[From *Contemporary Science and Natural Explanation. Commensense Conceptions of Causality.* D.J. Hilton (ed.) Harvester Press, 1988.]

2 The Problem of Causal Selection

Germund Hesslow

I THE PROBLEM OF CAUSAL SELECTION

1 Introduction: the plurality of causes

The problem to be discussed in this chapter arises from the fact that most of those events, facts, states or properties for which causal explanations are appropriate, have infinitely many causes. There are three reasons for this. Firstly, an event will normally depend on the immediately preceding occurrence of several different events. Secondly, it will usually be possible, at least in principle, to trace a causal chain backwards in time. Thirdly, it is generally possible to conceptualize the causes in infinitely many different ways.

To take a simple example, consider the event of my lifting a cup of coffee to my lips. Immediately preceding this there are several important conditions, such as the weight of the cup, its position in my hand, the position of my hand and arm, the contraction of the appropriate muscles etc., all of which are necessary conditions for the final event's taking place. If we attempt to trace the causal chain backwards, complexities increase exponentially. There will be causes of the cup's having the position it has, and causes of my arm's having the position it does. The muscular contraction will be preceded by calcium ions flowing into the muscle cells, because of the cell's being depolarized, because of the binding of acetylcholine to the receptors on the muscle cell membrane, because of transmitter release from nerve endings etc. And we still have a long way to go before reaching the neural events behind the intention to lift the cup.

However, when we explain why an event occurs, we never mention more than a few, usually just one, of the events making up

this complex web of causal antecedents. No one, for instance, would explain my action by mentioning the release of acetylcholine from the motor nerve endings. Nor do we explain the collapse of a bridge by the action of gravity or a fire by the presence of oxygen.

It could, of course, be maintained that gravity is not a cause of bridges collapsing or that oxygen does not cause fires, but this answer is unsatisfactory. One reason is that all philosophical theories about the nature of the causal connection would assign the same role to, for instance, the presence of oxygen, a heat source, inflammable material etc. as causes of a fire. Moreover, intuition suggests that, although there is certainly a difference between, say, oxygen and a lighted match, we nevertheless recognize *some* role for oxygen. Even if we are reluctant to do this in ordinary situations, like the fire in a house, we can easily imagine situations in which the presence of oxygen would be regarded as the crucial condition. Suppose, for instance, that we have a highly inflammable substance at very high temperature in an oxygen-free container. If we then were to let oxygen into the container, and the substance caught fire, we would surely regard the entry of oxygen into the container as an important cause.

Thus we are faced with the situation that a normal event has many, perhaps infinitely many, causes, but that only some of them are selected and cited in causal explanations. Sometimes we even speak of *the* cause of an event which actually has several causes. Why is this so? What determines the selection of the most important cause from the complete set of causal conditions? This is the selection problem with which we shall be concerned in this chapter.

It will be assumed, without detailed argument, that the selection of a cause from a set of conditions is a special case of the weighting of causes according to their relative importance. For instance, although we might explain someone's alcohol problems by their biochemical susceptibility to alcohol dependence, we might also concede that other factors, such as personal problems, were contributory. When the selection criterion unequivocally picks out one condition we call this *the* cause, but when other conditions come close to satisfying the criterion these are termed contributory, and the condition which best fits the criterion is considered more important than the others.

It should be stressed that the problem of understanding what is involved in causal selection is of double relevance to researchers in

psychology. It is a theoretical problem in cognitive psychology, because it concerns the way that common sense conceptualizes causes, but it is also a central methodological problem, because causal attribution is among the central activities of the scientist. Since any causal hypothesis in science depends on causal selections, one cannot understand the epistemological status of such hypotheses without an understanding of causal selection. Since it has been claimed that these selections are arbitrary and governed by subjective factors, the existence of causal selection constitutes a challenge to scientific objectivity. (see e.g. the exchange by Martin, 1974, 1978a, 1978b and Frey, 1976, 1978a, 1978b)

In the following, various attempts by philosophers to formulate criteria which govern causal selections and weightings will be reviewed. An alternative theory will also be outlined which, it will be argued, unifies these attempts but avoids some of their difficulties.

2 Two basic distinctions: selections versus connection, individual versus generic

Much of the philosophical discussion of the selection problem is based on the assumption that a causal statement rests on two distinct judgments, each with its corresponding conceptual problem, one concerning the existence of a causal relation between two events, the "connection problem", and one concerning the relative importance of these causes, the "selection problem". The connection problem is the problem of understanding the process by which we determine that, say, the presence of oxygen, combustible material and a heat source are all necessary conditions for houses catching fire. When the causal relevance of these conditions has been ascertained, there remains, however, the question of determining which of these conditions was, in a concrete individual case, the most important one. We do not say that a fire was caused by oxygen, in spite of the fact that we know that there is a causal connection between oxygen and fire.

The assumption that the two problems can be separated is not trivial and has been denied by some writers e.g. (Hart & Honoré 1959). We will return to this question later; for the time being our task is only to understand how and why, out of the set of causal conditions, the *complete cause*, we select one as *the* cause or as the most important cause.

A second important distinction is that between individual and

generic causal relations. The first kind are those relations which obtain between concrete individual occurrences of events, such as the house's catching fire at 9.05 pm yesterday because of the explosion in the television set a moment earlier or the fact that Smith's recent death was caused by a heart attack. The second kind are the relations which obtain between kinds of events (generic events) or between properties, such as the general propensity of explosions to cause fires, or the fact that heart attacks cause death.

There are differing views on the relationship between these two kinds of causal relation. The most common one is that we arrive at generic causal relations by generalizing from individual cases of co-occurrence and then apply this general knowledge to other individual occurrences. Thus, since a large proportion of those who have heart attacks die, we conclude that the disease is deadly. If Smith has an infarction and dies, we use our knowledge of the general causal relation to justify the belief that his death was caused by the infarction. (Note, however, that a general causal statement can be true while a corresponding individual statement is not. Smith's heart attack may not have killed him and he may have been killed by something else. Cf. Hesslow, 1981b)

Statements about the relative importance of causes may apply to both generic and individual causal relations, although the meaning of such statements will be very different. If it was said, for instance, that short-circuits are more important causes of fires than are explosions, the meaning would be that short-circuits cause a greater number of fires than explosions do, and it would be assumed that only one cause was important in each individual fire. Statements of this kind do not involve any theoretical problems beyond those of individual selection.

On this view, the connection problem is identical to the problem of how we arrive at general causal statements, while the selection problem is, in a sense, the problem of justifying various applications of general causal knowledge to particular individual events.

II EARLIER APPROACHES TO THE SELECTION PROBLEM

3 Review of selection criteria

The first philosopher to recognize that the conditions which

causally determine an event usually far outnumber the conditions mentioned in a causal explanation, was John Stuart Mill. In *A System of Logic* he advanced an analysis of causality which was essentially a development of Hume's so called "regularity theory". An event *A* is a cause of the event *B*, according to this theory, if events of the kind *A* are always followed by events of the kind *B*. One problem with this analysis is that few causes are invariably followed by their effects as the theory requires. It may be true that a fire was caused by a short-circuit, but short-circuits are not normally followed by fires. Mill recognized that "It is seldom, if ever, between a consequent and a single antecedent that this invariable sequence subsists." However, "It is usually between a consequent and the sum of several antecedents; the concurrence of all of them being requisite to produce . . . the consequent. In such cases it is very common to single out one only of the antecedents under the denomination of Cause, calling the others merely Conditions." (III, v, 3). Mill stressed that, from a scientific point of view, there could be no justification for a differential treatment of the causal conditions. "We have, philosophically speaking, no right to give the name of cause to one of them, exclusively of the others." (III, v, 3). It is important that causal selections, according to Mill and to many other philosophers, are logically *arbitrary*. Nevertheless, Mill did offer an explanation for the selections we actually make.

(a) Unexpected conditions. According to Mill,

> If we do not . . . enumerate all the conditions, it is only because some of them will in most cases be understood without being expressed, or because for the purpose in view they may without detriment be overlooked. For example, when we say, the cause of man's death was that his foot slipped in climbing a ladder, we omit as a thing unnecessary to be stated the circumstance of his weight, though quite as indispensable a condition of the effect which took place (III, v, 3).

Thus, on this view some conditions are not mentioned because they are presumed to be already known to the listener, and stating them explicitly would be superfluous. Consequently, we select as causes only such conditions that are unknown or unexpected.

A similar criterion, but with a different rationale, is given by William Dray. "To explain a thing", according to Dray, "is sometimes merely to show that it need not have caused surprise" (1957, p. 157). The idea is that we do not generally require explanations when things behave normally. We ask "why" mainly

when something unexpected happens, and a relevant explanation will then state those events which were unexpected, but which would have enabled us to predict the surprising event if we had known about them (see Gärdenfors, 1980 for a theory of explanation along these lines). The conditions which explain unexpected occurrences will generally themselves be unexpected, hence the selection of unexpected causes.

(b) Precipitating causes. In discussing the example of the man slipping on the ladder, Mill also notes that in this as in many other cases, "the fact which was dignified with the name of cause, was the one condition which came last into existence" (III, v, 3). It is often possible to divide the complete cause into more or less permanent *states* and instantaneous changes or *events.* We usually select the events immediately preceding the effect which we are trying to explain (cf. MacIver, 1952). This is one of the few selection criteria for which there is direct support in ordinary language, namely when we talk of "precipitating" causes. In such cases, we explicitly use the distinction between permanent conditions and the instantaneous event which "last came into existence". Furthermore, the fact that we explicitly qualify the causal statement by calling the cause precipitating, suggests that we are aware that the other conditions are also causes of a kind. Note that precipitating conditions are not necessarily the same as unexpected conditions. When a match catches fire, the precipitating cause will normally be the striking of the match, but this need not be unexpected.

(c) Abnormal conditions. A similar view is that we select conditions which are abnormal or unusual. This criterion was proposed in H.L.A. Hart and A.M. Honoré's book *Causation in the Law* which contains a very penetrating discussion of the distinction between causes and "mere" conditions with legal problems in mind.

> In a railway accident [mere conditions] . . . will be such factors as the normal speed and load and weight of the train and the routine stopping and acceleration. These factors are, of course, just those which are present alike both in the case where such accidents occur and in the normal cases where they do not; and it is this consideration that leads us to reject them as the cause of the accident, even though it is true that without them the accident would not have occurred.

It is important to notice the motivation given by the authors, that "to cite factors which are present both in the case of disaster and

normal functioning would explain nothing: such factors do not 'make the difference' between disaster and normal functioning, as the bent rail [does]" (Hart & Honoré, 1959, p. 32). It may seem over-sophisticated to make a distinction between unexpected and abnormal events, but these words reflect a substantial difference. Firstly, abnormality refers to objective facts; things are normal or abnormal independently of our knowledge of them, while unexpectedness refers to a subjective state. Secondly, the motivation given by Hart & Honoré in terms of explanatory relevance is quite different from that given by Mill, and it depends on an objective feature of situation. The weakness of the train's speed as an explanation for the accident is that it is the same both in cases where accidents occur and in cases where they do not. It is not that we expect the speed to be high.

(d) Variability. Selection of those conditions which are variable in contrast to more permanent conditions (cf. Nagel, 1961) is a blend of the first three, and it is doubtful if a defence of such a criterion could be made that would not also support the others.

(e) Deviation from theoretical ideal. A relevant observation, in this context, is that certain abstract theoretical concepts often guide causal selections. Examples are provided by Weber's concept of "ideal types", equilibrium models in social science, e.g. the perfectly working market economy in neo-classical economics, the definition of a "wild type" in bacterial genetics, the physiology of the healthy human organism in medicine etc. These *theoretical ideals,* as we may call them, define appropriate causal selections. For instance, in explaining a deviation from the market equilibrium or from physiological health, we select causes which are also deviations from the market or deviations from physiological health. This is very similar to Hart & Honoré's conception of selecting deviations from the normal course of events by causes which are also abnormal, but it differs in that no assumption need be made that market equilibrium is normal or that perfect health is normal.

Theoretical ideals are very similar to what Toulmin has called "ideals of natural order". In Aristotelian physics, a material body strives towards a state of rest on the ground, and the theory attempts to explain why certain bodies deviate from this state. The fact that a body is at rest on the ground does not need an explanation and, indeed, cannot be explained by the theory. Only deviations can. In Newtonian physics, a corresponding ideal of natural order is given

by the first law of motion as "a state of rest, or of uniform motion in a right line". The theory does not explain the motion of a body which conforms to this ideal, except in the vacuous sense that a body cannot change its state of motion because no force is acting on it. In a sense, forces are equivalent to causes in Newton's theory, and a state of motion in a straight line has no causes. These are defined away by the theory. Only deviations from this kind of motion have causes and can be explained.

(f) Responsibility. It has frequently been noted that causal statements may have an evaluative component. Indeed, the Greek word for cause, *aitia*, also means guilt. According to some historians (e.g. Kelsen, 1943) the Greeks modelled their idea of causation in nature by analogy with ideas about social organization. A cause was thought of as something that brings about a disturbance in state of harmonious equilibrium in nature, and the effect as something that restores this equilibrium, much as a punishment restores the social harmony after a crime. The idea that cause and effect somehow must equal each other also has a moral counterpart in the idea that the punishment should be proportional to the seriousness of the crime. The Latin word *causa* was originally a legal term. The moral term "responsibility" has a similar double meaning. In a scientific laboratory one may hear it said that "this or that factor was responsible for the failure of an experiment", even when it is clear that the issue is one of causality. If there is a standard view about this, it is that we identify the cause of a tragedy before assigning blame. However, it may be claimed that in selecting among the causal conditions we pick out those events or actions which deviate, not from what is normal, but from what is good, reasonable or appropriate. For instance, Dray maintains that "A cause will often be an omission which coincides with what is reprehensible by established norms of conduct" (1964, p. 56). Thus, when we say that a fire was caused by negligence of the authorities (who failed to notice the special dangers in the building), we are not denying that oxygen, a heat source etc. had something to do with it. Neither are we saying that negligence is abnormal. We are, rather, specifying what went wrong.

(g) Predictive value. According to a widespread view, an explanation for a certain event consists of information that, had we had access to it before the event to be explained occurred, would have enabled us to predict it (see e.g. Hempel, 1965; Gardiner,

1952; Gärdenfors, 1980; Nagel 1961; for critical discussions of this theory see e.g. Dray, 1957; von Wright, 1971; Toulmin, 1961). In view of this, and also for other reasons, a natural and intuitively compelling selection criterion would be that we select as the most important causes those that most effectively predict the effect. In formal terms, this means that C_1 is a more important cause of E than is C_2 if and only if,

$$P(E/C_1) > P(E/C_2)$$

i.e. if the probability of E is greater when C_1 occurs than when C_2 occurs. Criteria of this sort have been defended by Nagel (1961) and Martin (1972) (see also Hesslow, 1983 for a critical discussion).

(h) Replaceability and necessity. Most of us think about certain historical figures like Napoleon, Hitler or Lenin as being important causal factors in history. Historians sometimes take a different view and argue against the role of the individual in history that even if the person X had not done this or that, someone else would have done it instead, and therefore history would not have been much different. This argument does not deny that X did bring about certain things, only that X was not necessary. X's character, motives etc. may have been sufficient in the circumstances for whatever happened, but there were also other people with similar characters, motives etc., such that these too would have been sufficient. X was, we might say, *replaceable,* and therefore not as important a cause for historical developments as causes which were irreplaceable (cf. Mackie, 1974, p. 128).

Replaceability may be a matter of degree. In a famous and controversial book Robert Fogel (1964) argued that the American railroads were not as important as had previously been thought for the rapid economic development of the nineteenth century. The central argument was that if there had been no railroads, other means of transportation, for instance canals, would probably have replaced railroads and fulfilled the same role. Clearly, the strength of Fogel's argument hinges on the estimated probability that, in the absence of railroads, canals would be developed and would be able to take care of the necessary transportation. The more probable this is, the less important will the existence of the railroads seem. Causal importance in this situation will be inversely related to probability of replacement.

But a condition which is likely to be replaced is also one which is a

bad predictor of the effect. If railroads are likely to be replaced by some other condition which would have the same effect, then knowledge about the railroads is useless for predicting economic development. Thus the replaceability criterion reduces to the criterion of predictive value. The main difference between them is that predictive value focusses on the probability that the effect occurs, given the causal candidates, whereas replaceability focusses on the probability that the effect does not occur in the absence of the causal candidates. But, under normal conditions, these formulations are equivalent. (Another obvious difference is that the replaceability criterion is couched in "counterfactual" terms. That is, it requires a judgment about what *would* have happened, if something else *had* happened which in fact did not happen. Such judgments are not universally accepted as legitimate. For a discussion of the methodological problems involved, see Gerschenkron, 1968 and Lewis, 1973).

(i) Instrumental efficacy. Manipulability was suggested by Collingwood (1940) as a selection criterion. It is based on an instrumental view of the causal connection where cause and effect are related as means to ends. Thus, Collingwood refers to causes as levers by means of which we can produce or prevent certain effects. If causality is viewed in this way, it is very natural to think that we select those conditions which enable us to manipulate effects. If we want to bring about something, we will select conditions which come as close as possible to being sufficient for a desired end, and if we want to prevent something, we select conditions which come as close as possible to being necessary for whatever it is we wish to avoid. A related formulation, although not put forward as a selection criterion, is that of von Wright (1974). For von Wright, a cause is something that can be introduced as an intervention in nature by human action and that can be used to bring about other things, the effects. This is congenial to the view of the experimental scientist, who intervenes in nature by experimental manipulations and who typically selects the interventions to explain the experimental outcomes.

Instrumental effectiveness is similar in some ways to predictive value. The condition which is most likely to bring about a desired effect will usually, of course, be the one that gives the effect the highest probability. This is not quite the same thing as instrumental effectiveness, however, for some conditions, although they may be

very reliable predictors, may also be humanly impossible to influence, and thus useless as instruments.

(j) Interest. Carnap (1966) probably expresses a common view when he claims that causal selections are arbitrary and governed by the particular interests of the person giving an explanation. He gives the example of a car accident and the different explanations given by different people. Thus, a road engineer might point out that the road had a poor surface and that the cause of the accident was the slippery highway. A policeman might instead pick out some other factor, like the excessive speed of the car, and a psychologist yet another factor such as the driver's disturbed state of mind. According to Carnap, each person looks at the situation from a special point of view and singles out that factor that interests him or her most.

In a sense, Carnap's view amounts to a denial of one of our central presuppositions, namely that selection follows rules with a certain rationale. If this was correct, there would not be any *logical* problem of causal selection at all, only the trivial, from the logical point of view that is, problem of finding out how people's interests are shaped.

This list could no doubt be extended further, but I think that it is sufficient to illustrate the main approaches to the selection problem. It is clear that many of these suggestions are highly similar and it is probable that other suggestions will turn out to be variations on the themes outlined above.

4 Problems raised by theories of causal selection

There are two main problems raised by these proposed selection criteria. One concerns their epistemological status and a second their rationale.

Status and correctness of selection criteria. When it is said that selection proceeds according to some criterion or rule, it is implied that different people select causes in similar ways and that in doing this they adhere to some sort of convention. Thus, the idea of selection criteria is incompatible with the view that selection is arbitrary, and it was not quite correct in the review section to present as a proposed criterion the view that causal selection is guided by personal interests. However, most of the entries in the list have been proposed as criteria, and we must now address the question of how to evaluate them. Are any, or perhaps several, of them true or correct?

It is not self-evident what it would mean to say that any of the proposed criteria is "true" or "correct". One possibility is to interpret them as purely empirical, or "behaviouristic", *descriptions* of the causal attributions made by common sense. Most observers, when confronted with the list of selection criteria above, would probably find some truth in each of them. To those of us who like compromises, it is tempting to conclude that all, or at least most, of the criteria are true but that different criteria are used in different contexts (cf. van Fraassen, 1980). However, if we were to take this approach we would be faced with a new selection problem: how do we select, in each situation, the appropriate criterion of causal selection?

It seems clear that, although each criterion correctly describes selection in some contexts, none of them, taken by itself, correctly describes all causal selections. For each criterion, it is easy to find examples of causal selections which cannot be justified by that criterion alone (see e.g. the criticism of the "abnormality" criterion in Dietl, 1970). In a complete account, therefore, they would have to be supplemented by a description of which criteria are employed in which circumstances. But this would leave us in position scarcely better than the one we started from, for now we have to find out, not how causes are selected, but how causal selection rules are selected, and this is not very impressive progress.

Moreover, when the word "criterion" is employed, it is usually implicit that it is not meant only as description of "selection behaviour", but also that we somehow actually "use" the criterion. Rules of grammar, to take a simple analogy, are not just descriptions of the language actually produced by members of a linguistic community; the rules are also "employed", "followed" or "adhered to". It is for this reason that they are thought to have explanatory value. Claims of this sort are usually very difficult to support, however. Linguistic conventions are generally not consciously adhered to and are inferred from uniformities in linguistic behaviour. This problem is made worse by the fact that, although there are exceptions to most criteria, the majority of causal selections probably satisfy several criteria at once, and it will be a difficult task to say in such a case which criterion was actually operative. Again we will need an account of how selection criteria are selected.

A subsidiary issue concerns the epistemological status of those

conditions which are not selected. It is assumed by many writers, and explicitly maintained by some, that the individual normally knows about all the conditions. Indeed, it is implicit in the standard formulation of the problem as a problem of *selection*, that we are generally cognizant of those conditions from which the selection is supposed to be made. Thus, for instance, Carnap's suggestion that we select those conditions in which we are interested would be pointless if these were the only conditions we knew about.

This may often be the case, but quite as often, obviously, it is not. Some people probably do not know that oxygen is a necessary condition for a fire, but everyone knows several common causes of fires. According to Hart & Honoré (1959) it is not possible to separate the establishment of causal connections from the selection of causes, because we only discover the abnormal causes. For the experimental scientist, this is a natural view. In order to establish a causal connection between A and B, we need cases both where A is present and where A is absent. It is impossible to demonstrate the causal efficacy of permanent conditions.

This argument clearly has merit, but it does not solve the problem. It may be used to support several of the competing selection criteria. Proponents of variability, unexpected causes and manipulability could all use this argument. Nevertheless, a satisfactory theory of causal selection must account for the fact that although we often only know about one causal condition, we also continue to view this as the most important when science reveals other conditions.

Rationale of selection criteria. Assuming that one criterion is true in the sense that we really do select causes according to it, why do we do this? What purpose is served by focussing on, say, abnormal or manipulable conditions?

The "unexpectedness" criterion is mainly motivated by considerations of informational economy, at least as it is presented by Mill. Most of the others concentrate on one particular use of causal knowledge. There are three main reasons why the knowledge that A causes B is interesting to us. Firstly, knowledge of A may be used to *predict* B. Secondly, by manipulating A we may introduce or prevent, that is *manipulate*, B, and thirdly, A can be used to *explain* B. Some of the selection rules are straightforwardly explained in this way, e.g. manipulability and predictive value. Other cases are more problematic. It is difficult to

see, for instance, why abnormal causes should be particularly suited for any of these purposes.

The rationale of selection rules is interesting also because it might help us to explain the fact that we make different selections in different circumstances. It may be thought, for instance, that in a situation where the problem is to predict the future, predictive causes should be selected. When the problem is of a practical nature, such as determining the causes of disease, we should choose manipulable causes which will enable us to deal with these problems etc.

This suggestion has some initial plausibility, but it fails to solve the problem. Many causal problems are motivated by all potential uses, and causal selections have to be made before any specific use is thought of. Typically we look for an explanation for something that is important, and whose importance makes us want to both predict and manipulate it. The physician, for instance, seeks an explanation for the patient's symptoms, in order to be able to predict the course of the disease as well as to manipulate it. One selection must thus satisfy all three desiderata. I think it is clear then that the uses of causal knowledge cannot be separated.

III A NEW APPROACH TO THE SELECTION PROBLEM

5 The nature of events

In the following sections I will outline a different approach to the problem of causal selection and weighting which, it will be argued, better explains the selections we actually make, and which also overcomes the difficulties discussed above for other theories of selection. (Some aspects of the theory, including technical details, are omitted for reasons of space; cf. especially Hesslow, 1983 and also 1981a and 1984)

This theory rests on two ideas. The first is that the effect or the explanandum, i.e. the event to be explained, should be construed, not as an object's having a certain property, but as a *difference* between objects with regard to a certain property. The second idea is that selection and weighting of causes is determined by *explanatory relevance.* This may strike the reader as yet another addition to the already overlong list of proposals for selection criteria, but, as will be clear shortly, explanatory relevance is different.

In most writings on causality and explanation, philosophers have more or less taken it for granted that an event can be represented as a statement (or a "proposition", i.e. the meaning of a statement) having the form *Fa*, where *F* is a property and *a* is the individual object which has the property *F*. If we consult everyday languge, this may seem plausible. We ask why the bridge collapsed, the barn caught fire, the patient recovered etc. In these cases, objects (bridge, barn, patient) have properties (collapsing, catching fire, recovering), and it does seem to be the object's having these properties which we want to have explained.

However, if we think about it carefully, it is easy to find complications with this view. Consider, for example, a typical explanation for a friend's alcoholism, in terms of unemployment, depression or marital problems. We are not normally satisfied with such explanations, and the reason is that problems of this kind are very common, most of us have them to some extent; yet only some of us develop alcoholism. One way of putting it is that things like unemployment are at best partial explanations, and they do not, at least not completely, explain why this person became an alcoholic while that person did not. That is, they do not completely explain the difference between alcoholics and others.

But if we agree that this is a legitimate objection, then we must also agree that the explanandum is not simply one object's having a certain property, such as John's being an alcoholic, but a difference, such as John's being an alcoholic while Bill is not. We are, it seems, making a comparison between John and Bill. What I want to suggest, then, is that the explanandum should be construed as a relation which involves three things: an *object a*, an *object of comparison b* and an *explanandum property E* which *a* has and *b* does not have. We may abbreviate this

$$\langle a, E, b \rangle$$

In some cases the comparison will be between an object and a whole class of objects of comparison. When asking about Bill's alcoholism, we would normally compare him to all people who are not alcoholics. We may call this class of objects the *reference class*, and write the explanandum

$$\langle a, E, R \rangle$$

where *R* is the reference class.

Let me illustrate this point further with the help of another example. Consider the two fruit flies (M1 and N1) in the upper row

of Fig. 1 which have been raised under identical environmental conditions. If we now ask the question "Why does M1 have such short wings?", the natural answer is that the cause is genetic. Since the environment is the same, only genetic factors differentiate between M1 and N1. But suppose that we had never seen the normal flies, and that M1 had only been observed together with the mutant flies M2 and M3, both of which are genetically identical to M1 but have been raised in higher temperatures (27°C and 32°C respectively as compared to 22°C for M1). If we look at these flies and again ask why M1 has such short wings, the natural answer will be that it was raised in a lower temperature.

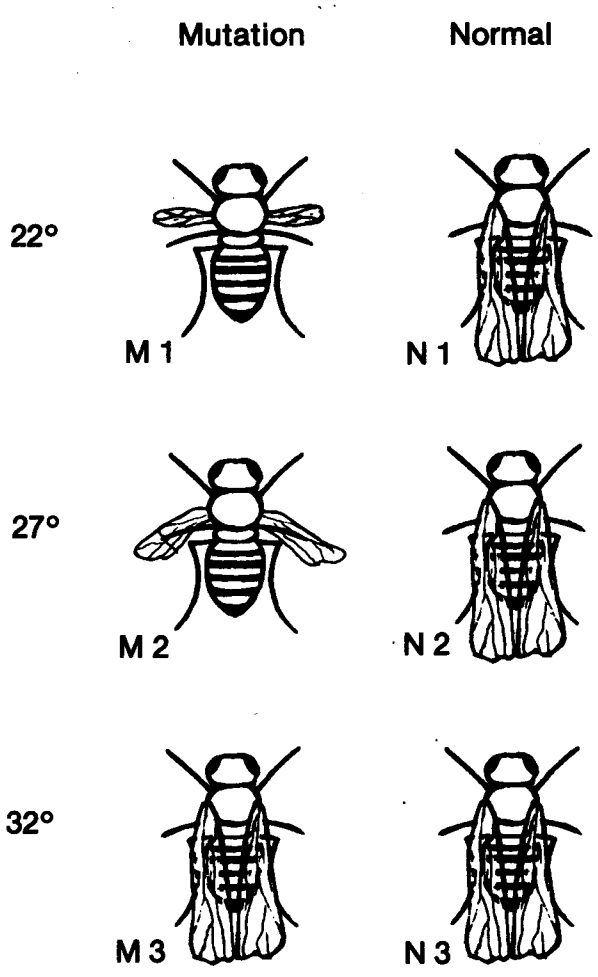

Figure 1

The apparent contradiction between these two different answers is resolved when it is realized that they are really answers to different questions. In the first case we are explaining, not why M1 has short wings, but why M1 has shorter wings than N1. In the second case we are explaining why M1 has shorter wings than M2 and M3. That is, different objects of comparison are related to different questions which in turn have different answers. (Cf. van Fraassen, 1980 for a different but related way of analyzing events.)

6 Explanatory relevance

Given the nature of explananda as differences, it is easy to see why some causal conditions, even if they are necessary conditions for the effect and even if they play a physical role in the outcome, should still lack explanatory relevance. Let us first consider the logical or formal aspects of the situation. Suppose that what we are interested in is Ea, and suppose that there are three conditions jointly sufficient and individually necessary for Ea, i.e. together they bring about Ea and the absence of any of them would ensure the absence of Ea ("$\neg$" indicates negation and "$\rightarrow$" the causal relation):

$$\text{I} \qquad \left.\begin{array}{l} C_1a \\ C_2a \\ C_3a \end{array}\right) \quad \rightarrow \quad Ea$$

Clearly, from a formal point of view, any selection from the set of conditions would be purely arbitrary. But suppose that we are interested in the difference between the above situation and the following specific case:

$$\text{II} \qquad \left.\begin{array}{l} \neg C_1b \\ C_2b \\ C_3b \end{array}\right) \quad \rightarrow \quad \neg Eb$$

Now our explanandum is not Ea but $\langle a, E, b\rangle$, i.e. the difference between I and II. If we want an explanation for the fact that a is different from b with respect to E, it is clear that C_1 is the only relevant explanation, for C_1 is the only condition that differs

between the two situations. Clearly we cannot explain a difference between two cases with a condition that is present in both. Furthermore, if we are explaining the difference between a and b with respect to E, we are not only explaining why a has E, but also why b does not have E, and we cannot explain why b does not have E with the presence of a condition that causes E.

For instance, if we want to explain why the fly M1 has shorter wings than N1, then the temperature in which the flies were raised is explanatorily irrelevant, since the temperature was the same in both cases. The mutated gene on the other hand was present in one case and absent in the other. It is, therefore, explanatorily relevant.

It is noteworthy that there is nothing arbitrary or subjective about this causal selection. Given that we construe the explanandum as a difference, the relevant selection becomes not only obvious but logically compelling. I have argued elsewhere (Hesslow, 1983) for the following analysis of what it means to explain a difference:

Ca explains the difference between a and b with respect to E, $\langle a, E, b \rangle$, if and only if,

(i) if Cb had been true, then Eb would have been true, and

(ii) if $\neg Ca$ had been true, then $\neg Ea$ would have been true.

For instance, that a condition C explains the difference between this barn, which caught fire, and that barn, which did not, implies that if C had not been present in this barn, it would not have caught fire, and if C had been present in that barn, then that barn would have caught fire.

7 The nature of causal selection

The advantage of this way of looking at explanation is that it dispenses with selection. This is not quite true, for I do claim that we select the explanatorily relevant conditions, but this is selection in a different sense from that discussed above. Let me explain this by discussing in turn traditional selection criteria and the choice of objects of comparison.

Many of the selection criteria listed in Section 3 can be construed as the result of choosing different objects of comparison or reference classes. Let us consider again the fire in the barn, and let us suppose that we have in the back of our minds the picture of a normal barn. This picture has been formed by previous experience and thus represents a kind of crude statistical average. If we ask why this barn caught fire, we will, unconsciously, be comparing this barn with the

statistically normal barn. Since the normal barn has not caught fire, it follows that an explanatorily relevant condition for this barn's catching fire must be abnormal. Thus, selection of abnormal conditions can be viewed as the result of comparing the explanandum object with a normal object.

Other criteria may be treated similarly. If we were comparing the barn that caught fire with the same barn before it caught fire, we would have to use a condition which was explanatorily relevant to the difference between this barn now and this barn yesterday, i.e. we would be selecting a precipitating cause. Selection of the unexpected may be viewed as the result of explaining the difference between an expected and an actual outcome. Selection according to responsibility follows from a comparison between actual and morally ideal behaviour. Selection of conditions which cause a deviation from a theoretical ideal involves a comparison between an actual and a theoretically ideal situation, and so on (cf. Hesslow, 1983).

A more difficult case is the situation where several conditions are explanatorily relevant. Say that we want to explain why the fly M1 has short wings, and that we compare M1 to the rest of the fly population. In such a case the explanandum may be broken down into a number of unique differences, ⟨*M1, E, N1*⟩, ⟨*M1, E, M2*⟩, ⟨*M1, E, N2*⟩ . . . It is plausible to say that the condition which explains the greatest number of such unique differences is the one that has the greatest explanatory power and should be chosen as the "most important" cause of the short wings. It has been shown (Hesslow, 1983) that under fairly reasonable assumptions, that condition which has the greatest explanatory power, in this sense, will also be the one with the greatest predictive power, i.e. it will be the one that most raises the probability of the effect.

So far, little has been said about how we choose the objects of comparison or the reference class. This is mainly because the reference class is often an unconscious entity, which is formed by a variety of logically irrelevant factors of which experience, norms and educational indoctrination are examples. We will often compare the explanandum object with what we perceive as normal, but we may also be biased by our education towards using a certain theoretical ideal as a reference class. An economist, when explaining unemployment, might compare the actual situation to a hypothetical one where perfect market equilibrium obtains,

whereas a politician is more likely to compare the situation today with that during the previous government.

Notice, however, that this does not make casual explanation subjective. It is not claimed that subjective factors make people answer the same questions differently, but rather that subjective factors *make people ask different questions* (see also Turnbull and Slugoski, this volume). Furthermore, nothing compels us to make a certain comparison or to ask a certain question. It may be true that our intellectual habits and expectancies tend to make us ask certain questions and not others, but we can and do understand questions of an unusual kind when they are explained to us. The important point is that different objects of comparison correspond to different explananda and give rise to and arise from different questions.

For these reasons it is misleading to construe the main problem we have been dealing with in this paper as a problem of selecting causes from a set of objectively equal conditions. What are being selected are essentially questions, and the causal selection that follows from this is determined by the straightforward criterion of explanatory relevance.

REFERENCES

Carnap, R. (1966). *The Philosophical Foundations of Physics.* New York: Basic Books.

Collingwood, R.G. (1940). *An Essay on Metaphysics.* Oxford: Oxford University Press.

Dietl, P.J. (1970). "Abnormalism". *Theoria, 36,* 93–9.

Dray, W. (1957). *Laws and Explanations in History.* Oxford: Oxford University Press.

Dray, W. (1964). *Philosophy of History.* Englewood Cliffs, N.J.: Prentice-Hall.

Fogel, R. (1964). *Railroads and American Economic Growth: Essays in Econometric History.* Baltimore: Johns Hopkins University Press.

Frey, R.G. (1976). "Judgements of causal importance in the social sciences". *Philosophy of the Social Sciences, 6,* 245–8.

Frey, R.G. (1978a). "Contributory causation and the objectivity of the social sciences". *Philosophy of the Social Sciences, 8,* 175–9.

Frey, R.G. (1978b). "Contributory causation and objectivity: a final instalment'. *Philosophy of the Social Sciences, 8,* 182–3.

Gardiner, P. (1952). *The Nature of Historical Explanation.* Oxford: Oxford University Press.
Gärdenfors, P. (1980). "A Pragmatic Theory of Explanations". *Philosophy of Science, 47,* 404–23.
Gerschenkron, A. (1968). *Continuity in History and other Essays.* Cambridge, Mass: Belknap Press.
Hart, H.L.A. and Honoré, A.M. (1959). *Causation in the Law.* Oxford: Oxford University Press.
Hempel, C.G. (1965). *Aspects of Scientific Explanation.* New York: The Free Press.
Hesslow, G. (1981a). "The transitivity of causation". *Analysis 41,* 130–3.
Hesslow, G. (1981b). "Causality and determinism". *Philosophy of Science, 48,* 591–605.
Hesslow, G. (1983). "Explaining differences and weighting causes", *Theoria, 49,* 87–111.
Hesslow, G. (1984). "What is a genetic disease? On the relative importance of causes". In L. Nordenfelt and B.I.B. Lindahl (Eds) *Health, Disease and Causal Explanations in Medicine,* Dordrecht: Reidel.
Hume, D. (1739). *A Treatise of Human Nature.*
Kelsen, H. (1943). *Society and Nature,* Chicago: University of Chicago Press.
Lewis, D. (1973). "Causation". *Journal of Philosophy, 70,* 556–67.
MacIver, R.M. (1952). *Social Causation,* Boston: Ginn & Co.
Mackie, J.L. (1965). "Causes and conditions". *American Philosophical Quarterly, 2:* 245–64.
Mackie, J.L. (1974). *The Cement of the Universe,* Oxford: Oxford University Press.
Martin, M. (1974). "Causal importance and objectivity". *Philosophy of the Social Sciences, 4,* 157–68.
Martin, M. (1978a). "Judgements of contributory causes and objectivity". *Philosophy of the Social Sciences, 8,* 173–75.
Martin, M. (1978b). "Contributory causes again". *Philosophy of the Social Sciences, 8,* 180–1.
Martin, R. (1972). "On weighting causes". *American Philosophical Quarterly, 9,* 291–9.
Mill, J.S. (1973). "System of Logic" (8th edn). In J.M. Robson (Ed.) *Collected Works of John Stuart Mill* (Vols. VII and VIII). Toronto: University of Toronto Press. (Original work published in 1872).
Nagel, E. (1961). *The Structure of Science.* New York: Harcourt, Brace and World.
Toulmin, S. (1961). *Foresight and Understanding.* London: Hutchinson.
van Fraassen, B.C. (1980). *The Scientific Image.* Oxford: Oxford University Press.

von Wright, G.H. (1971). *Explanation and Understanding*. London: Routledge & Kegan Paul.