already gave examples of this sort (*Posterior Analytics*, Book I, Chapter 13). Hempel mentions the phenomenon of the *red shift*: relative to accepted physics, the galaxies are receding from us if and only if the light received from them exhibits a shift toward the red end of the spectrum. While the receding of the galaxies can be cited as the reason for the red shift, it hardly makes sense to say that the red shift is the reason for their motion. A more simple-minded example is provided by the *barometer*, if we accept the simplified hypothesis that it falls exactly when a storm is coming, yet does not explain (but rather, is explained by) the fact that a storm is coming. In both examples, good grounds of belief are provided by either proposition for the other. The flagpole is perhaps the most famous asymmetry. Suppose that a flagpole, 100 feet high, casts a shadow 75 feet long. We can explain the length of the shadow by noting the angle of elevation of the sun, and appealing to the accepted theory that light travels in straight lines. For given that angle, and the height of the pole, trigonometry enables us to

deduce the length of the base of the right-angled triangle formed by pole, light ray, and shadow. However, we can similarly deduce the length of the pole from the length of the shadow plus the angle of elevation. Yet if someone asks us why the pole is 100 feet high, we cannot explain that fact by saying 'because it has a shadow 75 feet long'. The most we could explain that way is how we *came to know*, or how he might himself verify the claim, that the pole is indeed so high.

Second, not every explanation is a case in which good grounds for belief are given. The famous example for this is *paresis*: no one contracts this dreadful illness unless he had latent, untreated syphilis. If someone asked the doctor to explain to him why he came down with this disease, the doctor would surely say: 'because you had latent syphilis which was left untreated'. But only a low percentage of such cases are followed by paresis. Hence if one knew of someone that he might have syphilis, it would be reasonable to warn him that, if left untreated, he might contract paresis—but not reasonable to expect him to get it. Certainly we do not have here the high probability demanded by Hempel.

It might be replied that the doctor has only a partial explanation, that there are further factors which medical science will eventually discover. This reply is based on faith that the world is, for macroscopic phenomena at least, deterministic or nearly so. But the same point can be made with examples in which we do not believe that there is further information to be had, even in principle. The half-life of uranium U^{238} is $(4.5) \cdot 10^9$ years. Hence the probability that a given small enough sample of uranium will emit radiation in a specified small interval of time is low. Suppose, however, that it does. We still say that atomic physics explains this, the explanation being that this material was uranium, which has a certain atomic structure, and hence is subject to spontaneous decay. Indeed, atomic physics has many more examples of events of very low probability, which are explained in terms of the structure of the atoms involved. Although there are physicists and philosophers who argue that the theory must therefore be incomplete (one of them being Einstein, who said 'God does not play with dice') the prevalent view is that it is a contingent matter whether the world is ultimately deterministic or not.

In addition to the above, Wesley Salmon raised the vexing problem of *relevance* which is mentioned in the title of the first criterion,

but does not enter into its explication. Two examples which meet the requirements of providing good grounds are:

John Jones was almost certain to recover from his cold because he took vitamin C, and almost all colds clear up within a week of taking vitamin C.

John Jones avoided becoming pregnant during the past year, for he has taken his wife's birth control pills regularly, and every man who takes birth control pills avoids pregnancy.[9]

Salmon assumed here that almost all colds spontaneously clear up within a week. There is then something seriously wrong with these 'explanations', since the information adduced is wholly or partly irrelevant. So the criterion would have to be amended at least to read: 'provides good and *relevant* grounds'. This raises the problem of explicating relevance, also not an easy matter.

The second criterion, of testability, is met by all scientific theories, and by all the auxiliary information adduced in the above examples, so it cannot help to ameliorate these difficulties.

§2.2 *Salmon: Statistically Relevant Factors*

A number of writers adduced independent evidence for the conclusion that Hempel's criterion is too strong. Of these I shall cite three. The first is Morton Beckner, in his discussion of evolution. This is not a deterministic theory, and often explains a phenomenon only by showing how it could have happened—and indeed, might well have happened in the presence of certain describable, believable conditions consistent with the theory.

Selectionists have devoted a great deal of effort to the construction of models that are aimed at demonstrating that some observed or suspected phenomena are possible, that is, that they are compatible with the established or confirmed biological hypotheses ... These models all state strongly that if conditions were (or are) so and so, then, the laws of genetics being what they are, the phenomena in question must occur.[10]

Thus evolution theory explains, for example, the giraffe's long neck, although there was no independent knowledge of food shortages of the requisite sort. Evolutionists give such explanations by constructing models of processes which utilize only genetic and natural selection mechanisms, in which the outcome agrees with the actual phenomena.

In a similar vein, Putnam argued that Newton's explanations were *not* deductions of the facts that had to be explained, but rather

demonstrations of compatibility. What was demonstrated was that celestial motions could be as they were, given the theory and certain possible mass distributions in the universe.[11]

The distinction does not look too telling as long as we have to do with a deterministic theory. For in that case, the phenomena E are *compatible* with theory T if and only if there are possible preceding conditions C such that C plus T *imply* E. In any case, deduction and merely logical consistency cannot be what is at issue, since to show that T is logically compatible with E it would suffice to show that T is irrelevant to (has nothing to say about) E—surely not sufficient for explanation.

What Beckner and Putnam are pointing to are demonstrations that tend to establish (or at least remove objections to) claims of empirical adequacy. It is shown that the development of the giraffe's neck, or the fly-whisk tail fits a model of evolutionary theory; that the observed celestial motions fit a model of Newton's celestial mechanics. But a claim of empirical adequacy does not amount to a claim of explanation—there must be more to it.

Wesley Salmon introduced the theory that an explanation is not an argument, but an assembly of statistically relevant factors. A fact A is statistically relevant to a phenomenon E exactly if the probability of E *given* A is different from the probability of E *simpliciter*:

$$P(E/A) \neq P(E)$$

Hempel's criterion required $P(E/A)$ to be high (at least greater than $\frac{1}{2}$). Salmon does not require this, and he does not even require that the information A increases the probability of E. That Hempel's requirement was too strong, is shown by the paresis example (which fits Salmon's account very well), and that $P(E/A)$ should not be required to be higher than $P(E)$ Salmon argues independently.

He gives the example of an equal mixture of Uranium-238 atoms and Polonium-214 atoms, which makes the Geiger counter click in interval $(t, t+m)$. This means that one of the atoms disintegrated. Why did it? The correct answer will be: because it was a Uranium-238 atom: if that is so –although the probability of its disintegration is much higher relative to the previous knowledge that the atom belonged to the described mixture.[12] The problem with this argument is that, on Salmon's criterion, we can explain not only why there was a disintegration, but also why the disintegration occurred, let us say, exactly half-way between t and $t+m$. For the information

is statistically relevant to that occurrence. Yet would we not say that this is the sort of fact that atomic physics leaves unexplained?

The idea behind this objection is that the information is statistically relevant to the occurrence at $t+(m/2)$, but does not favour that as against various other times in the interval. Hence, if $E=$(a disintegration occurred) and $E_x=$(a disintegration occurred at time x), then Salmon bids us compare $P(E_x)$ with $P(E_x/A)$, whereas we naturally compare *also* $P(E_x/A)$ with $P(E_y/A)$ for other times y. This suggests that mere statistical relevance is not enough.

Nancy Cartwright has provided several examples to show that Salmon's criterion of statistical relevance also does not provide necessary or sufficient conditions for explanation.[13] As to sufficiency, suppose I spray poison ivy with defoliant which is 90 per cent effective. Then the question 'Why is *this* poison ivy now dead?' may correctly be answered 'Because it was sprayed with the defoliant.' About 10 per cent of the plants are still alive, however, and for those it is true that the probability that they are still alive was not the same as the probability that they are still alive *given* that they were sprayed. Yet the question 'Why is *that* plant now alive?' cannot be correctly answered 'Because it was sprayed with defoliant.'

Nor is the condition necessary. Suppose, as a medical fiction, that paresis can result from either syphilis or epilepsy, and from nothing else, and that the probability of paresis given either syphilis or epilepsy equals 0.1. Suppose in addition that Jones is known to belong to a family of which every member has either syphilis or epilepsy (but, fortunately, not both), and that he has paresis. Why did *he* develop this illness? Surely the best answer *either* is 'Because he had syphilis' *or* is 'Because he had epilepsy', depending on which of these is true. Yet, with all the other information we have, the probability that Jones would get paresis is already established as 0.1, and this probability is not changed if we are told in addition, say, that he has a history of syphilis. The example is rather similar to that of the Uranium and Polonium atoms, except that the probabilities are equal—and we still want to say that in *this* case, *the* explanation of the paresis is the fact of syphilis.

Let me add a more general criticism. It would seem that if either Hempel's or Salmon's approach was correct, then there would not really be more to explanatory power than empirical adequacy and empirical strength. That is, on these views, explaining an observed event is indistinguishable from showing that this event's occurrence

does not constitute an objection to the claim of empirical adequacy for one's theory, and in addition, providing significant information entailed by the theory and relevant to that event's occurrence. And it seems that Salmon, at that point, was of the opinion that there really cannot be more to explanation:

> When an explanation ... has been provided, we know exactly how to regard any A with respect to the property B ... We know all the regularities (universal or statistical) that are relevant to our original question. What more could one ask of an explanation?[14]

But in response to the objections and difficulties raised, Salmon, and others, developed new theories of explanation according to which there is more to explanatory power. I shall examine Salmon's later theory below.

§2.3 *Global Properties of Theories*

To have an explanation of a fact is to have (accepted) a theory which is acceptable, and which explains that fact. The latter relation must indubitably depend on what that fact is, since a theory may explain one fact and not another. Yet the following may also be held: it is a necessary condition that the theory, considered as a whole, has certain features beyond acceptability. The relation between the theory and *this* fact may be called a *local* feature of the theory, and characters that pertain to the theory taken as a whole, *global* features.

This suggestive geometric metaphor was introduced by Michael Friedman, and he attempted an account of explanation along these lines. Friedman wrote:

> On the view of explanation that I am proposing, the kind of understanding provided by science is global rather than local. Scientific explanations do not confer intelligibility on individual phenomena by showing them to be somehow natural, necessary, familiar, or inevitable. However, our overall understanding of the world is increased...[15]

This could be read as totally discounting a specific relation of explanation altogether, as saying that theories can have certain overall virtues, at which we aim, and because of which we may ascribe to them explanatory power (with respect to their primary domain of application, perhaps). But Friedman does not go quite as far. He gives an explication of the relation *theory T explains phenomenon P*. He supposes (p. 15) that phenomena, i.e. general uniformities, are represented by lawlike sentences (whatever those may be); that

we have as background a set K of accepted lawlike sentences, and that the candidate S (law, theory, or hypothesis) for explaining P is itself representable by a lawlike sentence. His definition has the form:

> S explains P exactly if P is a consequence of S, relative to K, and S 'reduces' or 'unifies' the set of its own consequences relative to K.

Here A is called a consequence of B relative to K exactly if A is a consequence of B and K together. He then modifies the above formula, and explicates it in a technically precise way. But as he explicates it, the notion of reduction cannot do the work he needs it to do, and it does not seem that anything like his precise definition could do.[16] More interesting than the details, however, is the form of the intuition behind Friedman's proposal. According to him, we evaluate something as an explanation relative to an assumed background theory K. I imagine that this theory might actually include some auxiliary information of a non-lawlike character, such as the age of the universe, or the boundary conditions in the situation under study. But of course K could not very well include all our information, since we generally know that P when we are asking for an explanation of P. Secondly, relative to K, the explanation implies that P is true. In view of Salmon's criticisms, I assume that Friedman would wish to weaken this Hempel-like condition. Finally, and here is the crux, it is the character of K plus the adduced information together, regarded as a complex theory, that determines whether we have an explanation. And the relevant features in this determination are global features, having to do with all the phenomena covered, not with P as such. So, whether or not K plus the adduced information provides new information about facts other than those described in P, appears to be crucial to whether we have an explanation of P.

James Greeno has made a similar proposal, with special reference to statistical theories. His abstract and closing statement says:

> The main argument of this paper is that an evaluation of the overall explanatory power of a theory is less problematic and more relevant as an assessment of the state of knowledge than an evaluation of statistical explanations of single occurrences ...[17]

Greeno takes as his model of a theory one which specifies a single probability space Q as the correct one, plus two partitions (or random variables) of which one is designated *explanandum* and the other

explanans. An example: sociology cannot explain why Albert, who lives in San Francisco and whose father has a high income, steals a car. Nor is it meant to. But it does explain delinquency in terms of such other factors as residence and parental income. The degree of explanatory power is measured by an ingeniously devised quantity which measures the information I the theory provides of the *explanandum* variable M on the basis of *explanans* S. This measure takes its maximum value if all conditional probabilities $P(M_i/S_j)$ are zero or one (D-N case), and its minimum value zero if S and M are statistically independent.

But it is not difficult to see that Greeno's way of making these ideas precise still runs into some of the same old difficulties. For suppose S and M describe the behaviour of barometers and storms. Suppose that the probability that the barometer will fall (M_1) equals the probability that there will be a storm (S_1), namely 0.2, and that the probability that there is a storm *given* that the barometer falls equals the probability that the barometer falls *given* that there will be a storm, namely 1. In that case the quantity I takes its maximum value—and indeed, does so even if we interchange M and S. But surely we do not have an explanation in either case.

§2.4 *The Difficulties: Asymmetries and Rejections*

There are two main difficulties, illustrated by the old paresis and barometer examples, which none of the examined positions can handle. The first is that there are cases, clearly in a theory's domain, where the request for explanation is nevertheless rejected. We can explain why John, rather than his brothers, contracted paresis, for he had syphilis; but not why he, among all those syphilitics, got paresis. Medical science is incomplete, and hopes to find the answer some day. But the example of the uranium atom disintegrating just then rather than later, is formally similar and we believe the theory to be complete. We also reject such questions as the Aristotelians asked the Galileans: why does a body free of impressed forces retain its velocity? The importance of this sort of case, and its pervasive character, has been repeatedly discussed by Adolf Grünbaum. It was also noted, in a different context, by Thomas Kuhn.[18] Examples he gives of explanation requests which were considered legitimate in some periods and rejected in others cover a wide range of topics. They include the qualities of compounds in chemical theory (explained before Lavoisier's reform, and not considered something

to be explained in the nineteenth century, but now again the subject of chemical explanation). Clerk Maxwell accepted as legitimate the request to explain electromagnetic phenomena within mechanics. As his theory became more successful and more widely accepted, scientists ceased to see the lack of this as a shortcoming. The same had happened with Newton's theory of gravitation which did not (in the opinion of Newton or his contemporaries) contain an explanation of gravitational phenomena, but only a description. In both cases there came a stage at which such problems were classed as intrinsically illegitimate, and regarded exactly as the request for an explanation of why a body retains its velocity in the absence of impressed forces. While all of this may be interpreted in various ways (such as through Kuhn's theory of paradigms) the important fact for the theory of explanation is that not everything in a theory's domain is a legitimate topic for why-questions; and that what is, is not determinable *a priori*.

The second difficulty is the asymmetry revealed by the barometer, the red shift, and the flagpole examples: even if the theory implies that one condition obtains when and only when another does, it may be that it explains the one in terms of the other and not vice versa. An example which combines both the first and second difficulties is this: according to atomic physics, each chemical element has a characteristic atomic structure and a characteristic spectrum (of light emitted upon excitation). Yet the spectrum is explained by the atomic structure, and the question why a substance has that structure does not arise at all (except in the trivial sense that the questioner may need to have the terms explained to him).

To be successful, a theory of explanation must accommodate, and account for, both rejections and asymmetries. I shall now examine some attempts to come to terms with these, and gather from them the clues to the correct account.

§2.5 *Causality: the* Conditio Sine Qua Non

Why are there no longer any Tasmanian natives? Why are the Plains Indians now living on reservations? Of course it is possible to cite relevant statistics: in many areas of the world, during many periods of history, upon the invasion by a technologically advanced people, the natives were displaced and weakened culturally, physically, and economically. But such a response will not satisfy: what we want is the story behind the event.

In Tasmania, attempts to round up and contain the natives were unsuccessful, so the white settlers simply started shooting them, man, woman, and child, until eventually there were none left. On the American Plains, the whites systematically destroyed the great buffalo herds on which the Indians relied for food and clothing, thus dooming them to starvation or surrender. There you see the story, it moves by its own internal necessity, and it explains why.

I use the word 'necessity' advisedly, for that is the term that links stories and causation. According to Aristotle's *Poetics*, the right way to write a story is to construct a situation which, after the initial parameters are fixed, moves toward its conclusion with a sort of necessity, inexorably—in retrospect, 'it had to end that way'. This was to begin also the hallmark of a causal explanation. Both in literature and in science we now accept such accounts as showing only how the events could have come about in the way they did. But it may be held that, to be an explanation, a scientific account must still tell a story of how things did happen and how the events hang together, so to say.

The idea of causality in modern philosophy is that of a relation among events. Hence it cannot be identified even with efficient causation, its nearest Aristotelian relative. In the modern sense we cannot say, correctly and non-elliptically, that the salt, or the moisture in the air, caused the rusting of the knife. Instead, we must say that certain events caused the rusting: such events as dropping of the salt on the knife, the air moistening that salt, and so on. The exact phrasing is not important; that the *relata* are events (including processes and momentary or prolonged states of affairs) is very important.

But what exactly is that causal relation? Everyone will recognize Hume's question here, and recall his rejection of certain metaphysical accounts. But we do after all talk this way, we say that the knife rusted because I dropped salt on it—and, as philosophers, we must make sense of explanation. In this and the next subsection I shall discuss some attempts to explicate the modern causal relation.

When something is cited as a cause, it is not implied that it was sufficient to produce (guarantee the occurrence) of the event. I say that this plant died because it was sprayed with defoliant, while knowing that the defoliant is only 90 per cent effective. Hence, the tradition that identifies the cause as the *conditio sine qua non*: had the plant not been sprayed, it would not have died.[19]

There are two problems with restating this as: a cause is a necessary

condition. In the first place, not every necessary condition is a cause; and secondly, in some straightforward sense, a cause may not be necessary, namely, alternative causes could have led to the same result. An example for the first problem is this: the existence of the knife is a necessary condition for its rusting, and the growth of the plant for its dying. But neither of these could be cited as a cause. As to the second, it is clear that the plant could have died some other way, say if I had carefully covered it totally with anti-rust paint.

J. L. Mackie proposed the definition: a cause is an insufficient but necessary part of an unnecessary but sufficient condition.[20] That sufficient condition must precede the event to be explained, of course; it must not be something like the (growth-plus death-plus rotting) of the plant if we wish to cite a cause for its death. But the first problem still stands anyway, since the existence of knife is a necessary part of the total set of conditions that led to its rusting. More worrisome is the fact that there may be no sufficient preceding conditions at all: the presence of the radium is what caused the Geiger counter to click, but atomic physics allows a non-zero probability for the counter not clicking at all under the circumstances.

For this reason (the non-availability of sufficient conditions in certain cases), Mackie's definition does not defuse the second problem either.

David Lewis has given an account in terms of counterfactual conditionals.[21] He simply equates '*A* caused *B*' with 'if *A* had not happened, *B* would not have happened'. But it is important to understand this conditional sentence correctly, and not to think of it (as earlier logicians did) as stating that *A* was a necessary condition for the occurrence of *B*. Indeed, the 'if . . . then' is not correctly identified with any of the sorts of implication traditionally discussed in logical theory, for those obey the law of *Weakening*:

1. If *A* then *B*
   hence
   if *A* and *C* then *B*.

But our conditionals, in natural language, typically do not obey that law:

2. If the match is struck it will light
   hence (?)

if the match is dunked in coffee and
struck, it will light;

the reader will think of many other examples. The explanation of why that 'law' does not hold is that our conditionals carry a tacit *ceteris paribus* clause:

3. If the plant had not been sprayed
(*and all else had been the same*)
then it would not have died.

The logical effect of this tacit clause is to make the 'law' of Weakening inapplicable.

Of course, it is impossible to spell out the exact content of *ceteris paribus*, as Goodman found in his classic discussion, for that content changes from context to context.[22] To this point I shall have to return. Under the circumstances, it is at least logically tenable to say, as David Lewis does, that whenever '*A* is the (a) cause of (or: caused) *B*' is true, it is also true that if *A* had not happened, neither would *B* have.

But do we have a sufficient criterion here? Suppose David's alarm clock goes off at seven a.m. and he wakes up. Now, we cite the alarm as the cause of the awakening, and may grant, if only for the sake of argument, that if the alarm had not sounded, he would not (then) have woken up. But it is also true that if he had not gone to sleep the night before, he would not have woken in the morning. This does not seem sufficient reason to say that he woke up because he had gone to sleep.

The response to this and similar examples is that the counterfactuals single out all the nodes in the causal net on lines leading to the event (the awakening), whereas 'because' points to specific factors that, for one reason or other, seem especially relevant (*salient*) in the context of our discussion. No one will deny that his going to sleep was one of the events that 'led up' to his awakening, that is, in the relevant part of the causal net. That part of the causal story is objective, and which specific item is singled out for special attention depends on the context—*every* theory of causation must say this.

Fair enough. That much context-dependence everyone will have to allow. But I think that much more context-dependence enters this theory through the truth-conditions of the counterfactuals themselves. So much, in fact, that we must conclude that there is nothing

in science itself—nothing in the objective description of nature that science purports to give us—that corresponds to these counterfactual conditionals.

Consider again statement (3) about the plant sprayed with defoliant. It is true in a given situation exactly if the 'all else' that is kept 'fixed' is such as to rule out death of the plant for other reasons. But who keeps what fixed? The speaker, in his mind. There is therefore a contextual variable—determining the content of that tacit *ceteris paribus* clause—which is crucial to the truth-value of the conditional statement. Let us suppose that I say to myself, *sotto voce*, that a certain fuse leads into a barrel of gunpowder, and then say out loud, 'If Tom lit that fuse there would be an explosion.' Suppose that before I came in, you had observed to yourself that Tom is very cautious, and would not light any fuse before disconnecting it, and said out loud, 'If Tom lit that fuse, there would be no explosion.' Have we contradicted each other? Is there an objective right or wrong about keeping one thing rather than another firmly in mind when uttering the antecedent 'If Tom lit that fuse . . .'? It seems rather that the proposition expressed by the sentence depends on a context, in which 'everything else being equal' takes on a definite content.

Robert Stalnaker and David Lewis give truth-conditions for conditionals using the notion of similarity among possible worlds. Thus, on one such account, 'if A then B' is true in world w exactly if B is true in the most similar world to w in which A is true. But there are many similarity relations among any set of things. Examples of the sort I have just given have long since elicited the agreement that the relevant similarity relation changes from context to context. Indeed, without that agreement, the logics of conditionals in the literature are violated by these examples.

One such example is very old: Lewis Carroll's puzzle of the three barbers. It occurs in *The Philosophy of Mr. B*rtr*nd R*ss*ll* as follows:

> Allen, Brown, and Carr keep a barber's shop together; so that one of them must be in during working hours. Allen has lately had an illness of such a nature that, if Allen is out, Brown must be accompanying him. Further, if Carr is out, then, if Allen is out, Brown must be in for obvious business reasons.[23]

The above story gives rise to two conditionals, if we first suppose that Carr is out:

1. If Allen is out then Brown is out
2. If Allen is out then Brown is in

the first warranted by the remarks about Allen's illness, the second by the obvious business reasons. Lewis Carroll, thinking that 1 and 2 contradict each other, took this as a *reductio ad absurdum* of the supposition that Carr is out. R*ss*ll, construing 'if *A* then *B*' as the material conditional ('either *B* or not *A*') asserts that 1 and 2 are both true if Allen is not out, and so says that we have here only a proof that if Carr is out, then Allen is in. ('The odd part of this conclusion is that it is the one which common-sense would have drawn', he adds.)

We have many other reasons, however, for not believing the conditional of natural language to be the material conditional. In modal logic, the strict conditional is such that 1 and 2 imply that it is not possible that Allen is out. So the argument would demonstrate 'If Carr is out then it is not possible that Allen is out.' This is false: if it looks true, it does so because it is easily confused with 'It is not possible that Carr is out and Allen is out.' If we know that Carr is out we can conclude that it is false that Allen is out, not that it is impossible.

The standard logics of counterfactual conditionals give exactly the same conclusion as the modal logic of strict conditionals. However, by noting the context-dependence of these statements, we can solve the problem correctly. Statement 1 is true in a context in which we disregard business requirements and keep fixed the fact of Allen's illness; statement 2 is true if we reverse what is fixed and what is variable. Now, there can exist contexts c and c' in which 1 and 2 are true respectively, only if their common antecedent is false; thus, like R*ss*ll, we are led to the conclusion drawn by common sense.

Any of the examples, and any general form of semantics for conditionals, will lend themselves to make the same point. What sort of situation, among all the possible unrealized ones, is more like ours in the fuse example: one in which nothing new is done except that the fuse is lit, or one in which the fuse is lit after being disconnected? It all depends—similar in what respect? Similar in that no fuse is disconnected or similar in that no one is being irresponsible? Quine brought out this feature of counterfactuals—to serve another purpose—when he asked whether, if Verdi and Bizet had been compatriots, would they have been French or Italian? Finally, even if

someone feels very clear on what facts should be kept fixed in the evaluation of a counterfactual conditional, he will soon realize that it is not merely the facts but the description of the facts—or, if you like, facts identified by non-extensional criteria—that matter: Danny is a man, Danny is very much interested in women, i.e. (?) in the opposite sex—if he had been a woman would he have been very much interested in men, or a Lesbian?

These puzzles cause us no difficulty, if we say that the content of 'all else being equal' is fixed not only by the sentence and the factual situation, but also by contextual factors. In that case, however, the hope that the study of counterfactuals might elucidate science is quite mistaken: scientific propositions are not context-dependent in any essential way, so if counterfactual conditionals are, then science neither contains nor implies counterfactuals.

The truth-value of a conditional depends in part on the context. Science does not imply that the context is one way or another. Therefore science does not imply the truth of any counterfactual—except in the limiting case of a conditional with the same truth-value in all contexts. (Such limiting cases are ones in which the scientific theory plus the antecedent strictly imply the consequent, and for them logical laws like Weakening and Contraposition are valid, so that they are useless for the application to explanation which we are at present exploring.)

There was at one point a hope, expressed by Goodman, Reichenbach, Hempel, and others, that counterfactual conditionals provide an objective criterion for what is a law of nature, or at least, a lawlike statement (where a law is a true lawlike statement). A merely general truth was to be distinguished from a law because the latter, and not the former, implies counterfactuals. This idea must be inverted: if laws imply counterfactuals then, because counterfactuals are context-dependent, the concept of law does not point to any objective distinction in nature.

If, as I am inclined to agree, counterfactual language is proper to explanation, we should conclude that explanation harbours a significant degree of context-dependence.

§2.6 *Causality: Salmon's Theory*

The preceding subsection began by relating causation to stories, but the accounts of causality it examined concentrated on the links between particular events. The problems that appeared may there-

fore have resulted from the concentration on 'local properties' of the story. An account of causal explanation which focuses on extended processes has recently been given by Wesley Salmon.[24]

In his earlier theory, to the effect that an explanation consists in listing statistically relevant factors, Salmon had asked 'What more could one ask of an explanation?' He now answers this question:

> What does explanation offer, over and above the inferential capacity of prediction and retrodiction . . . ? It provides knowledge of the mechanisms of *production* and *propagation* of structure in the world. That goes some distance beyond mere recognition of regularities, and of the possibility of subsuming particular phenomena thereunder.[25]

The question, what is the causal relation? is now replaced by: what is a causal process? and, what is a causal interaction? In his answer to these questions, Salmon relies to a large extent on Reichenbach's theory of the common cause, which we encountered before. But Salmon modifies this theory considerably.

A process is a spatio-temporally continuous series of events. The continuity is important, and Salmon blames some of Hume's difficulties on his picture of processes as chains of events with discrete links.[26] Some processes are causal, or genuine processes, and some are pseudo-processes. For example, if a car moves along a road, its shadow moves along that road too. The series of events in which the car occupies successive points on that road is a genuine causal process. But the movement of the shadow is merely a pseudo-process, because, intuitively speaking, the position of the shadow at later times is not caused by its position at earlier times. Rather, there is shadow *here* now because there is a car here now, and not because there was shadow *there* then.

Reichenbach tried to give a criterion for this distinction by means of probabilistic relations.[27] The series of events A_t is a causal process provided

(1) the probability of $A_{t+\tau}$ given A_t is greater than or equal to the probability of $A_{t+\tau}$ given $A_{t-t'}$, which is in turn greater than the probability of $A_{t-\tau}$, *simpliciter*.

This condition does not yet rule out pseudo-processes, so we add that each event in the series *screens off* the earlier ones from the later ones:

(2) the probability of $A_{t+\tau}$, given both A_t and $A_{t-t'}$ is just that of $A_{t+\tau}$ given A_t

and, *in addition*, there is no other series of events B_r which screens off A_{r+s} from A_r for all r. The idea in the example is that if A_{r+s} is the position of the shadow at time $r+s$, then B_r is the position of the car at time $r+s$.

This is not satisfactory for two reasons. The first is that (1) reminds one of a well-known property of stochastic processes, called the Markov property, and seems to be too strong to go into the definition of causal processes. Why should not the whole history of the process up to time r give more information about what happens later than the state at time r does by itself? The second problem is that in the addition to (2) we should surely add that B_r must itself be a genuine causal process? For otherwise the movement of the car is not a causal process either, since the movement of the shadow will screen off successive positions of the car from each other. But if we say that B_r must be a genuine process in this stipulation, we have landed in a regress.

Reichenbach suggested a second criterion, called the *mark method* and (presumably because it stops the threatened regress) Salmon prefers that.

> If a fender is scraped as a result of a collision with a stone wall, the mark of that collision will be carried on by the car long after the interaction with the wall occurred. The shadow of a car moving along the shoulder is a pseudo-process. If it is deformed as it encounters a stone wall, it will immediately resume its former shape as soon as it passes by the wall. It will not transmit a mark or modification.[28]

So if the process is genuine then interference with an earlier event will have effects on later events in that process. However, thus phrased, this statement is blatantly a causal claim. How shall we explicate 'interference' and 'effects'? Salmon will shortly give an account of causal interactions (see below) but begins by appealing to his 'at-at' theory of motion. The movement of the car consists simply in being *at* all these positions *at* those various times. Similarly, the propagation of the mark consists simply in the mark being there, in those later events. There is not, over and above this, a special propagation relation.

However, there is more serious cause for worry. We cannot define a genuine process as one that *does* propagate a mark in this sense. There are features which the shadow carries along in that 'at-at' sense, such as that its shape is related, at all times, in a certain topologically definable way to the shape of the car, and that it is black.

Other special marks are not always carried—imagine part of a rocket's journey during which it encounters nothing else. So what we need to say is that the process is genuine if, *were* there to be a given sort of interaction at an early stage, there *would be* certain marks in the later stages. At this point, I must refer back to the preceding section for a discussion of such counterfactual assertions.

We can, at this point, relativize the notions used to the theory accepted. About some processes, our theory *implies* that certain interactions at an early stage will be followed by certain marks at later stages. Hence we can say that, *relative to the theory* certain processes are classifiable as genuine and others as pseudo-processes. What this does not warrant is regarding the distinction as an objective one. However, if the distinction is introduced to play a role in the theory of explanation, and if explanation is a relation of theory to fact, this conclusion does not seem to me a variation on Salmon's theory that would defeat its purpose.[29]

Turning now to causal interactions, Salmon describes two sorts. These interactions are the 'nodes' in the causal net, the 'knots' that combine all those causal processes into a causal structure. Instead of 'node' or 'knot' Reichenbach and Salmon also use 'fork' (as in 'the road forks'). Reichenbach described one sort, the *conjunctive fork* which occurs when an event C, belonging to two processes, is the *common cause* of events A and B, in those separate processes, occurring after C. Here common cause is meant in Reichenbach's original sense:

(3) $P(A\ \&\ B/C) = P(A/C).P(B/C)$
(4) $P(A\ \&\ B/\bar{C}) = P(A/\bar{C}).P(B/\bar{C})$
(5) $P(A/C) > P(A/\bar{C})$
(6) $P(B/C) > P(B/\bar{C})$

which, as noted in Chapter 2, entails that there is a positive correlation between A and B.

In order to accommodate the recalcitrant examples (see Chapter 2) Salmon introduced in addition the *interactive fork*, which is like the preceding one except that (3) is changed to

(3*) $P(A\ \&\ B/C) > P(A/C).P(B/C)$

These forks then combine the genuine causal processes, once identified, into the causal net that constitutes the natural order.

Explanation, on Salmon's new account, consists therefore in exhibiting the relevant part of the causal net that leads up to the events that are to be explained. In some cases we need only point to a single causal process that leads up to the event in question. In other cases we are asked to explain the confluence of events, or a positive correlation, and we do so by tracing them back to forks, that is, common origins of the processes that led up to them.

Various standard problems are handled. The sequence, barometer falling–storm coming, is not a causal process since the relevance of the first to the second is screened off by the common cause of atmospheric conditions. When paresis is explained by mentioning latent untreated syphilis, one is clearly pointing to the causal process, whatever it is, that leads from one to the other—or to their common cause, whatever that is. It must of course be a crucial feature of this theory that ordinary explanations are 'pointers to' causal processes and interactions which would, if known or described in detail, give the full explanation.

If that is correct, then each explanation must have, as cash-value, some tracing back (which is possible in principle) of separate causal processes to the forks that connect them. There are various difficulties with this view. The first is that to be a causal process, the sequence of events must correspond to a continuous spatio-temporal trajectory. In quantum mechanics, this requirement is not met. It was exactly the crucial innovation in the transition from the Bohr atom of 1913 to the new quantum theory of 1924, that the exactly defined orbits of the electrons were discarded. Salmon mentions explicitly the limitation of this account to macroscopic phenomena (though he does discuss Compton scattering). This limitation is serious, for we have no independent reason to think that explanation in quantum mechanics is essentially different from elsewhere.

Secondly, many scientific explanations certainly do not look as if they are causal explanations in Salmon's sense. A causal law is presumably one that governs the temporal development of some process or interaction. There are also 'laws of coexistence', which give limits to possible states or simultaneous configurations. A simple example is Boyle's law for gases (temperature is proportional to volume times pressure, at any given time); another, Newton's law of gravitation; another, Pauli's exclusion principle. In some of these cases we can say that they (or their improved counterparts) were later deduced from theories that replaced 'action at a distance'

(which is not action at all, but a constraint on simultaneous states) with 'action by contact'. But suppose they were not so replaceable—would that mean that they could not be used in genuine explanations?

Salmon himself gives an example of explanation 'by common cause' which actually does not seem to fit his account. By observations on Brownian motion, scientists determined Avogadro's number, that is, the number of molecules in one mole of gas. By quite different observations, on the process of electrolysis, they determined the number of electron charges equal to one Faraday, that is, to the amount of electric charge needed to deposit one mole of a monovalent metal. These two numbers are equal. On the face of it, this equality is astonishing; but physics can explain this equality by deducing it from the basic theories governing both sorts of phenomena. The common cause Salmon identifies here is the basic mechanism—atomic and molecular structure—postulated to account for these phenomena. But surely it is clear that, however much the adduced explanation may deserve the name 'common cause', it does not point to a relationship between events (in Brownian motion on specific occasions and in electrolysis on specific occasions) which is traced back via causal processes to forks connecting these processes. The explanation is rather that the number found in experiment A at time t is the same as that found in totally independent experiment B at *any* other time t', because of the *similarity* in the physically independent causal processes observed on those two different occasions.

Many highly theoretical explanations at least look as if they escape Salmon's account. Examples here are explanations based on principles of least action, based on symmetry considerations, or, in relativistic theories, on information that relates to space–time as a whole, such as specification of the metric or gravitational field.

The conclusion suggested by all this is that the type of explanation characterized by Salmon, though apparently of central importance, is still at most a subspecies of explanations in general.

§2.7 *The Clues of Causality*

Let us agree that science gives us a picture of the world as a net of interconnected events, related to each other in a complex but orderly way. The difficulties we found in the preceding two sections throw some doubt on the adequacy of the terminology of cause and

causality to describe that picture; but let us not press this doubt further. The account of explanation suggested by the theories examined can now be restated in general terms as follows:

(1) Events are enmeshed in a net of causal relations
(2) What science describes is that causal net
(3) Explanation of why an event happens consists (typically) in an exhibition of salient factors in the part of the causal net formed by lines 'leading up to' that event
(4) Those salient factors mentioned in an explanation constitute (what are ordinarily called) the *cause*(*s*) of that event.

There are two clear reasons why, when the topic of explanation comes up, attention is switched from the causal net as a whole (or even the part that converges on the event in question) to 'salient factors'. The first reason is that any account of explanation must make sense of common examples of explanation—especially cases typically cited as scientific explanations. In such actual cases, the reasons cited are particular prior events or initial conditions or combinations thereof. The second reason is that no account of explanation should imply that we can never give an explanation—and to describe the whole causal net in any connected region, however small, is in almost every case impossible. So the least concession one would have to make is to say that the explanation need say no more than that *there is* a structure of causal relations of *a certain sort*, which could *in principle* be described in detail: the salient features are what picks out the 'certain sort'.

Interest in causation as such focuses attention on (1) and (2), but interest in explanation requires us to concentrate on (3) and (4). Indeed, from the latter point of view, it is sufficient to guarantee the truth of (1) and (2) by *defining*

the causal net = whatever structure of relations science describes

and leaving to those interested in causation as such the problem of describing that structure in abstract but illuminating ways, if they wish.

Could it be that the explanation of a fact or event nevertheless resides solely in that causal net, and that *any* way of drawing attention to it explains? The answer is *no*; in the case of causal explanation, the *explanation* consists in drawing attention to certain

('special', 'important') features of the causal net. Suppose for example that I wish to explain the extinction of the Irish elk. There is a very large class of factors that preceded this extinction and was statistically relevant to it—even very small increases in speed, contact area of the hoof, height, distribution of weight in the body, distribution of food supply, migration habits, surrounding fauna and flora—we know from selection theory that under proper conditions any variation in these can be decisive in the survival of the species. But although, if some of these had been different, the Irish elk would have survived, they are not said to provide the explanation of why it is now extinct. The explanation given is that the process of sexual selection favoured males with large antlers, and that these antlers were, in the environment where they lived, encumbering and the very opposite of survival-adaptive. The other factors I mentioned are not spurious causes, or screened off by the development of these huge and cumbersome antlers, because the extinction was the total effect of many contributing factors; but those other factors are not the salient factors.

We turn then to those salient features that are cited in explanation—those referred to as 'the cause(s)' or 'the real cause(s)'. Various philosophical writers, seeking for an objective account of explanation, have attempted to state criteria that single out those special factors. I shall not discuss their attempts. Let me just cite a small survey of their answers: Lewis White Beck says that the cause is that factor over which we have most control; Nagel argues that it is often exactly that factor which is not under our control; Braithwaite takes the salient factors to be the unknown ones; and David Bohm takes them to be the factors which are the most variable.[30]

Why should different writers have given such different answers? The reason was exhibited, I think, by Norwood Russell Hanson, in his discussion of causation.

> There are as many causes of x as there are explanations of x. Consider how the cause of death might have been set out by a physician as 'multiple haemorrhage', by the barrister as 'negligence on the part of the driver', by a carriage-builder as 'a defect in the brakeblock construction', by a civic planner as 'the presence of tall shrubbery at that turning'.[31]

In other words, the salient feature picked out as 'the cause' in that complex process, is salient to a given person because of his orientation, his interests, and various other peculiarities in the way he approaches or comes to know the problem—contextual factors.

It is important to notice that in a certain sense these different answers cannot be combined. The civic planner 'keeps fixed' the mechanical constitution of the car, and gives his answer in the conviction that regardless of the mechanical defects, which made a fast stop impossible, the accident need not have happened. The mechanic 'keeps fixed' the physical environment; despite the shrubbery obscuring vision, the accident need not have happened if the brakes had been better. What the one varies, the other keeps fixed, and you cannot do both at once. In other words, the selection of the salient causal factor is not simply a matter of pointing to the most interesting one, not like the selection of a tourist attraction; it is a matter of *competing* counterfactuals.

We must accordingly agree with the Dutch philosopher P. J. Zwart who concludes, after examining the above philosophical theories,

> It is therefore not the case that the meaning of the sentence 'A is the cause of B' depends on the nature of the phenomena A and B, but that this meaning depends on the context in which this sentence is uttered. The nature of A and B will in most cases also play a role, indirectly, but it is in the first place the orientation or the chosen point of view of the speaker that determines what the word cause is used to signify.[32]

In conclusion, then, this look at accounts of causation seems to establish that explanatory factors are to be chosen from a range of factors which are (or which the scientific theory lists as) objectively relevant in certain special ways—but that the choice is then determined by other factors that vary with the context of the explanation request. To sum up: no factor is explanatorily relevant unless it is scientifically relevant; and among the scientifically relevant factors, context determines explanatorily relevant ones.

§2.8 *Why-questions*

Another approach to explanation was initiated by Sylvain Bromberger in his study of why-questions.[33] After all, a why-question is a request for explanation. Consider the question:

1. Why did the conductor become warped during the short circuit?

This has the general form

2. Why (is it the case that) *P*?

where *P* is a statement. So we can think of 'Why' as a function that turns statements into questions.

Question 1 *arises*, or *is in order*, only if the conductor did indeed become warped then. If that is not so, we do not try to answer the question, but say something like: 'You are under a false impression, the conductor became warped much earlier,' or whatever. Hence Bromberger calls the statement that *P* the *presupposition* of the question *Why P?* One form of the rejection of explanation requests is clearly the denial of the presupposition of the corresponding why-question.

I will not discuss Bromberger's theory further here, but turn instead to a criticism of it. The following point about why-questions has been made in recent literature by Alan Garfinkel and Jon Dorling, but I think it was first made, and discussed in detail, in unpublished work by Bengt Hannson circulated in 1974.[34] Consider the question

3. Why did Adam eat the apple?

This same question can be construed in various ways, as is shown by the variants:

3a. Why was it Adam who ate the apple?
3b. Why was it the apple Adam ate?
3c. Why did Adam *eat* the apple?

In each case, the canonical form prescribed by Bromberger (as in 2 above) would be the same, namely

4. Why (is it the case that) (Adam ate the apple)?

yet there are three different explanation requests here.

The difference between these various requests is that they point to different contrasting alternatives. For example, 3b may ask why Adam ate *the apple* rather than some other fruit in the garden, while 3c asks perhaps why Adam *ate* the apple rather than give it back to Eve untouched. So to 3b, 'because he was hungry' is not a good answer, whereas to 3c it is. The correct general, underlying structure of a why-question is therefore

5 Why (is it the case that) *P in contrast to* (other members of) *X*?

where *X*, the *contrast-class*, is a set of alternatives. *P* may belong to *X* or not; further examples are:

Why did the sample burn green (rather than some other colour)?
Why did the water and copper reach equilibrium temperature 22.5 C (rather than some other temperature)?

In these cases the contrast-classes (colours, temperatures) are 'obvious'. In general, the contrast-class is not explicitly described because, *in context*, it is clear to all discussants what the intended alternatives are.

This observation explains the tension we feel in the paresis example. If a mother asks why her eldest son, a pillar of the community, mayor of his town, and best beloved of all her sons, has this dread disease, we answer: because he had latent untreated syphilis. But if that question is asked about this same person, immediately after a discussion of the fact that everyone in his country club has a history of untreated syphilis, *there is no answer*. The reason for the difference is that in the first case the contrast-class is the mother's sons, and in the second, the members of the country club, contracting paresis. Clearly, an answer to a question of form 5 must adduce information that *favours P in contrast to* other members of *X*. Sometimes the availability of such information depends strongly on the choice of *X*.

These reflections have great intuitive force. The distinction made is clearly crucial to the paresis example and explains the sense of ambiguity and tension felt in earlier discussion of such examples. It also gives us the right way to explicate such assertions as: individual events are never explained, we only explain a particular event *qua* event of a certain kind. (We can explain *this* decay of a uranium atom *qua* decay of a uranium atom, but not *qua* decay of a uranium atom at *this* time.)

But the explication of what it is for an answer to favour one alternative over another proves difficult. Hannson proposed: answer *A* is a good answer to (Why *P* in contrast to *X*?) exactly if the probability of *P* given *A* is higher than the average probability of members of *X* given *A*. But this proposal runs into most of the old difficulties. Recall Salmon's examples of irrelevancy: the probability of recovery from a cold *given* administration of vitamin C is nearly one, while the probability of not recovering *given* the vitamins is nearly zero. So by Hannson's criterion it would be a good answer—even if taking vitamin C has no effect on recovery from colds one way or the other.

Also, the asymmetries are as worrisome as ever. By Hannson's criterion, the length of the shadow automatically provides a good explanation of the height of the flagpole. And 'because the barometer fell' is a good answer to 'why is there a storm?' (upon selection of

the 'obvious' contrast-classes, of course). Thus it seems that reflection on the contrast-class serves to solve some of our problems, but not all.

§2.9 *The Clues Elaborated*

The discussions of causality and of why-questions seem to me to provide essential clues to the correct account of explanation. In the former we found that an explanation often consists in listing salient factors, which point to a complete story of how the event happened. The effect of this is to eliminate various alternative hypotheses about how this event did come about, and/or eliminate puzzlement concerning how the event could have come about. But salience is context-dependent, and the selection of the correct 'important' factor depends on the range of alternatives contemplated in that context. In N. R. Hanson's example, the barrister wants this sort of weeding out of hypotheses about the death relevant to the question of legal accountability; the carriage-builder, a weeding out of hypotheses about structural defects or structural limitations under various sorts of strain. *The context*, in other words, *determines relevance* in a way that goes well beyond the statistical relevance about which our scientific theories give information.

This might not be important if we were not concerned to find out exactly how having an explanation goes beyond merely having an acceptable theory about the domain of phenomena in question. But that is exactly the topic of our concern.

In the discussion of why-questions, we have discovered a further contextually determined factor. The range of hypotheses about the event which the explanation must 'weed out' or 'cut down' is not determined solely by the interests of the discussants (legal, mechanical, medical) but also by a range of contrasting alternatives to the event. This *contrast-class* is also determined by context.

It might be thought that when we request a *scientific* explanation, the relevance of possible hypotheses, and also the contrast-class are automatically determined. But this is not so, for both the physician and the motor mechanic are asked for a scientific explanation. The physician explains the fatality *qua* death of a human organism, and the mechanic explains it *qua* automobile crash fatality. To ask that their explanations be scientific is only to demand that they rely on scientific theories and experimentation, not on old wives' tales. Since any explanation of an individual event must be an explanation of

that event *qua* instance of a certain kind of event, nothing more can be asked.

The two clues must be put together. The description of some account as an explanation of a given fact or event, is incomplete. It can only be an explanation with respect to a certain *relevance relation* and a certain *contrast-class*. These are contextual factors, in that they are determined neither by the totality of accepted scientific theories, nor by the event or fact for which an explanation is requested. It is sometimes said that an Omniscient Being would have a complete explanation, whereas these contextual factors only bespeak our limitations due to which we can only grasp one part or aspect of the complete explanation at any given time. But this is a mistake. If the Omniscient Being has no specific interests (legal, medical, economic; or just an interest in optics or thermodynamics rather than chemistry) and does not abstract (so that he never thinks of Caesar's death *qua* multiple stabbing, or *qua* assassination), then no why-questions ever arise for him in any way at all –and he does not have any explanation in the sense that we have explanations. If he does have interests, and does abstract from individual peculiarities in his thinking about the world, then his why-questions are as essentially context-dependent as ours. In either case, his advantage is that he always has all the information needed to answer any specific explanation request. But that information is, in and by itself, not an explanation: just as a person cannot be said to be older, or a neighbour, except in relation to others.

§3. *Asymmetries of Explanation. A Short Story*

§3.1 *Asymmetry and Context: the Aristotelian Sieve*

That vexing problem about paresis, where we seem both to have and not to have an explanation, was solved by reflection on the contextually supplied contrast-class. The equally vexing, and much older, problem of the asymmetries of explanation, is illuminated by reflection on the other main contextual factor: contextual relevance.

If that is correct, if the asymmetries of explanation result from a contextually determined relation of relevance, then it must be the case that these asymmetries can at least sometimes be reversed by a change in context. In addition, it should then also be possible to account for specific asymmetries in terms of the interests of questioner and audience that determine this relevance. These considera-

tions provide a crucial test for the account of explanation which I propose.

Fortunately, there is a precedent for this sort of account of the asymmetries, namely in Aristotle's theory of science. It is traditional to understand this part of his theory in relation to his metaphysics; but I maintain that the central aspects of his solution to the problem of asymmetry of explanations are independently usable.[35]

Aristotle gave examples of this problem in the *Posterior Analytics* I, 13; and he developed a typology of explanatory factors ('the four causes'). The solution is then simply this. Suppose there are a definite (e.g. four) number of types of explanatory factors (i.e. of relevance relations for why-questions). Suppose also that relative to our background information and accepted theories, the propositions *A* and *B* are equivalent. It may then still be that these two propositions describe factors of different types. Suppose that in a certain context, our interest is in the mode of production of an event, and 'Because *B*' is an acceptable answer to 'Why *A*?'. Then it may well be that *A* does not describe any mode of production of anything, so that, *in this same context*, 'Because *A*' would not be an acceptable answer to 'Why *B*?'.

Aristotle's lantern example (*Posterior Analytics* II, 11) shows that he recognized that in different contexts, verbally the same why-question may be a request for different types of explanatory factors. In modern dress the example would run as follows. Suppose a father asks his teenage son, 'Why is the porch light on?' and the son replies 'The porch switch is closed and the electricity is reaching the bulb through that switch.' At this point you are most likely to feel that the son is being impudent. This is because you are most likely to think that the sort of answer the father needed was something like: 'Because we are expecting company.' But it is easy to imagine a less likely question context: the father and son are re-wiring the house and the father, unexpectedly seeing the porch light on, fears that he has caused a short circuit that bypasses the porch light switch. In the second case, he is *not* interested in the human expectations or desires that led to the depressing of the switch.

Aristotle's fourfold typology of causes is probably an over-simplification of the variety of interests that can determine the selection of a range of relevant factors for a why-question. But in my opinion, appeal to some such typology will successfully illuminate the asymmetries (and also the rejections, since no factor of a *particular* type

may lead to a telling answer to the why-question). If that is so then, as I said before, asymmetries must be at least sometimes reversible through a change in context. The story which follows is meant to illustrate this. As in the lantern (or porch light) example, the relevance changes from one sort of efficient cause to another, the second being a person's desires. As in all explanations, the correct answer consists in the exhibition of a single factor in the causal net, which is made salient in that context by factors not overtly appearing in the words of the question.

§3.2 *'The Tower and the Shadow'*

During my travels along the Saône and Rhône last year, I spent a day and night at the ancestral home of the Chevalier de St. X ..., an old friend of my father's. The Chevalier had in fact been the French liaison officer attached to my father's brigade in the first war, which had—if their reminiscences are to be trusted—played a not insignificant part in the battles of the Somme and Marne.

The old gentleman always had *thé à l'Anglaise* on the terrace at five o'clock in the evening, he told me. It was at this meal that a strange incident occurred; though its ramifications were of course not yet perceptible when I heard the Chevalier give his simple explanation of the length of the shadow which encroached upon us there on the terrace. I had just eaten my fifth piece of bread and butter and had begun my third cup of tea when I chanced to look up. In the dying light of that late afternoon, his profile was sharply etched against the granite background of the wall behind him, the great aquiline nose thrust forward and his eyes fixed on some point behind my left shoulder. Not understanding the situation at first, I must admit that to begin with, I was merely fascinated by the sight of that great hooked nose, recalling my father's claim that this had once served as an effective weapon in close combat with a German grenadier. But I was roused from this brown study by the Chevalier's voice.

'The shadow of the tower will soon reach us, and the terrace will turn chilly. I suggest we finish our tea and go inside.'

I looked around, and the shadow of the rather curious tower I had earlier noticed in the grounds, had indeed approached to within a yard from my chair. The news rather displeased me, for it was a fine evening; I wished to remonstrate but did not well know how, without overstepping the bounds of hospitality. I exclaimed,

'Why must that tower have such a long shadow? This terrace is so pleasant!'

His eyes turned to rest on me. My question had been rhetorical, but he did not take it so.

'As you may already know, one of my ancestors mounted the scaffold with Louis XVI and Marie Antoinette. I had that tower erected in 1930 to mark the exact spot where it is said that he greeted the Queen when she first visited this house, and presented her with a peacock made of soap, then a rare substance. Since the Queen would have been one hundred and seventy-five years old in 1930, had she lived, I had the tower made exactly that many feet high.'

It took me a moment to see the relevance of all this. Never quick at sums, I was at first merely puzzled as to why the measurement should have been in feet; but of course I already knew him for an Anglophile. He added drily, 'The sun not being alterable in its course, light travelling in straight lines, and the laws of trigonometry being immutable, you will perceive that the length of the shadow is determined by the height of the tower.' We rose and went inside.

I was still reading at eleven that evening when there was a knock at my door. Opening it I found the housemaid, dressed in a somewhat old-fashioned black dress and white cap, whom I had perceived hovering in the background on several occasions that day. Courtseying prettily, she asked, 'Would the gentleman like to have his bed turned down for the night?'

I stepped aside, not wishing to refuse, but remarked that it was very late—was she kept on duty to such hours? No, indeed, she answered, as she deftly turned my bed covers, but it had occurred to her that some duties might be pleasures as well. In such and similar philosophical reflections we spent a few pleasant hours together, until eventually I mentioned casually how silly it seemed to me that the tower's shadow ruined the terrace for a prolonged, leisurely tea.

At this, her brow clouded. She sat up sharply. 'What exactly did he tell you about this?' I replied lightly, repeating the story about Marie Antoinette, which now sounded a bit far-fetched even to my credulous ears.

'The *servants* have a different account', she said with a sneer that was not at all becoming, it seemed to me, on such a young and pretty face. 'The truth is quite different, and has nothing to do with ancestors. That tower marks the spot where he killed the maid with whom he had been in love to the point of madness. And the height of the

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:

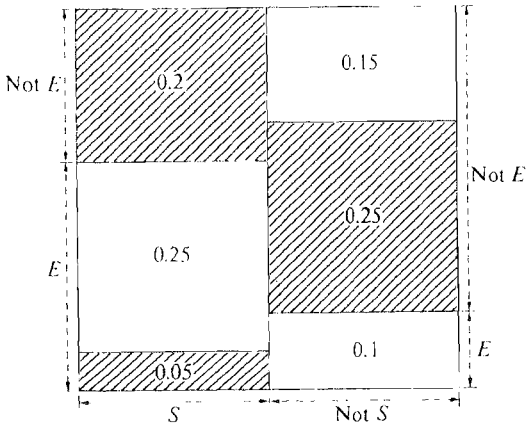
Why E_1 rather than $E_2, \ldots, E_{1000}$?
Because A.
$Prob\ (E_1) = \ldots = Prob\ (E_{10}) = 99/1000 = 0.099$
$Prob\ (E_{11}) = \ldots = Prob\ (E_{1000}) = 1/99{,}000 \doteq 0.00001$
$Prob\ (E_1/A) = 90/1000 = 0.090$
$Prob\ (E_2/A) = \ldots = Prob\ (E_{1000}/A) = 910/999{,}000 \doteq 0.001$

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 deter-

ministic 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.