A CAUSAL APPROACH TO EXPLANATIONS

Erik WEBER

1. *Introduction*

The analysis of explanations has become one of the most important topics in recent philosophy of science. This paper is meant as a contribution to this analysis, and will consist of three main parts. First of all, I will deal with some traditional problems (determinism, explanatory asymmetries) without referring to causal relations (section 3). This non-causal analysis is only a first step in the construction of a causal model. My first aim is to develop such causal model, and to show that a causal approach is to be preferred (section 4). As we will see, the criteria on explanations resulting from a causal approach have to be identical with the results of a non-causal approach. Therefore, the preliminary analysis of section 3 is indispensable. The second aim of this paper is to solve a problem which hitherto has been neglected in studies of explanations. It will appear that some explanations are extensional, while other explanations are not. Each explanation has to describe the process that caused its explanandum. An explanation is extensional if the way in which this causal mechanism is described is not important: if the description of the mechanism is replaced by another description which phrases the same causal relation in other terms, we still have an adequate explanation. If an explanation is not extensional, only some descriptions of the relevant causal mechanism allow us to give an adequate explanation. Additional requirements which are to be met by this second class of explanations, will be formulated in section 5.

2. *Espistemic versus causal models*

When we try to classify the different models of explanation which have been developed since Hempel, we discover that there are two research traditions. In the first tradition, an explanation is seen as a derivation. In the other, an explanation is considered an identification of causal mechanisms.

The models of the first tradition may be called epistemic models, since the explanation seeking question "Why E?" is regarded as a request for an argument which allows us to conclude that E is true. Explanation seeking questions only arise if one already has *empirical* grounds to believe that the explanandum is true, i.e. if *observation* has already shown that E is true. They are answered by giving a *rational* ground for E, i.e. by constructing an argument: we are looking for "Vernunftgründe". [1] The differences between the models of this tradition arise because different criteria are introduced to exclude derivations that are not adequate as explanations: high probability requirement (C.G. Hempel 1965), statistical relevance (W. Salmon 1971), positive statistical relevance (P. Gärdenfors 1980) etc. The models of the second tradition may be called ontic or causal models. An explanation seeking why-question is not a request for an argument to derive E. We want to understand how the fact described in the explanandum sentence came into existence. We want to know which processes led to the phenomenon that is to be explained: we are asking for causes, "Realgründe". Each ontic model is the result of a particular analysis of the concepts "cause", "effect" and "causal relation". Epistemic and ontic models have a common view on the context of explanations: an explanation is necessary when one is confronted with a phenomenon that looks strange, whose features are found surprising. The first and most fundamental difference between the two approaches lies in their general characterisation of what it means to understand a phenomenon. According to the epistemic approach, we have to give arguments, derivations; according to the ontic approach, we have to give causes, we must identify the relevant causal mechanism.

3. *An epistemic model of explanations*

3.1. *Covering law explanations*

A traditional way of describing the general structure of an explanation is as follows [2]:

[1] See e.g. W. STEGMÜLLER (1983) p. 910.

[2] This scheme corresponds to Hempel's deductive-nomological explanations (C.G. Hempel, 1965, pp. 335-411).

$$\frac{\begin{array}{c} C_1, C_2, \ldots, C_n \\ L_1, L_2, \ldots, L_n \end{array}}{E} \qquad \text{(I)}$$

The single line indicates that E is deductively derivable from a set of laws $L_1, L_2, \ldots, L_n$ and a set of singular statements $C_1, C_2, \ldots, C_n$. This scheme represents three criteria:

1° Each explanation is a derivation.

2° An explanation must contain at least one law and at least one singular statement.

3° The derivation has to be deductive (statistical explanations are excluded).

Can this deductive-nomological scheme be justified? That each explanation is a derivation has even to be accepted by the proponents of the second tradition, as each causal relation entails a conditional relation. If no derivation is possible, there can't be any causal relation. Of course, in practice not every explanation consists of both a set of singular statements and a set of law statements. For instance, if we want to explain a plane crash, it is possible that we are satisfied with the observation that there was ice-accretion on the wings. Indeed, if the relevant aerodynamical laws (Bernoulli principle) are already known, they don't have to be mentioned in the explanans. So the concrete form of an explanation depends on the background knowledge of the people involved. But since we want to construct a general model of explanations, this background knowledge has to be neglected. So this is not an argument against the first criterion. Does this mean that every explanation conforms to the covering law model, as is required in the second criterion? No. The function of the law is to show that there is a sufficient condition relationship (in the strong counterfactual sense: "A is a sufficient condition for B" means that "If A, then *always* B") between the set of singular statements and E. Since a law cannot connect two (or more) particular events, but only two (or more) types of events, only explanations of types of events conform to the covering law model. If we want to explain a particular event, all we find is a trivial sufficient condition relation ("If A, then B"). So explanations of particular events are derivations, but don't use any law statements. I will confine my attention to explanations of types of events. For these, at least one law is required to derive the explanandum from a set of singular statements, and hence the covering law model is adequate. If one wants to analyse explanations of particular events, scheme (I) has to be modified. We may conclude that the first two criteria don't cause any problems if the above-mentioned restrictions are respected. The third criterion will be examined in the subsequent section 3.2.

3.2. *Statistical explanations*

Consider the following example(3):

> A breeding experiment on pea plants produces a filial population in which 75 % of the plants have red blossoms, and 25 % white blossoms. A plant has red blossoms if it has genotype RR or Rr (so R is dominant with respect to r). A plant has white blossoms if it has genotype rr. The parent population consists of plants with genotype Rr. Each parent transmits an r with probability 0.5. No details about the fertilization process are known, so we can't predict when an r will be transmitted and when not. b_1 is a plant with red blossoms, b_2 is a plant with white blossoms. Can we say that we have explained why B_1 has red blossoms, though the probability of b_1 having red blossoms is 0.75, and not 1? Can we say that we have explained why b_2 has white blossoms, though the probability of b_2 having white blossoms is low (0.25)?

I think it is impossible to answer these questions, since they are not clear enough. As a matter of fact, there are two problems, viz. (i) Can this example be accepted as an ideally complete explanation?, and (ii) Can this example be accepted as a minimally sufficient explanation? Ideally complete explanations eliminate all our ununderstanding on a certain topic. It is obvious that the example is not an ideally complete explanation. The initial and most important questions, "Why does b_1 have red blossoms, and not white ones?" and "Why does b_2 have white blossoms, and not red ones?", are not answered. In general, statistical explanations don't answer questions of the type "Why X, and not Y?". But the explanation does answer some other questions, e.g. "Why is the probability of a pea plant having red blossoms 0.75?", "Why is the probability of a pea plant having white blossoms 0.25?", "How can a parent population consisting of pea plants with red blossoms produce some off-springs with white blossoms?", etc. So our understanding is advanced by explanations like

(3) P. KITCHER (1985) p. 634.

the one given in the example. Consequently, they are minimally sufficient. We can conclude that determinism is an ideal, but not a necessary condition; when only a statistical explanation can be given, there is a conflict between our ideals and what is known about the problem. Scheme (I) is a description of the logical structure of ideally complete explanations. It has to be modified in order to include all minimally sufficient explanations.

Are there any criteria to be imposed on statistical explanations in order to be minimally sufficient? It is obvious that no "high probability requirement" is needed: the explanation helps us to understand why b_2 has white blossoms just as much as it helps us to understand why b_1 has red ones. Hence, high a posteriori probability is neither a minimal requirement, nor an indicator of greater explanatory power. What about statistical relevance? The a posteriori probability of b_2 having white blossoms (0.25) is less than the a priori probability (0.5). So Gärdenfors' criterion of positive statistical relevance isn't appropriate either. Why don't we return to Salmon's original criterion (positive *or* negative statistical relevance)? Indeed, in order to answer questions like those we mentioned above, the reference class (in this case the class of all pea plants) must be divided into subclasses. That is exactly what we do when we discover statistically relevant factors. To my view, the best solution is to delineate the set of questions and explanations should answer to be minimally sufficient, and the set of questions which will be taken into account in determining the degree of explanatory power. Once such classification is made, the use of probabilities is superfluous. This solution is the most straightforward one: since the aim of the explanation is to answer those questions, and since there are no practical (i.e. metrical) problems, there is no reason why we should look for another criterion.

3.3. *Explanatory asymmetries*

Consider the following example, which we owe to W. Stegmüller (4):

(A) L: Everyone who is hit by a meteorite weighing 2 tons, dies almost immediately.

C: Mister X.Y. was hit by a meteorite weighing 2 tons yesterday.

E: Mister X.Y. died yesterday.

(4) W. STEGMÜLLER (1983) p. 911.

(B) L′: No one who is still alive has ever been hit by a meteorite weighing 2 tons.
C′: Mister X.Y. was still alive yesterday.

E′: Mister X.Y. wasn't hit by a meteorite weighing 2 tons yesterday.

Both (A) and (B) meet the requirements imposed on explanations till now, but only (A) is in agreement with our pre-analytic notion of explanations. (B) is intuitively unacceptable. So an additional criterion has to be found, to cope with this asymmetry. Many solutions have been proposed. ([5]) The most interesting suggestion is to see the asymmetry as temporal: (A) is an adequate explanation because the falling of the meteorite precedes mister X.Y.'s death. In general, the phenomena described by $C_1 \ldots C_n$ have to precede those described by E. This proposal is the most interesting one because it can be integrated in a causal model too (see section 4). Since it is impossible to comment on each proposal, I will only examine the time criterion.

Many examples may be given to falsify the time criterion. These counterexamples can be of two kinds: some examples show that the criterion is too restrictive, some others show that it is too wide as well. An example of the first kind is the following. I want to explain why my car has a certain acceleration. The following explanation would be completely satisfying:

C_1: My car weighs 1500 kilos. (M = 1500 kg)
C_2: The force exerted by its engine equals 4500 N. (F = 4500 N)
L: F = M.a (a = acceleration)

E: The acceleration of my car equals 3 m/sec^2.

This explanation is in complete agreement with our intuitions, though c_1, c_2 and e (the events described by resp. C_1, C_2 and E) occur simultaneously. Only the mass of my car at the moment of the acceleration is relevant, not the mass it had before. The same may be said about the force exerted by the engine. A well-known example, resembling the present one, is the explanation of the period of a pendulum. The relation between the length of a pendulum (L) and its period (T) is given by the

([5]) See e.g. B. VAN FAASSEN (1980) pp. 130-133, J. WOODWARD (1984), W. SALMON (1984) pp. 94-96.

formula $T = 2\pi \cdot \sqrt{L/g}$ ($g = 9.8$ m/sec^2). One can explain the period of a pendulum by referring to its length, though only the length of the pendulum while it is oscillating is relevant. Since in these cases no temporal order may be discovered, whereas the explanations are intuitively acceptable, the time criterion is too narrow. An example indicating that the time criterion is too wide, is the following:

C: Patient P exhibits symptoms S_1, ..., S_n on Monday.
L: Each patient who exhibits S_1, ..., S_n, will subsequently exhibit symptom S_E.

E: Patient P exhibits symptom S_E on Wednesday.

Although the requirements are met, this is not an adequate explanation. The occurrence of the symptom should have been explained in terms of an offending organism or material.

An adequate criterion to solve the problem of explanatory asymmetries is, in my opinion, the following: C_1, ..., C_n, L_1, ..., L_n is an adequate explanation of E if and only if by doing c_1, ..., c_n, we can bring about e. This requirement implies that (i) C_1, ..., C_n must be a sufficient condition for E, and (ii) that there has to be a productive power, a possibility of manipulation in one direction, viz. from explanans to explanandum. The asymmetry of explanations is a "strategic" asymmetry. For statistical explanations, the formulation has to be slightly different: "... we probably can bring about e". This criterion can cope with the problems which arose for the time criterion. Since I can act upon the acceleration of my car by changing its mass or its power, it is in perfect agreement with our criterion that the car example was intuitively adequate. On the other hand, the following example violates both our intuitions and the criterion:

C_1: $a = 3$ m/sec^2
C_2: $F = 4500$ N
L: $F = M.a$

E: $M = 1500$ kg.

No one would say that this helps us to understand why a car has its characteristic weight, though this construction is surely useful when we want to derive the weight if it is not known. The other examples don't cause any problems either. One may change the period of a pendulum by changing its length, but not vice versa. Consequently, the length can explain the period, but not conversely. Symptoms of a disease are brought about by offending organisms or materials, not by other symptoms. However, a new problem seems to arise: the criterion can only be used if we don't exceed the domain in which human intervention is possible. This forces us to modify our criterion: C_1, ..., C_n, L_1, ..., L_n is an adequate explanation of E if and only if: if I could do c_1, ..., c_n, I could bring about *e*. This formulation refers to possible human action, not to real human intervention.

In a recent book, F. Wilson denies the existence of explanatory asymmetries. (6) One of his examples is the pendulum example we already mentioned. One can change the period of a pendulum by changing its length, but not conversely. Wilson agrees that such a strategic asymmetry exists. What he denies is that this asymmetry allows us to conclude that there is also an explanatory asymmetry. Wilson distinguishes two kinds of explanations: explanations of how a system came into existence, and explanations of how a system works. Explanations of the first kind have to respect the strategic asymmetry, those of the second kind don't. In the first case, the length of the pendulum has to be explained by e.g. "Because I made it so", in the other case the length can be explained by the period. Wilson's mistake is that he mixes up two different aims of science by calling them both "explanations". One aim of science is to help us understand certain problematic phenomena we are confronted with. In this case, our starting-point is the problematic phenomenon: we know that E is true, otherwise we wouldn't ask for an explanation. The other aim is to predict or control the occurence of certain events. In this case, our starting-point is the fact that a set of singular statements and a set of law statements are true. Though the logical structure is the same as for explanations, there is a pragmatic difference: the truth of the conclusion E has to be derived. If E were already known, the prediction would be useless. These pragmatic differences may cause structural differences, so we can't assume that the structure of explanations and predictions is completely identical (though there are some similarities: predictions are derivations, have to conform to the covering law model, and deterministic predictions are ideal; so the logical basis is identical, but other differences may appear). Consequently, explanations and predictions have to be analysed separately. In this paper, the aim is to determine the structure of explanations; the

(6) F. WILSON (1985) pp. 96-105.

structure of predictions is not examined. What Wilson calls an "explanation of how a system came into existence" is really an explanation: the starting-point is the state of the fundamental part(s) of the system (i.e. the part or parts whose state influences the state of all other parts of the system: one can change these latter states by changing the state of the fundamental part(s), but not vice versa); the state of this fundamental part is known, but surprising, so an explanation is asked. On the contrary, Wilson's "explanations of how a system works" are not homogeneous. An analysis of how a system works is meant to answer three kinds of questions. First of all, the states of the other parts (which are not fundamental) are to be explained. These states are the starting-point, and the strategic asymmetry is to be respected. Secondly, questions of the type "How may I change the state of part X?" must be answered. It is obvious that the strategic asymmetry is to be respected here too, though this is not an explanation. Finally, our problem may be that we want to determine an unknown state. In this case, our question is not a why-question, but a what-question. For instance, we don't ask "Why is the length of the pendulum 1 meter?", but "What is the length of the pendulum?". Wilson is right when he says that in this case the strategic asymmetry isn't always respected, but these are not explanations.

3.4. *Conclusions*

The conclusions of this preliminary analysis can be stated as follows:

1° All explanations are derivations and conform to the covering law model.

2° In order to be complete, an explanation has to deductive-nomological. Statistical explanations are useful ("minimally sufficient") when they answer some relevant questions, which still are to be specified.

3° Explanations must respect the strategic asymmetry. This requirement is not just an ideal. It is a necessary condition for an explanation to be minimally sufficient.

4. *An ontic model of explanations*

The epistemic model in section 3 was developed by adding restrictions to the initial requirement that an explanation has to be a derivation. Each restriction is imposed in order to solve a problem which can be discovered by means of an example which points out that some intuitively acceptable explanations don't meet our criteria, or that some intuitively unacceptable explanations are not excluded. Every epistemic model is totally dependent on these examples: it is the only way to discover problems and to argue in support of or against a criterion. An ontic model is developed in a completely different way: the criteria are derived from a general cognitive value, viz. that an explanation has to identify the causal mechanism that led to the event that is to be explained. In other words: the characteristic features which are found when we analyse causal relations are transformed into restrictions on explanations. Therefore, causal models can also be called "value-driven", while espistemic models can be called "problem-driven". Though the way criteria are introduced is different, the result should be identical: the criteria that can be derived from our general cognitive value must be sufficient to cope all known problems, and the criteria we introduce to solve the problems we discovered have to be derivable from the requirement that explanations must be causal. Since I think that the analysis in section 3 is basically correct (though not complete: see section 5), I am convinced that a causal model has to lead to the conclusions formulated in 3.4. The analysis of causation I will give below will show that this requirement is fulfilled.

The first criterion to be examined is the requirement that every explanation has to be a derivation. The corresponding characteristic of causal relations is that they are based on a sufficient condition relation (in the trivial or the counterfactual sense): a cause must be a sufficient condition for its effect. Is this really a characteristic of causal relations? Some people will deny this and hold that a cause is only a part of a sufficient condition. According to J. Mackie(7), a cause is an INUS condition for its effect: an Insufficient but Non redundant part of a Unnecessary but Sufficient condition. Sufficient conditions are usually complex: a set of simple events A_1, ..., A_n is jointly sufficient for B. A cause is one of these events, which is for some reason preferred, while the others are neglected. So two kinds of causes can be defined: a sufficient cause, based on a sufficient condition relation, and a cause based on Mackie's INUS condition, which can be called an "excellent" cause. Since the decision to prefer one INUS condition is not based on objective criteria, but depends on the background knowledge of the people involved, excellent causes

(7) J. MACKIE (1974) p. 62.

can't be taken into account if we want to develop a general model of explanations (see 3.1). Consequently, only sufficient causes are relevant. This allows us to conclude that the first criterion can be derived from an analysis of causation.

Since we want to analyse explanations of types of events (i.e. explanations of why an event belongs to a certain type, not explanations which explain all the particular features of an event), causes and effects will be types of events too. We don't have to give a (sufficient) cause of all the particular features of the explanandum event. All we need is a (sufficient) cause of some of its features, viz. those that are mentioned on the explanandum sentence. Since causal relations between types of events are repeatable phenomena, they are expressed by regularities. Consequently, the causes we are looking for are not based on a trivial sufficient condition. The relation that is required, is: If A, then always B (where A and B are types of events). This means that the necessity of the covering law model can be derived. We can now formulate a first definition (instead of "sufficient cause", I will simply write "cause"):

(Dfl) A is a cause of B if and only if: if A, then always B.

This definition assumes that causes are always deterministic: the idea of a probabilistic cause is a contradiction. It is obvious that such a definition is unacceptable. But though the definition should be modified in such a way as to include probabilistic causation, deterministic causation still can be seen as the strict sense, while a probabilistic cause is only a cause in a wider sense. In other words: when we say that an explanation has to identify the causal mechanism which leads to the events that it is to be explained, the term "leads to" can be interpreted strictly (referring to a deductive derivation) or more loosely (referring to probabilistic cases too). From this, we can derive our conclusions about statistical explanations (see 3.2).

The last criterion that is to be derived is that explanations have to respect the strategic asymmetry. The corresponding characteristic we have to find is that the idea of causation is intrinsically linked with concepts like production, productive power and manipulation. Causal relations can't be defined in terms of conditional relations alone. A well-known example to illustrate this is that both "If it's raining, the streets get wet" and "If the streets are dry, it isn't raining" are true, while only the first sufficient condition is a cause. We can try to solve this problem by introducing a time criterion: a cause has to precede its effect. In this way a second definition is obtained:

(Df2) A_t is a cause of $B_{t'}$, if and only if: (i) if A_t, then always $B_{t'}$, and (ii) $t < t'$.

This solution was chosen by e.g. P. Suppes for the definition of this probabilistic causes and by W. Salmon to develop his ontic model of explanations. (8) Two arguments can be given against this time criterion. Each of them corresponds to one of the two kinds of counterexamples against the time criterion as a solution of the problem of explanatory asymmetries (see 3.3). Firstly, many examples of simultaneous causation can be found. For instance, the exertion of a force upon an object and the acceleration resulting from it are simultaneous. Other examples are the relation between the temperature and the pressure of a gas (when mass and volume are constant) and the pendulum example we already mentioned. These cases show that the time criterion is too restrictive. On the other hand, it is possible that, though the requirements of (Df2) are fulfilled, there is still no causal relation: A and B can be non-simultaneous effects of a common cause.

Causes can be correctly distinguished from mere sufficient conditions if the following definition is accepted:

(Df3) A is a cause of B if and only if by doing A, we can produce (or: bring about) B.

In this definition, which we owe to G.H. Von Wright (9), two requirements are combined, viz. (i) if A, then always B, and (ii) the existence of a productive power of A in respect of B. This ability to produce is the distinctive mark of causal relations. That this criterion solves the problems which arose for the time criterion, is quite obvious. So we can conclude that in our world-view the ideas of causation and production are intrinsically linked. However, this definition is only useful in the domain where human intervention is possible. Our last definition solves this problem:

(8) P. Suppes (1970) p. 12 and 80; W. Salmon (1984).
(9) G.H. Von Wright (1971) pp. 64-76.

(Df4) A is a cause of B if and only if: if we could do A, we could produce (bring about). B.

Of course, another modification is necessary in order to include probabilistic causes, but this is not important here.

This analysis of causation shows that there is no difference between the conclusions of the epistemic approach and the requirement that every explanation has to be causal: the set of adequate explanations is identical. So there seems to be no reason at all to prefer an ontic model. This conclusion would be overhasty. I think the causal approach has three advantages, which justify our preference:

1° A causal approach reveals the interconnections between the criteria: instead of ad hoc criteria without any coherence, we get a system of interrelated criteria, the coherence of which is the result of their being all concrete consequences of the same abstract cognitive value. So, only a causal model enables us to understand why certain criteria are to be introduced and why others don't fit out intuitions.

2° An epistemic model is completely dependent on the discovery of counterexamples. New criteria can only be introduced when a problem is discovered. Moreover, there is no guiding principle to introduce these new criteria: each proposal is an unreasoned conjecture. In a causal model, new criteria can be found even if there are no known problems for the old criteria. Consequently, the probability of the set of criteria being complete is very high. Moreover, the causal approach tells us where to look for our criteria: correct criteria can be found by transforming characteristic features of causal relations into requirements on explanations. So, a causal model provides us with heuristic rules for developing criteria. Of course, these heuristic rules are to be tested to see if they are succesful. But as we have shown in our analysis of causation (and also will show in section 5), the result of this test is positive.

3° Though the arguments of an epistemic model surely are conclusive (if a counterexample is found, one of the criteria is to be changed or an additional one is to be found), this argumentation is entirely negative. One can falsify a criterion (or a set of criteria), but it is impossible to verify one: no matter how many examples can be given to support a criterion, ther is no warrant that these examples differ enough to be representative. On the contrary, if it is assumed that explanations must be causal, positive evidence can be derived from this principle.

5. *Two kinds of explanations*

Though I am convinced that the models I proposed don't contain any criteria which are not correct, I think that some additional requirements are necessary. These will be examined in the present section. The causal and the epistemic approach will be combined. To begin with, I will give an example to illustrate why additional criteria are necessary. Secondly, I will derive a solution for the problem from the principle that explanations must be causal. Finally, this solution will be examined by means of some examples.

Consider the following example:

C: The annual income of mister M. has constantly increased during the last twenty years.
L: When someone's income increases, the relative part of this income spent to buy and maintain cars (expressed as a percentage of the total income) increases too.

E: The relative part of his income, spent by mister M. to by and maintain cars, has constantly increased during the last twenty years.

Explanation seeking why-questions can be of two sorts. Firstly, it is possible that the questioner is perfectly acquainted with the different mechanisms that can cause the event which is to be explained. His only concern is to know which mechanism really functioned in the case he is examining. Here, the laws used in the explanation are not important: all the questioner wants to know is whether certain singular statements can be accepted as true. On the other hand, the questioner may be convinced that certain events (which have caused the event that is to be explained) have happened, without seeing these events as explanatory, i.e. without knowing that they are causes of the problematic event. In this case, not the singular statements, but the laws are important. Consequently, those laws will have to meet certain additional criteria, as we will see later. The example given above can clarify this differentiation of explanation seeking questions. When someone asks why mister M. has spent a greater part of his income to buy cars, the questioner may be perfectly acquainted with the mechanisms which cause changes in a consumer's pattern of spending. He just wonders whether it is reasonable to assume that mister M.'s income has increased. But it is also possible that the questioner knows for sure that mister M.'s income has increased, but fails to see the relevance of this fact. In this latter case, the explanation given above would not be adequate: the law used in it only mentions input and output of the mechanism, without clarifying what happens in between. However, a why-question of the first kind would be perfectly answered by this explanation. So answers to why-questions of the first kind don't have to meet further requirements, while explanations of the second kind have to use laws which give a sufficiently detailed analysis of the causal mechanism involved, and make it understandable for the questioner. The same idea can be stated as follows. In the first case, the explanations are extensional: the description of the causal mechanism can be replaced by every other description with exactly the same extension. Explanations of the second kind are not extensional: the way in which the causal mechanism is described determines whether the explanation is adequate or not.

Explanations of the second kind have to analyse the causal mechanism in a way that satisfies the questioner. It is obvious that such a satisfactory description has some individual aspects (for instance, the same explanation can be easily understandable for some people, but far too difficult for others). But as we will see, there are some characteristics which are common to all satisfactory descriptions. If the arguments we gave in section 4 to prefer a causal approach are correct, it should be possible to derive these common characteristics from an analysis of causation. Causal relations exhibit two features that were neglected till now. Firstly, causal mechanisms are continuous: there are no great gaps between the successive stages of the process. Secondly, each scientific discipline (or each scientific research tradition) only regards some characteristic processes as real. Each causal mechanism is then supposed to be a chain of such fundamental processes (examples will be given below).

The first additional criterion that can be derived now is that explanations of the second kind have to describe the causal mechanism in such a way as to guarantee the highest possible degree of continuity between the successive stages of this mechanism. This criterion functions as an ideal: the more continuity there is, the greater the explanatory power. But it is also a necessary condition: if there is no continuity at all, the explanation is worthless. Let's return to the example to examine this criterion. Cars are, unlike clothes or food, not necessary to survive. It is hardly surprising that if the income is low – this means: when one has to choose which products one will buy and which not – the consumer's first concern is to buy enough necessary products. So luxury products like cars are only bought when the income is sufficiently high to buy both necessary products and luxury goods. This mixture of economical facts and psychological mechanisms establishes the relevance of mister M's higher income. Why is this description so revealing? Because there is a lot of continuity in the mechanism as it is described now. There are no big cleavages between two successive stages; only slight changes occur, which are easy to grasp. Many other examples can be constructed to illustrate this point. For instance, if a higher pressure in a gas container is explained by saying that the temperature rised while volume and mass were constant, one may wonder why the temperature of the gas is relevant here. Indeed, temperature and pressure are not phenomena that are at first sight closely related to each other. But when someone tells you that the temperature is in fact the average kinetic energy of the molecules, and the pressure the force exerted by these molecules, only slight differences between the successive stages of the process are left.

This continuity requirement has an important consequence. For each discipline there is a characteristic level of reality one cannot go beyond. If entities of a lower level were taken into account, it would become impossible to formulate adequate laws. For instance, social sciences mustn't go beyond the level of individual human actors, because social or economic laws can't be formulated in terms of cells or molecules. But within these limits, the lowest level of reality must be chosen, i.e. the laws used in the explanation must be laws on that lowest level. Why? Because there is more continuity on the relative micro-level than on higher levels. The lower the level of description is, the more continuity is possible. This is illustrated by e.g. the gas example above. So when as much continuity as is possible is to be attained, a necessary, but not sufficient condition to meet this requirement is that the explanation uses laws on the lowest possible level of reality.

The second additional criterion we can formulate is that explanations of the second kind have to analyse the relevant causal mechanism in terms of the processes that are considered as fundamental and not analysable by the scientists of the discipline involved. This criterion is not a minimal requirement, but an ideal: the explanatory power increases when more fundamental processes are used in an explanation. Why do we have to prefer these mechanisms? Because they are supposed to be intelligible at first sight: further analysis of these mechanisms is not necessary. For instance, most physicist will regard transformation of kinetic energy into potential energy, and vice versa, as a process that is not in need of further clarification. Neither is it to be explained how gravitation works. Behaviouristic psychologists didn't require any further analysis of the stimulus-response-scheme. Analogous examples can be found for each discipline. It should be noticed that the criteria on explanations are to a certain extent individualised: different disciplines have different additional criteria.

One problem with respect to these additional criteria is still to be solved. If all explanations have to be causal, why do only some explanations have to meet requirements, and why are these requirements just ideals? The answer is that we have to distinguish between what the relation between explanans and explanandum *is* on the one hand, and the *description* that is given of this relation on the other hand. If we say that all explanations have to be causal, this means that the relation between explanans and explanandum must be of a certain type (this requirement includes the additional features that were dealt with in this section). However, it does not follow that in each explanation we have to *demonstrate* that the relation exists in the instance that is examined. Demonstrating that a causal relation exists is only relevant if we have to give an explanation of the second kind.

6. *Conclusions*

Two conclusions can be added to those formulated in 3.4, viz. (i) explanations of the second kind have to meet the continuity requirement and the "fundamental mechanism requirement", and (ii) these requirements, as well as the conclusions of 3.4, are consequences of the principle that explanations must be causal.

The analysis of explanations given in this paper was not intended to be complete. First of all, explanations of particular events were neglected. Next, some problems about statistical explanations are still to be solved (e.g. the classification of the questions that are relevant to determine the explanatory power of statistical explanations). Finally, the heuristic value of the causal approach has to be exploited further: it has to be applied to problems that were not treated here, and we must examine if problems that are hitherto unknown, can be discovered.

Rijksuniversiteit Gent
Seminarie voor Logica en
Wijsbegeerte van de Wetenschappen

Erik WEBER

REFERENCES

Achinstein P., (1983), *The nature of explanation*, New York & Oxford.

Gärdenfors P., (1980), *A pragmatic approach to explanations*, in: Philosophy of Science 47 pp. 405-423.

Hempel C.G., (1965), *Aspects of scientific explanation and other essays in the philosophy of science*, New York.

Kitcher P., (1985), *Two approaches to explanations*, in: The Journal of Philosophy 82 pp. 632-639.

Laudan L., (1983), *Science and values*, Berkeley, Los Angeles & London.

Mackie J., (1974), *The cement of the universe. A study of causation.*, Oxford.

Pollock J., (1976), *Subjunctive reasoning*, Dordrecht.

Papineau D., (1985), *Probabilities and causes*, in: The Journal of Philosophy 82 pp. 57-74.

Salmon W., (1971), *Statistical explanation and statistical relevance*, Pittsburgh.

Salmon W., (1984), *Scientific explanation and the causal structure of the world*, Princeton, New Jersey.

Salmon W., (1985), *Conflicting conceptions of scientific explanations*, in: The Journal of Philosophy 82 pp. 651-654.

Stegmüller W., (1983), *Probleme und Resultate der Wissenschaftstheorie I. Erklärung, Begründung, Kausalität*, Berlin (second, improved and enlarged edition).

Suppes P., (1970), *A probabilistic theory of causality*, Amsterdam.

Van Fraassen B., (1985), *Salmon on explanation*, in: The Journal of Philosophy 82 pp. 639-651.

Von Wright G.H., (1971), *Explanation and understanding*, Ithaca, New York.

Wilson F., (1985), *Explanation, causation and deduction*, Dordrecht.

Woodward J., (1984), *Explanatory asymmetries*, in: Philosophy of Science 51 pp. 421-442.