

JAEGWON KIM

CAUSES AS EXPLANATIONS: A CRITIQUE*

Causes certainly are connected with effects; but this is because our theories connect them, not because the world is held together by cosmic glue.¹

Norwood Russell Hanson

ABSTRACT. This paper offers a critique of the view that causation can be analyzed in terms of explanation. In particular, the following points are argued: (1) a genuine explanatory analysis of causation must make use of a fully epistemological-psychological notion of explanation; (2) it is unlikely that the relatively clear-cut structure of the causal relation can be captured by the relatively unstructured relation of explanation; (3) the explanatory relation does not always parallel the direction of causation; (4) certain difficulties arise for any attempt to construct a nonrelativistic relation of causation from the essentially relativistic relation of explanation; and (5) to analyze causation as explanation is to embrace a form of "causal idealism", the view that causal connections are not among the objective features of the world. The paper closes with a brief discussion of the contrast between the two fundamentally opposed viewpoints about causality, namely causal idealism and causal realism.

It is little more than a truism to say that causes explain their effects, or that to ask for the cause of an event is to ask for an explanation of why or how the event occurred. This close association between causation and explanation is amply mirrored in language: we answer 'why'-questions with 'because'-statements, and surely there is more than an orthographic resemblance between 'cause' and 'because'. The association is also ancient: it goes back to Aristotle, who characterized true, scientific knowledge as knowledge of the 'why' of things, that is, knowledge of the cause that makes a thing what it is and not something else.² More recently Hanson wrote:³

The primary reason for referring to the cause of x is to explain x . There are as many causes of x as there are explanations of x .

Thus the connection between causes and explanations seems more than a

contingent matter; it strikes us as being rooted in the very concepts of cause and explanation. It becomes attractive, therefore, to attempt an explicit *analysis* of causation as explanation. If there is a single characteristic that binds together the diverse range of things we commonly identify as 'causes', it seems to be that they explain their effects, that is, they can serve in an account of the existence and properties of whatever it is they are said to be causes of.

It is somewhat surprising, therefore, that, apart from some scattered exceptions to be mentioned below, no sustained effort has been made to formulate a systematic and detailed analysis of causal concepts in terms of explanatory concepts. Nor has there been a systematic exploration of the precise nature of the connection between causes and explanations. That there are some intimate connections is often simply assumed or asserted; but these connections have never been catalogued or elucidated. The present paper has the modest aim of examining some issues involved in the program of analyzing causation as explanation. More specifically, I intend to examine and comment on a recent explanatory analysis of causation, and then to develop some larger issues and problems that seem to arise for any genuine explanatory analysis of causal relations. If I am right in my major conclusions, the project of analyzing causes as explanations not only runs into a variety of difficulties in its execution, but it also forces upon us a choice between two fundamentally opposed viewpoints with regard to the place of the causal relation in the world and the role it plays in our knowledge of the world, that is, between 'causal realism' and 'causal idealism'. I shall claim that to view causation merely as explanation is to embrace a form of causal idealism, a viewpoint I believe we should reject.

I

It will be useful to focus our discussion on some recent proposals made by Scriven toward an analysis of causation as explanation. These proposals are not always easy to interpret consistently; this is in part because they are bound up with Scriven's attempts to refute the traditional regularity account of causality, and in part because Scriven at the same time holds the view that causation *cannot* be analyzed. All the same Scriven's account of causation as explanation can serve as a framework within which we could develop a general critique of the explanatory approach to causality. Scriven writes:⁴

Now suppose we assert that requests for the cause of something are like requests for directions or guidance (What should I do now?). More specifically, they are requests for *single-factor explanations* (of a certain kind).

The correlative grammar of other causal concepts would then go something like this. A cause is an explanatory factor (of a particular kind). Causation is the relation between explanatory factors (of this kind) and what they explain.

Scriven goes to explain 'explanation':⁵

We can amplify the concept of explanation on which this depends by saying, crudely speaking, that (i) explanations are symbolic vehicles for conveying understanding; (ii) understanding is acquired whenever the capacity for solving a certain appropriate range of problems is achieved without learning the solutions for each problem separately; (iii) understanding may thus be acquired by all sorts of input, scientific or not, depending simply on whether they produce this result; and (iv) the determination of the "appropriate range of problems" is a contextual requirement.

I believe Scriven will be the first to admit that the proposals he makes here are only a skeleton of a theory, and that detailed work remains to be done to convert this skeleton into a full-fledged analysis. However, exactly how the details are worked out is not our present concern. Our purpose is not to produce specific counterexamples to a particular explanatory account of causation, but rather to evaluate the general approach embodied in Scriven's discussion. For this reason, it is best not to worry about the qualification 'of a certain kind' or 'of a particular kind' in Scriven's thesis that a cause is an explanatory factor of a *certain kind*. Ultimately one would need to ask whether the kind in question can be characterized in a noncircular way, but at this point our attention should be directed to the central core of the thesis that *causes are explanations*.

What is of crucial importance is Scriven's account, sketchy as it is, of explanation in terms of 'understanding'. This is important because the mere assertion that causes are explanations does not commit one to any specific theory of causation unless the concept of explanation involved is made clear. It is not required that a full and complete theory of explanation be given; what we need is, at a minimum, an indication of the direction in which one wishes to go in explaining 'explanation'. The reason for this is simple: if by 'explanation' one has in mind something like Hempelian deductive-nomological (D-N) arguments, then the explanatory analysis of causation would quickly collapse to the standard nomological or regularity analysis. Thus, Hempel's own remarks about causes, which tie singular causal

statements to D-N explanations, cannot be considered as forming even the beginning of a genuine explanatory theory of causation;⁶ in fact, Hempel may be viewed as having attempted a reduction of both explanation and causation to nomological connections between events, and on his essentially Humean regularity view of lawlikeness, nomological connections themselves reduce to the Humean relation of 'constant conjunction'.

As is evident in the second of the two passages quoted from Scriven above, the D-N model is the furthest thing in Scriven's mind when he talks about explanation. This passage has the virtue of bringing out clearly his commitment to an explicitly psychological and epistemological view of explanation, and this is what makes his account interesting: it is what makes the account a genuine explanatory theory of causation. Any genuine account of causation as explanation should make use of the full-blooded notion of explanation with all its epistemological and psychological implications and associations. I do not claim that explanation is merely a psychological affair, but it is difficult to see how a full account of explanation could avoid reference to such epistemological and quasi-psychological notions as 'understanding', 'making intelligible', 'puzzle', 'knowledge', etc. It is likely that there is always some 'cognitive' or 'theoretical content' in any explanation, and this content may just be a D-N argument. But, then, as we have seen, an account of causation as explanation in this sense of 'cognitive content' may simply turn out to be the standard Humean account. Although this is in a sense a terminological point, it is important nonetheless; for our aim is to discuss genuine explanatory theories of causality, not some disguised regularity account; our aim is to bring out the strengths and weaknesses of the explanatory analysis insofar as it is a distinctive approach to causality in its own right.

The point being emphasized here can be appreciated when we look at the ambiguous nature of the account of singular causal statements offered by Morton White, who advocates the following thesis:⁷

A statement of the form '*A* is a contributory cause of *C*' is true if and only if there is an explanatory deductive argument containing '*A*' as a premise and '*C*' as its conclusion.

The critical expression in this formulation is 'explanatory deductive argument'. According to White, who is basically sympathetic with the Hempelian nomological account of explanation, there is an important difference between

a *mere* law and an *explanatory* law. The motivation for this distinction is twofold: first, the apparent fact, noted by White, that some perfectly legitimate D-N arguments fail to be explanatory, and second, White's belief that this is due to the use of a nonexplanatory law, a 'mere law', in the explanans of the argument. An example of such a nonexplanatory law offered by White is "Whenever there is lightning, there is thunder very soon afterwards".⁸ It isn't entirely clear whether this particular statement is a law or whether no D-N argument making use of this statement is explanatory, but this is unimportant. The point to be noted is that in the absence of further explanation of the concept of 'explanatory law' it is not possible to categorize White's analysis either as an explanatory or as a nonexplanatory account. White himself says little by way of explicating this concept, but there are a few things he says that are of interest. First, he thinks that whether a law is explanatory is not a matter of 'judgment'. Being a matter of judgment, I would suppose, is somewhat like being 'subjective', a situation in which no objective criteria exist as to being correct or incorrect about the matter involved. Second, White says that the concept of an explanatory law could be 'relative in character', that is, it is a covert relation that is fully expressible by the locution 'explanatory law for person or culture X'.⁹ Thus, the concept of explanation White has in mind could be relative though not subjective. (We shall recur to this matter of relativity later.) In any event it is clear that pending further clarification of the notion of 'explanatory law', and hence of the notion of 'explanation', White's theory of causal statements cannot be definitively classified.

Hanson, too, has closely associated causation with explanation. While Hanson does not offer a systematically developed and articulated account of explanation, his conception of explanation evidently goes much beyond the standard D-N account in taking the epistemological and conceptual aspects of our explanatory endeavors as central to successful scientific theorizing. Hanson's views do not lend themselves to a short encapsulation, but some of the highlights can be mentioned.¹⁰ According to Hanson, cause-words, such as 'wound' (as contrasted with 'scar'), explain because they are 'theory-laden'; they derive their explanatory power from the interlocking patterns of concepts in which they enter as part of a scientific theory, and a theory is explanatory because it provides a multi-layered system of concepts, the layers corresponding to the different degrees of explanatory power

attaching to these concepts. And a theory, for Hanson, is not a mere abstract calculus that is given a 'partial interpretation' in observation by 'correspondence rules'. The foremost goal of a scientific theory is to *explain observed phenomena* and Hanson would say that the positivist conception of theories as partially interpreted formal systems is incapable of accounting for this central aspect of theories. Hanson emphasizes *conceptual* congruence between cause and effect (no mere Humean constant conjunctions will provide such connections), and also the context-dependency of explanation. Regardless of the question of detail, it should be clear that Hanson's explanatory account of causation does not reduce to any sort of regularity account; Hanson takes the epistemological aspects of explanation as essential, and it is the use of these epistemological factors that makes an approach to causation a genuine explanatory approach.

Let us now turn to some details of Scriven's proposals. This will help us uncover some fundamental problems that any explanatory theory of causation would have to face. According to Scriven, *A* is a cause of *B* just in case *A* is an explanatory factor (of a certain kind) for *B*. Scriven amplifies this by adding that an explanation is a conventional vehicle of communicating 'understanding', and that understanding is acquired when the capacity for solving a class of problems is obtained. Throughout his discussion Scriven repeatedly stresses the context dependency of the notion of understanding and the marginal role played, if any at all, by the 'objective correlative' of a causal statement. What, according to Scriven, are the advantages to be gained from this explanatory analysis of causation?

(1) Scriven claims that this analysis explains why causation is irreflexive and transitive:

It is obviously useless to provide information that the inquirer already has. One does not come to understand a phenomenon described by oneself as *X* by being told that it exists or that it is called *X*. . .

If the phenomenon-to-be-explained (*X*) is explained by reference to another phenomenon, *P*, and that phenomenon is explained by reference to a phenomenon *A*, then of course *X* has to be explained in terms of *A* (hence causation is, generally speaking, transitive).¹¹

Few will question that causal relations are irreflexive and transitive; but are explanatory relations also irreflexive and intransitive? Take irreflexivity first. As Scriven says, it would be pointless to reply '*P*' when asked "Why is it

the case that *P*?”. But this isn’t enough: to conclude that explanation is irreflexive one needs to show that we cannot explain *X* by reference to *X* (note Scriven’s wording in the quotation above), and as long as the same *X* can be described in nonequivalent ways, self-explanation need not take the form ‘*X* is so because *X* is so’. In this connection, William Dray’s notion of *explanation by redescription*¹² readily comes to mind as a possible legitimate way of explaining something by reference to itself, i.e., by redescribing it. As examples of such a mode of explanation the following examples come to mind: we redescribe the scratchy lines on a photographic plate exposed in a Wilson cloud chamber as ‘tracks of such-and-such particles colliding with such-and-such particles’; we redescribe a peasant unrest as ‘the beginning of a revolution’. Now Scriven himself seems to leave open this sort of possibility when he talks of “the phenomenon to be explained and the necessarily different (*at least in description*) phenomena that provide the explanation”,¹³ and in comparison with, say, the D-N model, his broad and relatively unstructured conception of explanation should be particularly congenial to the possibility of explanation by redescription. In any case, the very possibility of this sort of explanation throws doubt on Scriven’s claim that explanation, on his account, is irreflexive, and that hence so is causation.

(2) As for the alleged transitivity of explanation similar difficulties arise. To be sure, a nonepistemic conception of explanation, such as the D-N account, may very well make explanation a transitive relation; even under an epistemic conception, the *logical* relation holding between the explanandum and the explanans (‘the theoretical substance of the explanation’) may very well be transitive. But it is by no means an obvious or trivial matter that explanation understood in its full-fledged epistemic and psychological sense is transitive; in fact, it is doubtful that the explanatory relation as conceived by Scriven can be transitive. For from the fact (i) that reference to *X* would impart, to a suitably qualified person, understanding of *Y*, or a problem-solving capacity with regard to *Y*, and (ii) that reference to *Y* would do the same for a suitably qualified person with regard to *Z*, it by no means follows logically that reference to *X* would impart to such a person understanding or some problem-solving capacity with regard to *Z*. Perhaps, a ‘suitably qualified person’ could not see the connection between *X* and *Z* *without actually going through the intermediate Y*. The situation here is similar to this: entailment is transitive but knowledge of entailment may very well not be, and probably isn’t.

(3) What the foregoing shows is that once epistemological and psychological elements are brought into explanation, we can no longer count on the explanatory relation to exhibit neat and clear-cut structures such as transitivity and irreflexivity. Since the causal relation appears to be more clearly structured than the explanatory relation – at least, the causal relation must be transitive and irreflexive – there is *prima facie* doubt that the explanatory relation could be made to mirror the structure of the causal relation. When we consider the question of asymmetry, the difficulty becomes evident. For a certain *X* and *Y*, there may be no good reason to think that the explanatory relation could not obtain both ways, from *X* to *Y*, and from *Y* to *X*; this could clearly be the case for different persons, and it seems that we could imagine the situation obtaining for one person. That is, a person may understand *X* by reference to *Y*, and *Y* by reference to *X*; in fact, if *X* and *Y* are lawfully correlated so that each could be manipulated through the manipulation of the other, then under Scriven's account of understanding in terms of the acquisition of a problem solving capacity it seems perfectly possible for *X* and *Y* to explain each other.

Now Scriven himself leaves open such a possibility when he says that “if *X* can be explained in terms of *P*, it's an open question whether the converse is true”.¹⁴ The problem, however, is that the causal relation is asymmetric: if *X* is the cause of *Y*, then *Y* can't be the cause of *X*. It will not do, as Scriven does, blithely to conclude that causation isn't asymmetric, because explanation isn't. Causation is generally thought to be asymmetric, and if a proposed analysis of causation doesn't make is asymmetric, an explanation of why causation isn't, or needn't be, asymmetric, should be provided.

(4) Our discussion so far indicates that while there may be a close connection between causation and explanation, the two may not always run in parallel directions, and this, I believe, is indeed the case. Consider these facts:

(i) A full understanding of an event, or kind of event or process, may involve not only information concerning its causative factors but also information concerning its effects. Consider the problem of attaining a proper understanding of such events and processes as the Industrial Revolution, the gravitational collapse of a star, atomic fissions and fusions, and the early deprivation of a mother or mother-substitute for an infant. It would seem that in any ordinary sense of ‘understanding’, or Scriven's sense of it, a

full and proper understanding of these phenomena must attend to their effects, as well as to their causes.

(ii) One pattern of explanation that has lately received a good deal of attention is the so-called 'teleological explanation' or 'functional analysis', a mode of explanation which some believe is prevalent in the biological and social sciences.¹⁵ There has been much controversy about the precise nature of explanations of this type, but it is safe to say that these explanations are typically, if not always, explanations of the presence or occurrence of a biological or sociological trait, process, or condition in terms of its causal *effects* that are relevant to the occurrence or attainment of a certain 'goal-state', such as the survival or maintenance of the organism or social group. Here causal and explanatory connections criss-cross each other, and if an overall pattern is discernible, it is that a typical teleological explanation of a phenomenon is a causal explanation in terms of the *effects* of the phenomenon, not in terms of its causes.

(iii) One important variety of theoretical explanation in science is micro-explanation of macro-phenomena; theoretical science abounds in such explanations (e.g., explanation of thermal phenomena in terms of kinetic molecular phenomena, molecular explanation of genetic phenomena, etc.). These explanations aren't *noncausal* – at least, they needn't be – but it isn't happy to regard the micro-state as a *cause* of the macro-phenomenon it explains. For one thing it would obscure important differences between this model of reductive explanation and the more familiar causal explanation in which an event is explained by reference to temporally antecedent causal determinants; for another, some philosophers would regard the relation between the micro-processes and their macro-counterparts (e.g., molecular phenomena and thermal phenomena) as that of *identity* ("Heat is nothing but molecular motion"), and to adopt the position that one is the cause of the other is to preclude, without argument, this interesting philosophical interpretation of the relation whereby micro-processes determine the nature of the macro-phenomena they underlie.¹⁶

We must conclude that there is no neat parallel relationship between causation and explanation. Compared with the rather clearly structured relation of causality, the explanatory relation is a free-wheeling affair; it can ride on the causal relation backward as well as forward, and sometimes it does not have a clear directional relationship at all with respect to causality.

If we take explanation in the broad sense of making some phenomenon intelligible, better understood, and so forth, then we should expect that knowledge of the *whole causal network* in which the phenomena is a link will be relevant. The usual preference for the causal origin of whatever it is that is the object of explanation may simply reflect our interest in the control and prediction of that sort of phenomenon. Of course, our concept of understanding may be connected in certain important ways with control and prediction, but this is a point that needs further exploration. Scriven's own account of understanding in terms of problem-solving capacities is not definite enough to rule out understanding in terms of effects as a valid way of achieving understanding.

II

As Scriven and others have emphasized, what works as an explanation for one person, that is, what 'conveys understanding' to one person, may fail as an explanation for another person with different cognitive resources and capacities, and with different queries or puzzles. Indeed, the so-called 'pragmatic' and 'context-dependent' character of explanation has often been stressed in the philosophical literature on explanation, and philosophers like Sylvain Bromberger and Peter Achinstein have done valuable work in delineating the epistemological and others kinds of factors that constitute the context of a given explanatory episode.¹⁷ There is no question that explanation, broadly understood, is a heavily context-dependent affair, and no full account of explanation could ignore this relativity of explanation with respect to contextual factors.

It will be premature to conclude, however, that any explanatory account of causation inevitably renders the causal relation a similarly relativistic relation. If something like "*P* is an explanation of *Q* for the person *A* in the context *C*" (where '*C*' could be a conjunction of contextual factors) is the basic schematic expression for the explanatory relation, the explanatory theory of causation need not take as the basic expression of the causal relation the similarly relativized expression "*P* is a cause of *Q* for the person *A* in the context *C*". Causal relativism of this overt sort is surely intolerable. Now, in order to get a nonrelative causal expression from the relativized explanatory relation one of the following two procedures may be followed:

(1) we may choose some standard person A and standard context C , say A^* and C^* (the expression ' A^* ' and ' C^* ', therefore, are constants, not variables), and explain ' P causes Q ' as ' P explains Q for A^* in C^* '; alternatively, (2) we may existentially quantify away the variables ' A ' and ' C ' and explain ' P causes Q ' as "For some (possible?) person A in some (possible?) context C , P is an explanation of Q ". Scriven himself hints at the first procedure when he says "the explanation is thus whatever *would* impart full understanding to a *suitably qualified* listener/reader".¹⁸ In fact, we may take Scriven's formulation as a hybrid of the two procedures picking a reference audience (a 'suitably qualified' audience) and existentially quantifying over the context C .

It is easy to see, however, that neither of the two approaches yields satisfactory results. On the first approach, various questions arise for the choice of the reference standards. Scriven's notion of 'suitably qualified listener/reader' is itself heavily context-dependent and any choice of a definite reference standard will be subject to the challenge: "Why this standard and not some other?" The possibility of such a challenge is not to be taken lightly, for the choice of reference standards is going to *determine what causal relations are to hold in the world*. It is dubious that *any* choice of reference standards could be justified, unless one is willing to embrace the view that causal relations after all are not objective characteristics of the world. For as long as we take the realist view of causality that causal connections are among the real facts of the world independent of our cognition, any choice of reference audience A and the context C would have to be justified by a demonstration that explanatory connections *just* for *this* audience in *this* sort of context constitute causal connections. It is difficult to see how such a justification could be successfully formulated.

On the second approach mentioned above of existentially quantifying away audience and context, causal connections will multiply and proliferate beyond any reasonable bounds. For, given the variety of possible theoretical and conceptual puzzles, queries and predicaments and the diverse range of cognitive resources and capacities different persons could bring to an explanatory context, the question what facts, events, etc. will explain what other is a wide-open question. As Achinstein observes, "given many E 's and q 's that to us seem quite unrelated, we can at least imagine an A and a situation S such that A would attempt to explain q to those in S by citing E ",¹⁹ and

many of these explanatory attempts might succeed. Indeed, the possibilities can be seen to be truly limitless when we realize that humans are not the only possible cognitive beings, and that there may well be other cognitive beings whose requirements for 'understanding' are entirely different from ours. On the explanatory account of causality, how could such possibilities be discounted? It would seem that the only way to discount them would be openly to adopt an anthropomorphic relativism with respect to causality, a position that would be repugnant to a realist and objectivist position on causation. We shall recur to this problem below.

We now turn to another general problem with the explanatory account of causation which is not often recognized. The notion of understanding, hence the notion of explanation, clearly admit of degrees. We can speak of *how well* something is understood by some one, and similarly *how well* an explanation succeeds in explaining something. In fact, the degree to which an explanation is successful as an explanation can be measured using a variety of distinct, though related, criteria; Achinstein mentions relevance, correctness, depth, completeness, unification and manner of presentation, and all of these criteria permit graduated evaluations.²⁰ On the other hand, what might it mean to attach degrees to causal relations? Or are there just the right degrees of these properties such that an explanation having them just to those degrees will yield a causal relation? It would seem that what needs to be done is for the advocate of the explanatory approach to show that different explanations of the same phenomenon that do not rate equally on these criteria can, and often do, make references to the same explanatory facts or events. This, I do not believe, is a very difficult problem; it is, nonetheless, a problem that should be dealt with by those who advocate the explanatory approach to causation.

We now come to our final point, a point that I believe is of considerable philosophical interest and importance. It concerns two philosophical positions which could be called 'causal realism' and 'causal idealism' respectively. The two viewpoints I have in mind can be explained by reference to Hume, because Hume can be considered either as a causal realist or idealist depending on which of the two interpretations of Hume's views on causation is chosen. An historically accurate interpretation of Hume's theory is not our main concern; what is of interest is the fact that strands of both realism and idealism are clearly present in Hume. For, on the one hand, he refers to

causation (along the resemblance and contiguity) as “the cement of the universe” on which the “operations of the mind must in a great measure depend”.²¹ Hume qualifies these remarks by saying that causation is, ‘to us’, the cement of the universe, but Hume’s statement, if taken literally, suggests the view that causal connections are *objectively real connections* in the world that are independent, in their existence and essential nature, of human cognition and belief. Whether or not there are humans, or indeed any other sorts of beings capable of cognition, rising temperatures would cause ice to melt, gravity would cause bodies to accelerate, and sunlight reflected from the surface of the earth would cause earthlight on the moon. At least, so thinks the causal realist.

But there is another aspect of Hume’s account of causation. Of the notion of causal necessity, Hume says, “Upon the whole, necessity is something that exists in the mind, not in objects,”²² and goes on to give the following vivid description of a causal realist’s reaction to his account of causal necessity.²³

But though this be the only reasonable account we can give of necessity, the contrary notion is so riveted in the mind from the principles above-mentioned, that I doubt not but my sentiments will be treated by many as extravagant and ridiculous. What! the efficacy of causes lie in the determination of the mind! As if causes did not operate entirely independent of the mind, and would not continue their operation, even though there was no mind existent to contemplate them, or reason concerning them. Thought may well depend on causes for its operation, but not causes on thought. This is to reverse the order of nature, and make that secondary, which is really primary.

Hume’s reply to this is essentially that it’s no use to protest:²⁴

I can only reply to all these arguments that the case here is much the same as if a blind man should pretend to find a great many absurdities in the supposition that the colour of scarlet is not the same with the sound of a trumpet, nor light the same with solidity. If we have really no idea of a power or efficacy in any object, or of any real connection betwixt causes and effects, it will be to little purpose to prove that an efficacy is necessary in all operations.

The point of view Hume seems to hold in these passages is what I call ‘causal idealism’. To be sure, Hume’s target here is the idea of ‘necessary connection’, and perhaps not that of causal connection itself. But we can say this: on the interpretation of Hume according to which necessary connection is an essential ingredient in the concept of causation, Hume is a causal idealist. Causal connections are the projections of our mind with no genuine objective correlates in the world; they are mere creatures of our belief, custom and association.

On the other hand, if we take Hume's position on causation as expressed in the famous statement "we may define a cause to be an object followed by another, and where all the objects, similar to the first, are followed by objects similar to the second",²⁵ that is to say, if we take Hume's initial account of causation *sans* necessary connection, then we have a realist account of causation. For, presumably, the relations of temporal precedence, spatio-temporal contiguity and constant conjunction are objective relations in the world that are independent of the 'operations of the mind'.

I hope that the outlines of the two fundamentally opposed viewpoints on causation that I have called 'causal realism' and 'causal idealism' have been made clear enough for our present purposes. Among the historical figures we can classify, I believe, Aristotle and Kant as causal realists; as we saw, Hume's allegiance to one or the other school will depend on the interpretation of his doctrines; Mill is probably a causal realist. Some clear examples of causal idealists come from contemporary philosophy: Wittgenstein in the *Tractatus*²⁶ (5.1361: "Belief in causal nexus is *superstition*"), Hanson (as the quotation from him at the beginning of the paper attests), and, by stretching things a bit, Goodman²⁷ (see his explication of lawlikeness in terms of 'entrenchment'). Contemporary positivists are generally Humeans and share Hume's ambivalent attitude toward causal efficacy (if causal efficacy is part of the causal relation, they would be causal idealists; if law-like constant conjunctions are sufficient for causal relations, they would be causal realists).²⁸ Philosophers like Collingwood, von Wright, and Gasking,²⁹ who attempt to explain causation in terms of concepts pertaining to human actions, are best classified, I believe, as *relativistic* causal realists – relativistic because of the injection of *human* agency into causal relation.

The point of bringing out all this is that I believe the explanatory approach to causation is a form of causal idealism. Explanation is essentially and primarily an epistemological concept; it has to do with 'understanding', 'intelligibility', and integration and systematization of our beliefs and knowledge about the world. To base the analysis of causation on the concept of explanation is to project an aspect of our knowledge ('the operations of the mind') onto the world. Compare this with the kind of account a causal realist would give of the relation between causal relations and explanatory relations. A fundamental tenet of causal realism would be the claim that causal connections are the *objective correlates* of explanatory relations – at least, the

relations embodied in the so-called 'causal explanations'. Explanations serve to explain, and correct explanations are correct *because* they reflect the objectively existent determinative connections among events, states, and processes, just as our beliefs are correct *because* they 'correspond to' objective facts and conditions. To explain causation in terms of explanation would, on the realist view, be somewhat like basing an account of the concept of *fact* on the concept of *true belief*. Explanations, as we noted earlier, are subject to evaluation on various criteria such as appropriateness and relevance, but what makes an explanation *correct* as an explanation is the fact that it mirrors what's *out there* independently of our cognition. The causal realist believes that trying to explain certain properties of the causal relation, such as transitivity and asymmetry, in terms of the (alleged) fact that the explanatory relation has these properties is going the wrong way. The right way is to explain why explanations have these properties (to the extent that they do) by reference to the corresponding properties of the causal relations. Where these properties break down for explanation we would need to refer to the specifically epistemological and psychological character of explanation. On the realistic view, the relation between explanation and causation is rather like the relation between seeing an object and the properties of the object, or that between true belief and fact. The idealist, too, could agree that these relations are rather similar, but he will say, for example, that the object perceived and its properties are mere logical constructions out of our 'ideas'.

This is not the place to give an extensive argument for causal realism but I wish to note a few points that would be relevant in assessing the comparative merits of the two opposing viewpoints. First, on the idealist view it would be difficult, I think, to give a plausible and defensible account of the 'correctness' of an explanation. Perhaps, the idealist would want to say that such a concept as 'correctness' has no application to explanation, but this in itself strikes me as an implausible view. I believe that the correctness or accuracy of an explanation has nothing logically to do with the success of an explanation in creating a sense of relief from puzzlement or 'intellectual satisfaction'. In any successful explanatory episode there must be some theoretical content conveyed that is 'correct' or 'accurate', and which can be judged to be so on an objective ground. Second, it seems to me that whatever plausibility that attaches to various 'causal theories' of philosophical

topics, such as the causal theories of perception, perceptual objects, objects of emotion, naming and reference, inferential belief, knowing, etc. will quickly vanish if causation is understood nonrealistically, as a mere projection of some internal epistemological state with no corresponding objective basis. Finally, the split between causal idealism and realism obviously ties in with the larger division between idealism and realism in general. Although causal idealism appears to be logically consistent with general realism concerning knowledge and reality, it is also clear that there is a high degree of affinity and congruence between causal realism and global realism, and also between causal idealism and global idealism. Our discourse, both scientific and ordinary, appears to be thoroughly imbued with causal concepts, as witness the prevalence of causative verbs such as 'kill' and 'break', and the marriage between causal idealism and general realism is likely to result in a rather strange and unattractive picture of the world.

It seems to me that the twin issue of causal relativism vs. absolutism and of causal idealism vs. realism are of central importance and interest in assessing the explanatory approach to the analysis of causation. It is the standpoint one takes on these issues, rather than the degree of success or failure in formulating a detailed account or constructing counterexamples, that will shape one's overall view of the nature of causality and its role in our knowledge of the world.³⁰

The University of Michigan

NOTES

* Work on this paper was supported in part by a National Science Foundation grant.

¹ N. R. Hanson, *Patterns of Discovery* (University Press, Cambridge, 1958), p. 64.

² See, e.g., *Posterior Analytics*, Bk. I, ch. 2. For a brief but clear statement of Aristotle's views, see William A. Wallace, *Causality and Scientific Explanation* (University of Michigan Press, Ann Arbor, 1972), pp. 11 ff.

³ *Op. cit.*, p. 54.

⁴ Michael Scriven, 'Causation as Explanation', *Noûs*, IX (1975), 10. An earlier article by Scriven, 'The Logic of Cause', appeared in *Theory and Decision* 2 (1971), 49–66.

⁵ *Ibid.*, p. 11.

⁶ Carl G. Hempel, *Aspects of Scientific Explanation* (Macmillan, New York, 1965), pp. 347 ff.

⁷ *Foundations of Historical Knowledge* (Harvard University Press, Cambridge, 1965), p. 60.

⁸ *Ibid.*, p. 73.

⁹ *Ibid.*, p. 76.

¹⁰ See for details ch. 3 on 'Causality' in *Patterns of Discovery*.

¹¹ 'Causation as Explanation', p. 13.

¹² "'Explaining What" in History', in Patrick Gardiner (ed.), *Theories of History* (Free Press, Glencoe, Ill., 1959).

¹³ 'Causation as Explanation', p. 13.

¹⁴ *Ibid.* (We may note that any transitive and irreflexive relation is necessarily asymmetric, so that the question of asymmetry is not open to Scriven.)

¹⁵ The classic sources include Ernest Nagel, *The Structure of Science* (Harcourt, New York, 1961) and Carl G. Hempel, 'The Logic of Functional Analysis', in *Aspects of Scientific Explanation*. For current references see Peter Achinstein, 'Functional Statements', *Philosophy of Science*, 44 (1977), 341–67; Robert Cummins, 'Functional Analysis', *Journal of Philosophy*, 72 (1975), 741–65; David Hull, *The Philosophy of Biological Sciences* (Prentice-Hall, Englewood Cliffs, N.J., 1974).

¹⁶ See, e.g., Lawrence Sklar, 'Types of Inter-theoretic Reduction', *British Journal for the Philosophy of Science*, 18 (1967), 109–24; Kenneth Shaffner, 'Approaches to Reduction', *Philosophy of Science*, 34 (1967), 137–47; Robert Causey, *Unity of Science* (Reidel, Dordrecht, 1977).

¹⁷ Achinstein, *Law and Explanation* (The Clarendon Press, Oxford, 1971); Bromberger, 'An Approach to Explanation', R. J. Butler (ed.), *Analytic Philosophy* (Blackwell, Oxford, 1965).

¹⁸ 'Causation as Explanation', p. 12 (my emphasis).

¹⁹ *Law and Explanation*, p. 67.

²⁰ *Ibid.*, pp. 78 ff. See also p. 72.

²¹ *An Abstract of A Treatise of Human Nature*.

²² *A Treatise of Human Nature*, ed. L. A. Selby-Bigge, p. 165.

²³ *Ibid.*, p. 167.

²⁴ *Ibid.*, p. 168.

²⁵ *An Enquiry Concerning Human Understanding*, Section VII.

²⁶ *Tractatus Logico-Philosophicus*. See also 6.32, 6.36, 6.3611. A precise interpretation of Wittgenstein of course is a difficult matter. William Zulch called my attention to these passages in the *Tractatus*.

²⁷ Nelson Goodman, *Fact, Fiction and Forecast* (Harvard University Press, Cambridge, 1955).

²⁸ It is noteworthy that logical positivists in general eschewed serious discussion of causation; their overt concern was directed to such related problems as lawlikeness, dispositions, conditionals, and explanation.

²⁹ R. G. Collingwood, 'On the So-Called Idea of Causation', *Proceedings of the Aristotelian Society* 38 (1938); G. H. von Wright, *Explanation and Understanding* (Ithaca, N.Y.: Cornell University Press, 1971); Douglas Gasking, 'Causation and Recipes', *Mind* 64 (1955), 479–87.

³⁰ This paper incorporates some material from my comments on Michael Scriven's 'Causation as Explanation' in a symposium at the meetings of the American Philosophical Association, Western Division in 1975 in Chicago. For a more detailed systematic discussion of causal realism and idealism, see my 'Causal Realism vs. Causal Idealism' under preparation.