ELLIOTT SOBER

EQUILIBRIUM EXPLANATION*

(Received 14 April, 1982)

Familiar counterexamples to Hempel's DN model of explanation (e.g., Bromberger's, 1966) strongly suggest that the explanation of a particular occurrence must cite its cause. When a building casts a shadow, the sun's position and the shadow's length do not explain the building's height because, it would seem, they do not cause it. This plausible diagnosis has been made a matter of principle in Salmon's (1971, 1975, 1978) SR model of explanation, in spite of the puzzles posed by so-called laws of coexistence; if the length of a pendulum at a given time explains its period at that time (as Hempel, 1965, believes), then the causal requirement is inappropriate. But intuitions and issues are divided or unclear in this class of cases, so the causal condition remains tenable, if not unproblematic.

Equilibrium explanations, as I will call them, present a distinct set of counterexamples for the causal requirement. Or, at the very least, they suggest that more attention needs to be paid to specifying exactly what it is for an explanation to be causal. Additionally, equilibrium explanations have an interesting bearing on the role of the statistical relevance idea in the theory of explanation, in that equilibrium explanations show how the cause of an event can be (statistically) *ir*relevant to its explanation.

R. A. Fisher (1931) formulated an equilibrium explanation of the fact that the sex ratio at reproductive age is 1:1 in many species. The main idea of his characteristically terse formulation (see Hamilton, 1968; Crow and Kimura, 1970; or Maynard Smith, 1979 for elaboration and discussion) is that if a population ever departs from equal numbers of males and females, there will be a reproductive advantage favoring parental pairs that overproduce the minority sex. A 1:1 ratio will be the resulting equilibrium point. The ratio of male to female progeny has an impact on a parent's fitness in virtue of the number of *grand*children that are produced. If males are now in the majority, an individual who produces all female offspring will on average

Philosophical Studies 43 (1983) 201–210. 0031–8116/83/0432–0201$01.00

have more grandchildren than one that produces all males or a mixture of sons and daughters.[1]

A causal explanation of the observed 1:1 ratio in a population at a given time would presumably describe some earlier state of the population and the evolutionary forces that moved the population to its present configuration. But Fisher's explanation describes no such thing. In fact, Fisher's account shows why the *actual* initial conditions and the *actual* selective forces don't matter; whatever the actual initial sex ratio *had been*, the selection pressures that *would have* resulted would have moved the population to its equilibrium state. Where causal explanation shows how the event to be explained was in fact produced, equilibrium explanation shows how the event would have occurred regardless of which of a variety of causal scenarios actually transpired.

Fisher's argument does cite a number of contingent facts about any population to which it applies. It asserts that if any force appreciably influences the sex ratio it will be selection for individuals (or more accurately, for parental pairs) that overproduce the minority sex.[2] Furthermore, the required heritable variation in reproductive strategy is simply assumed to be available. The explanation also requires that the population not drift into either of the two absorbing states of all males or all females, since extinction would then follow. Can any of these facts about the population be naturally construed as the cause of its 1:1 sex ratio? Is Fisher's explanation a causal one because it asserts or presupposes such facts? To answer these questions, we must clarify the concept of *causal explanation*.

One way to trivialize the idea of causal explanation is to focus on the inevitable deployment of *ceteris paribus* clauses that must figure in any explanation. A necessary condition for the occurrence of an event E is the nonoccurrence of events that would prevent E from happening. Any explanatory story will presuppose that such preventors failed to intervene. A hollow victory can be won for the causal requirement by focusing on this *ceteris paribus* presupposition and claiming that it constitutes a cause of the event to be explained. This vindication is nearly vacuous in that it ignores what the explanation actually asserts. Constructing causal explanations is not so easy a task. But on this accounting, as soon as one knows that an event occurred one knows what caused it.

A second trivialization of the idea of 'causal explanation' arises from demanding of a causal explanation only that it provide 'information about' the *explanandum's* causal history. On this reading, Fisher's argument fills the

bill, but then so do apparent counterexamples to the DN model. The pseudo-explanation mentioned before of the building's height does tell us something about the cause – namely that it produced a building that allowed the sun to cast the length shadow it did. And if one wishes to argue that laws of coexistence do not provide causal explanation (as Salmon, 1978, does) then, for the same reason, the concept of causal explanation requires a stronger construal.[3]

Can the idea of causal explanation be made more substantial? Grounds for pessimism can be found in the following consideration. A causal explanation describes what the cause is. But 'describing what the cause is' is a species of reporting in which success or failure is context relative. When Holmes asks Watson what caused the death of the murder victim they are examining, Watson provides information about the cause when he says "he was murdered". But Holmes' withering stare quickly indicates that Watson's answer is not wholly satisfactory. It is one thing to answer the *verbal question*, another to answer the *question in mind*. Providing a causal explanation – saying 'what the cause is' – must do more than give a true answer to the verbal question. When a question is asked, there typically will be expectations as to the level of information that any alternative answer must convey. This suggests that the proper unit of analysis is not the idea of causal explanation *tout court*, but the idea of causal explanation relative to some set of alternative causes. If we have in mind some set of alternative causes that we think might have produced the occurrence requiring explanation, then a causal explanation (relative to those alternatives) will tell us which of the alternatives actually produced the outcome.

This analysis may be thought to open the door to an all-permissive relativism. Won't it turn out that any 'information about the cause' will count as providing a causal explanation, since one can always cook up a background context relative to which such information, no matter how impoverished, uniquely determines one of the alternative answers? The prospect of such philosophical inventiveness should not trouble us. Different sciences typically have taxonomies of what can count as a possible cause and if we are thinking about causal explanations in science, we can simply let the sciences decide what does and does not count as a specification of 'the cause'.

Evolutionary theory, and particularly population genetics, which is the science within which Fisher elaborated his equilibrium explanation of the sex ratio, describes the possible causes of evolution. Natural selection, mutation,

migration, random drift and properties of population structure are described, singly and in conjunction, in terms of their impact on gene and genotype frequencies. A causal explanation in this science will explain the present gene or genotype frequencies found in a population by specifying the earlier frequencies and the configuration of evolutionary forces that acted on the population. Fischer's argument is thereby not a causal explanation in population genetics. One might generalize on this and conjecture that an equilibrium explanation in a science is never a causal explanation in that science.

Hamilton (1968) made explicit some of the empirical presuppositions of Fisher's argument. He showed that Fisher's argument requires the following, not always plausible, biological assumptions: (i) that there be population-wide rather than local competition for mates; (ii) that genes controlling the sex ratio not be located on the sex chromosomes of the heterogametic sex; (iii) that in haplodiploidy, the males must fertilize all the females. If we revise Fisher's argument in the light of these findings, do we find ourselves giving *causal* explanations of 1:1 sex ratios for populations to which it applies? Fisher specified how the fitness of a parental pair producing a given frequency of male and female offspring is a function of the sex ratio found in the population. Hamilton made clear what some of the necessary conditions are for this fitness function. But Hamilton's result was a law of coexistence: a population with the required structure *simultaneously* instantiates the Fisherian relationship. When this fitness function obtains, the population will eventually attain a 1:1 sex ratio, if it doesn't already have one. The explanation leaves open the possibility that the population be at its equilibrium value *forever*, in which case the conditions articulated by Fisher and Hamilton can hardly be cited as the cause of the 1:1 sex ratio. Hamilton's argument, just as much as Fisher's, does not say what actual changes a population will undergo in attaining a 1:1 ratio.

Equilibrium explanations present *disjunctions* of possible causal scenarios; the actual cause is given by one of the disjuncts, but the explanation doesn't say which. In Fisher's sex ratio argument there are three disjuncts. If a population is at its equilibrium 1:1 ratio at a given time, the argument says that at an earlier time one of three processes began: (i) at the earlier time, there was an excess of males, a parental pair produced a female biased group of offspring, and selection for a female biased ratio brought the population to 1:1; (ii) at the earlier time, the sex ratio was 1:1, and if there were any parental pairs that deviated from this ratio among their progeny, they would

be selected against; (iii) at the earlier time, there was an excess of females, a parental pair produced a male biased group of offspring, and selection for a male biased ratio brought the population to 1:1. Notice that each of these disjuncts itself covers a range of alternatives; for example, (ii) includes the possibility that the population remained at 1:1 because each parental pair produced this ratio and so there was no selection at all.

A disjunction of causal scenarios clearly provides 'information about the cause' (namely, that the actual cause was one of the several mentioned); but disjunctions of causal scenarios will sometimes fail to say what the cause is. An ice cube melts in a warm room. We might say that the cube's being made of water caused it to melt. Suppose that there were another substance, call it X, that has the same melting point; if the cube had been made of X, it would have melted just the same. In what sense is it true that the cube's being made of water or of X caused it to melt? Only in the sense that its being made of water caused it to melt, or its being made of X caused it to melt. The disjunction of possible causes fails to say what the cause is.

This example suggests that causality abhors an ineliminable disjunction:

(DIS) If ϕ or ψ causes something, then ϕ causes it or ψ does.

But there is room to doubt that this distributive principle is always true. Other issues are involved. When the ice cube melts, its being shaped like a cube was not part of the cause. So its being made of water and shaped like a cube didn't cause it to melt. Conjunction addition is an incorrect principle:

(C) If an object's being ϕ caused A, and if the object's being ψ did not cause A, then its being ϕ and ψ did not cause A.

But now consider the fact that, as a matter of logic, something is made of water if and only if it is made of water and is shaped like a cube, or is made of water and is not shaped like a cube. If substitution of logical equivalents in the predicate position of a causal context is legitimate, then we have the following result. Since the object's being made of water caused the object to melt, it will also be true that the object's being made of water and shaped like a cube, or being made of water and not shaped like a cube, was the cause. If an object's being ϕ is causally efficacious, its being (ϕ and ψ) or (ϕ and $\overline{\psi}$) is too. But neither of the disjuncts is itself causally efficacous in our example. The first disjunct is causally inert, given principle (C); and the second was not the cause, since it, in fact, is false. So if (C) is true and if substitution of logical equivalents is legitimate, (DIS) is false.[4]

Whether or not causality obeys this distributive law, it still seems true that some disjunctive predicates succeed in 'saying what the cause is'. When a tree falls and crushes a bush, we might say that it was the bush's being under the tree that caused it to get crushed. This counts as saying what the cause is, even though being under the tree is a disjunctive state: being under the tree means being *somewhere* under the tree. Perhaps this causal claim is true only because it is unwritten by the truth of one of its disjuncts; maybe the bush's having the precise location it had caused it to get crushed.[5] Either way, I think one is compelled to grant that you are saying what the cause is when you say that the bush was under the tree. However, matters change when we revert to our water or X case. If you say that the object melted because it was made of water or of X, you are not saying what the cause is. Your claim will be true only because either the cube's being made of water was the cause, or its being made of X was; but you didn't say which.

So the situation is complex. There is first of all the question about the reality (irreducibility) of disjunctive causes (i.e., the question of whether DIS is true). But a separate question concerns what constitutes success and failure in saying what the cause is. Even if causation abhors an ineliminable disjunction, a disjunctive predicate (like 'being under the tree') can succeed in picking out the cause. But there are cases in which a disjunction fails to do this (as in the 'water or X' case).

Perhaps the difference here reflects a contrast between uniqueness versus diversity of *processes*. The position of the bush would have led to its being crushed by the same process, regardless of exactly where it was under the tree. But if the ice cube had been made of X rather than of water, a different process would have led it to melt. Nature's abhorence of disjunction is not syntactic; there is a real difference, which may or may not be reflected in surface structure, between 'being under the tree' and 'being made of water or of X'. This speculation presents the obvious question of seeing what progress can be made in the project of individuating processes.[6]

It is a further question to connect these observations about causation with the issue of explanation. If you say that the cube was made of water or of X, did you explain why the cube melted? You didn't say what the cause of the melting was. It would be an easy (and hollow) victory for my position to call this an explanation, and conclude that not all explanation is causal explanation. But my intuition is that in this case, explanation and causation go hand in hand: your remark fails to explain precisely because and in so far as you've failed to pinpoint the cause.

Equilibrium explanations are different. Compare a *non*equilibrium of why the sex ratio in some population is 1:1. Suppose we record the reproductive output of males and females from some earlier time and show how these numbers brought the population from 73% females to 1:1. This scenario, though it pinpoints the cause, is a lot less explanatory than the equilibrium story. The causal explanation focuses exclusively on the actual trajectory of the population; the equilibrium explanation situates that actual trajectory (whatever it may have been) in a more encompassing structure. It is in this way that equilibrium explanations can be more explanatory than causal explanations even though they provide less information about what the actual cause was. This difference arises from the fact that explanations provide *understanding*, and understanding can be enhanced without providing more details about what the cause was.

Equilibrium explanations are made possible by theories that describe the dynamics of systems in certain ways. The objects in the domain can satisfy a range of values of a certain parameter. In population genetics, the parameter is gene (and genotype) frequencies, which take values between 0 and 1 inclusive. When an object is subject to certain forces, there will be three sorts of states that may be located at different values of the parameter. The three types of states we want to define are absorbing states, stable equilibria, and unstable equilibria.. It is a delicate matter to characterize these, so perhaps an example from population genetics will be useful. Consider the fitness function shown in Figure 1 of two traits A and B in a population. The function specifies the fitness (in terms of expected number of offspring) that each type has, given its frequency in the population. Notice that each of these traits is favored when it is in the minority.[7]

As the diagram suggests, point E is an equilibrium value. Here, the fitnesses are identical, and so natural selection ceases. If the population is anywhere else (except for 100% A and 100% B), selection favoring the minority trait will take the population to its equilibrium value. E is a *stable* equilibrium point, since if the system is in the neighborhood of E, the relevant force will move the system to E. The frequencies of 0% and 100%, however, are called absorbing points. The reason is not simply that once the system is in either of these states, selection can't budge it (the model doesn't include the possibility of mutation); if this were the only reason, there would be no difference between absorbing points and stable equilibria. Nor does the difference seem to be that at equilibrium there is no variation in fitness,

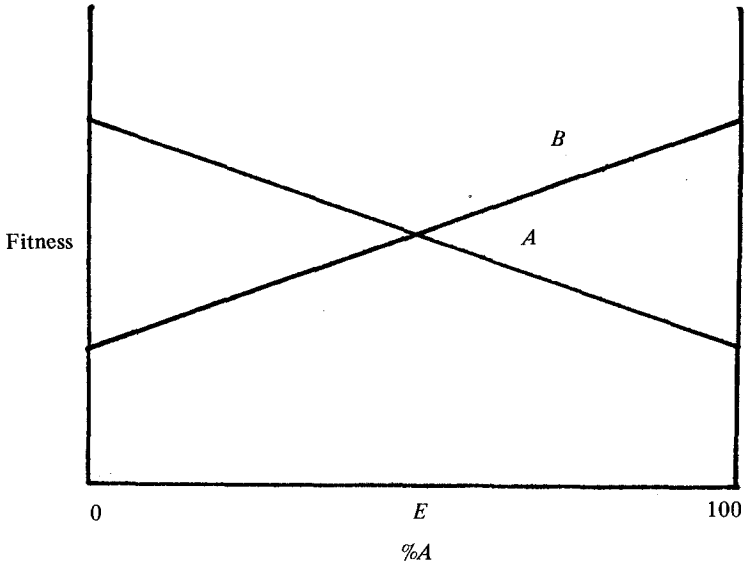

Fig. 1.

whereas at the absorbing states of 100% and 0% there still is. There can't be variation in fitness where there is no variation, and selection doesn't exist without there being variation in fitness (for it, unlike some physical forces, *esse est percipe*). Rather, the difference seems to involve the notion of chance. Chance (i.e., random drift) can move a population out of its equilibrium state; if the equilibrium is a stable one, the deterministic forces involved (i.e., selection in this case) will return the population to equilibrium. If the equilibrium is unstable, the population will be carried further away from the equilibrium by the forces involved. (Switch the fitness functions of A and B to obtain an unstable equilibrium at E.) But the point is that at the absorbing states, chance can't bring about a deviation; it takes another deterministic force to do so (in this case, mutation).

Whether a given parameter value is a stable equilibrium depends on what happens in its neighborhood. Because of this, a stable equilibrium can be either local or global. A global equilibrium is one that the system approaches, regardless of its initial state. A local equilibrium is one that the system will

move towards, if its initial conditions fall within the required local neighborhood. The fitness function given in the above diagram implies that point (E) is a global stable equilibrium.

The distinction between local and global equilibria can be supplemented, in an obvious way, by a quantitative measure of *how* local a given equilibrium is. This may simply be identified with the range of initial parameter values all sharing the same equilibrium state. The more local an equilibrium is, the more information about the initial conditions will be required in explaining its equilibrium state. As the equilibrium becomes more and more local, there will be less and less difference between what a causal explanation and an equilibrium explanation need to say about the initial conditions. But when we are at one end of the continuum – when the equilibrium is a global one – an event can be explained in the face of considerable ignorance of the actual forces and initial conditions that in fact caused the system to be in its equilibrium state. In this circumstance, we are, in one natural sense, ignorant of the event's cause, but explanation is possible nonetheless.[8]

University of Wisconsin-Madison

NOTES

* This work was supported by grants from the John Simon Guggenheim Foundation and from the University of Wisconsin Graduate School, which I acknowledge with thanks. I also am grateful to the Museum of Comparative Zoology, Harvard University, for its hospitality during 1980–81.

[1] In point of fact, Fisher argued that it was not the number of females and males that would be equal, but rather the amount of parental expenditure on males and females. This way of formulating Fisher's argument would not affect the points to be made above, which can be understood as implicitly assuming that the costs of daughters and sons are equal.

[2] A simple genetical model making more explicit Fisher's idea might naturally be interpreted as implying *genic* selection. See Crow and Kimura (1970, pp. 288ff.) for such a model, and Sober and Lewontin (1982) for discussion of the difference between genic and organismic selection.

[3] Similar comments apply to Salmon's remark (1978, p. 421) that "such laws as conservation of energy and momentum are causal laws in the sense that they are regularities exhibited by causal processes and interactions".

[4] In Sober (1982), I argue that there are counterexamples to this substitution principle. Note that Quantum Mechanics may provide a class of counterexamples to (DIS).

[5] I am inclined to say that the bush's exact location caused it to get crushed. I thereby deny that causes need be necessary for their effects. For the contrary position see, for example, Lewis (1973).

[6] Salmon (1978) argues that the concept of a physical process is crucial for understanding causality.

[7] This fitness function has the mathematical form of a constant viability model of heterozygote superiority. See Sober and Lewontin (1982) for details.

[8] In a way, equilibrium explanations are the mirror images of what David Hull (1975) has called historical narrative explanations. The items explained by both these methods have their causes and also fall under natural laws. The difference between the kinds of explanation, however, lies in what information is cited and does the explanatory work. Equilibrium explanations do not say what the cause is (though such a thing doubtless exists); historical narrative explanations do not say what the law is (though presumably one exists).

BIBLIOGRAPHY

[1] Bromberger, S.: 1966, 'Why-questions', in R. Colodny (ed.), Mind and Cosmos University Series in the Philosophy of Science, Vol. III (University of Pittsburgh Press); reprinted in B. Brody (ed.), 1970, Readings in the Philosophy of Science (Prentice-Hall, Englewood Cliffs).

[2] Crow, J. and Kimura, M.: 1970, Introduction to Population Genetics Theory (Burgess, Minneapolis).

[3] Fisher, R.: 1931, The Genetical Theory of Natural Selection (Dover, New York, 1958).

[4] Hamilton, W.: 1968, 'Extraordinary sex ratios', Science 156, pp. 477–488.

[5] Hempel, C.: 1965, 'Aspects of scientific explanation' in Aspects of Scientific Explanation and Other Essays in the Philosophy of Science (Free Press, New York).

[6] Hull, D.: 1975, 'Central subjects and historical narratives', History and Theory 14, pp. 253–274.

[7] Lewis, D.: 1973, 'Causation', Journal of Philosophy 70, pp. 556–567.

[8] Maynard Smith, John: 1979, The Evolution of Sex (Cambridge University Press, Cambridge).

[9] Salmon, W.: 1971, Statistical Explanation and Statistical Relevance (University of Pittsburgh Press, Pittsburgh).

[10] Salmon, W.: 1975, 'Theoretical explanation', S. Korner (ed.), Explanation (Basil Blackwell, Oxford).

[11] Salmon, W.: 1978, 'Why ask "why?" An inquiry concerning scientific explanation', in W. Salmon (ed.), Hans Reichenbach: Logical Empiricist (Reidel, Dordrecht).

[12] Sober, E.: 1982, 'Why logically equivalent predicates may pick out different properties', American Philosophical Quarterly 19, pp. 183–189.

[13] Sober, E. and Lewontin, R.: 1982, 'Artifact, cause, and genic selection', Philosophy of Science 19, pp. 157–180.