

Resisting Explanation¹

Introduction

During the last few decades the concept of explanation has received intense philosophical study within the philosophy of science. Beyond the borders of this subdiscipline, however, it remains a neglected topic. This is despite the fact that explanation is generally recognized as a central concept in the study of rationality. In epistemology, for instance, essential reference to explanation is made by those who base their theory of epistemic justification on the inductive principle known as 'inference to the best explanation'. In informal logic, too, it has long been conventional to distinguish between argumentative and explanatory reasoning. In neither of these contexts, however, has the nature of explanation been satisfactorily articulated.²

In this paper I suggest that there are several important reasons that explanation has been neglected, even resisted, outside the context of the philosophy of science. The reasons are diverse, but they are all in some way related to the desire to comprehend the epistemological significance of explanation in a priori terms. Despite the fact that most philosophical study of rationality now claims to be in some sense 'naturalistic' there is still much uncertainty and discomfort about what this actually entails. This is especially in evidence in the study of explanation where it has become increasingly clear that most of the criteria for producing and evaluating explanations are fundamentally pragmatic, and hence largely immune from purely a priori considerations.

After proposing three reasons that explanation has been resisted I'll suggest two changes that would significantly improve our chances of treating this concept effectively. The first is a simple adjustment to the traditional vocabulary, one that facilitates the attempt to represent explanation and argument as making distinct and equally fundamental contributions to rational inquiry. In support of this move I will argue that the conventional vocabulary reflects the mistaken assumption that the concept of rational support can be fully expressed as the relation of entailment between premises and a conclusion. Although on my view discerning the logical relation between premises and conclusions remains essential to any rational evaluation, full evaluation requires us to focus attention on the relationship between arguments and explanations.

Secondly, I will suggest a model of inquiry that represents the relationship between argument and explanation as one of mutual generation and support. This is something that is difficult to appreciate when argument and explanation are represented

as distinct epistemic relations, but easily seen when they are represented as integrated cognitive processes. Briefly, argument and explanation are linked at both ends so that the product of one serves as the raw material for the other. Fueled by environmental input, this relationship can be understood as the engine of rational inquiry. I will also suggest that this approach to explanation provides the basis for a clear articulation and defense of the principle of inference to the best explanation.

To Know and Understand

The fundamental intuition supporting all independent inquiry into the nature of explanation is that we can *know* something without *understanding* it. This view is held by philosophers with very different ideas about the nature of explanation itself. Pierre Duhem expressed it in his conventionalist interpretation of scientific theories when he claimed that explanation is the task of metaphysics:

To explain is to strip reality of appearances covering it like a veil to see the bare reality itself. ([1906] 1962, p. 7)

According to Duhem, science allows us to codify observational knowledge, but it does not afford understanding. On the other hand, in the opening lines of their modern classic "Studies in the Logic of Explanation," Hempel and Oppenheim claim that:

To explain the phenomena in the world of our experience, to answer the question "Why?" rather than only the question "What?", is one of the foremost objectives of all rational inquiry; and especially scientific research in its various branches strives to go beyond a mere description of its subject matter by providing an explanation of the phenomena it investigates. (1948, p. 135)

Clearly, the meaning of the distinction between knowledge and understanding is the source of profound philosophical differences, but the distinction itself is foundational. Most philosophers of science acknowledge that the need for a theory of explanation is as basic as the need for a theory of knowledge, and few would claim that these projects are identical.

From this point of view, then, it should be surprising that outside the philosophy of science the study of rationality is generally conducted as an inquiry into the nature of knowledge and epistemic justification rather than an inquiry into the nature of explanation and understanding. How do we account for this?

Reason 1: Equivocal Use of 'Argument'

For most philosophers the study of rational inquiry can be defined as the study of *argument*. It is through argument that things become known, as opposed to merely believed. Even those who, like Hume, hold that argument by itself lacks the power to convince, will agree that to challenge the rationality of a claim is, at bottom, to ask for an argument in support of it. But the term 'argument' has at least two distinct meanings, both of which are employed in philosophy, and the acceptability of this definition depends critically on which sense is being employed.

In most logic books an argument is defined as a finite sequence of sentences, called premises, together with another sentence, called the conclusion, in which the premises are intended (or taken) to entail the conclusion. This definition is perfectly adequate for formal logic. In fact, given the basic aim of formal logic- to develop and deploy generalized procedures for proof and interpretation in formal languages - it is more than adequate. Strictly speaking there is no need to refer to the *intention* that the premises should entail the conclusion. While this may help to block the pointless evaluation of sequences of sentences not so intended, no formal problem arises from considering any sequence of sentences at all to be an argument. In fact, speaking even more strictly, in the context of formal logic, the term 'argument,' is strictly eliminable. The concept of entailment (which may be explicated either proof-theoretically or model-theoretically) is all that is really needed.

Outside the context of formal logic, however, our purposes extend well beyond proof and model theory to questions about rational agency. In this sense, the study of rational inquiry does require explicit reference to intentions, for it is impossible to say whether an individual is thinking or behaving rationally without making use of some theory about what she is trying to accomplish.³ Moreover, as the term argument is ordinarily used in this context, the relevant intention is not just that the premises be understood to entail the conclusion, but rather that premises provide a rational justification for *believing* the conclusion. (The distinction is essential; to see why, consider that while P entails P, P is not a *reason* for believing P.) But whereas the formal sense of the term is epistemically neutral, this justificatory or evidential sense of the term is loaded in such a way as to exclude explanation from the category of argument. For explanation does not involve the intention to provide reasons for believing a conclusion;

rather, in an explanation the conclusion is understood to represent an accepted fact, and the reasons are intended to provide some understanding of why the fact actually occurs.

Now, what sense of the term 'argument' do we have in mind when we use it to explicate the nature of rational inquiry? I think most philosophers would, when they are being careful, say that it is the formal sense. The question whether an argument is valid, i.e., whether the premises entail the conclusion, is essential to all rational evaluation. This, as we just noted, in no way excludes explanation from the study of rational inquiry. The problem is that even in philosophy it is not the formal sense, but the ordinary evidential sense of the term 'argument' that is by far the most commonly employed. Almost invariably when we request an argument we are not simply asking for premises that entail a conclusion, but rather for premises that will convince us of the truth of the conclusion. Indeed, under normal circumstances it would actually be confusing to ask for an argument for a conclusion, when the truth of the conclusion is not at issue and what is really wanted is an explanation.

Obviously *if* we were to fail to keep these two senses of 'argument' clear and distinct then equivocal usage would be more likely to result. But there are two very different ways in which the equivocation could occur. One is that the formal sense of 'argument' would be assumed in contexts in which the evidentiary sense is appropriate. This is very unlikely simply because the evidentiary sense of argument is the predominant usage and therefore supplies us with the natural default. Second, the evidential sense of argument could be assumed in contexts in which the formal sense is appropriate. This is likely for exactly the same reason. Given that the evidential sense provides the default usage, only a very strong grasp of, and attentiveness to, the difference between these two senses of argument will prevent us from applying the evidentiary sense of the term when the formal sense is intended.

This situation is further exacerbated by the fact that often there is no real penalty for the equivocation. Consider the following statement:

A person believes S rationally only if (a) S is self-evident or (b) there is a strong argument in support of S.

Obviously, someone could make this claim with the formal sense of argument clearly in mind, though no great confusion would result if a reader interpreted it according to the evidentiary sense.⁴ On the other hand, consider a statement like:

It is pointless to produce an argument in support of a conclusion whose truth is already accepted.

This claim, if made with the formal sense of 'argument' in mind, is false. It is completely appropriate to produce *explanations* in support of accepted conclusions- in fact, it is *only* appropriate to produce them in such contexts- and explanations are arguments in the formal sense of the term. But the statement is quite plausible if made in accord with the evidentiary sense of 'argument'. Hence, anyone who receives this statement according to the natural default would be unable to recognize that, as intended, it is false.

For these reasons I suggest that that the simple predominance of the evidential sense of argument can foster an uncritical acceptance of the idea that rationality is fundamentally a matter of having justified beliefs. Although I think that this claim has a great deal of prima facie plausibility, it is reasonable to ask for some actual evidence in support of it. It is an empirical claim, after all. Unfortunately I have little of substance to offer. There is, however, one telling piece of evidence to be found in introductory logic textbooks.

It is common in such texts to begin by introducing the formal definition of argument, and to follow this with exercises that are examples of ordinary language reasoning from which the student is asked to reconstruct formal arguments. If the authors of such texts worked with a clear understanding that explanations were arguments in this sense, then, since explanations are no less common in ordinary reasoning than arguments, we should expect to find them well-represented among the examples. In fact, however, the vast majority of these examples actually turn out to be arguments in the evidentiary sense of the term. The theory that authors equivocate on the two senses of argument predicts exactly this result; i.e., a person who defines 'argument' formally, but uses it evidentially would fail to recognize explanations as arguments.

Of course, this analysis reduces to the claim that in some degree our prevailing concept of rational inquiry rests on a simple equivocation, and I realize that this is a prima facie implausible complaint to lodge against competent logicians and epistemologists. However, the alternatives are no more appealing. Let's be clear about what they are.

On the one hand, we might claim that the normative tasks of epistemology and informal logic can be fully understood by reference to the formal concept of entailment. This, of course, is highly problematic since intentional states resist truth-functional analysis. I would also require us to regard all irreducibly pragmatic considerations as fundamentally irrelevant to the task of logical and epistemological inquiry. This is an

extremist position, though it might reasonably be attributed to the school of Logical Positivism, at least in some of its early formulations.

On the other hand, one might defend the idea that the evidential sense of argument does, in fact, provide an adequate characterization of the aims of rational inquiry; this, either because the nature of explanation is not a relevant subject, or because explanation can be fully understood by reference to the evidential sense of argument. The first alternative has to be rejected. Explanation is an essential topic in the study of scientific rationality and explanation can not plausibly be construed as a uniquely scientific concern. The second alternative deserves to be taken more seriously, however, and we consider it in the following section.

Reason 2: The Traditional Model of Explanation

According to the traditional model (Hempel and Oppenheim 1948) an explanation is just a certain kind of argument. This view is now widely rejected in the philosophy of science, but, as no other model has achieved the status of a received view, it continues to be influential. If correct, it would justify any reluctance to treat explanation as a distinct concern in the study of rationality.

The traditional deductive-nomological (DN) model of explanation draws on both senses of the term 'argument' discussed above. Officially, the model employs only the formal sense of the term. It states that an explanation is a deductive argument in which a conclusion representing the phenomenon to be explained is derived from premises representing certain initial conditions and at least one general law. However, Hempel is clear that this analysis is an attempt to justify the intuition that an explanation is fundamentally "an argument to the effect that the phenomenon to be explained...was to be expected in virtue of certain explanatory facts" (Hempel 1965, p. 336). Here, Hempel clearly employs the evidential sense of 'argument' to provide a counterfactual analysis of explanation. Where an argument does justify us in predicting a certain event, an explanation *would have justified* us in predicting it, had it not already occurred. In a sense, an explanation is just an argument that's late for work.

The problems of Hempel's model are amply documented (Salmon et al 1992, p. 20-23) . Our purposes will be served by looking briefly at the criticism that directly challenges the identification of explanation and argument. This is the observation that many arguments would justify a prediction which nevertheless do not seem to explain their conclusion. For example, one could argue that the presence of a skunk was to be expected on the basis of a sudden stench though, intuitively at least, the stench does not

explain the presence of the skunk. The usual way of framing this problem is to claim that explanations are necessarily asymmetrical: if E explains C, then C doesn't explain E. Arguments, however, are not necessarily asymmetrical. Depending on the context, a stench may be as much a basis for predicting a future encounter with a skunk as a skunk is for predicting a future encounter with a stench.

Hempel's response to this criticism is chastening and instructive. The DN model is based on the view that explanations and predictions are structurally identical. Both, that is, have the logical form of a deductive-nomological argument. In Hempel's view, the problem of asymmetry (as well as most other criticisms of the DN model), is a red herring in the sense that it dwells on purely pragmatic differences between prediction and explanation. There are no *a priori* reasons why a skunk stench can't explain a future encounter with a skunk.⁵ If we believe otherwise, it is because we accept certain physical theories the content of which is not determined by logic, but by the current state of scientific knowledge. A theory of explanation that rules out the possibility that skunk stench may explain the appearance of skunks, would have overstepped what can be known on *a priori* grounds.

Hempel's response fits perfectly within the view of philosophical inquiry that was current at the time that he made it. Under the influence of Logical Positivism, philosophical inquiry was understood to be limited to the formal analysis of concepts. If we accept these limitations, then the asymmetry criticism actually has very little bite. To the extent that the power of explanation to counterfactually justify prediction can be understood in purely formal terms, the deductive-nomological model does seem to account for it. It seems, then, that what is properly criticized in the DN model is not so much the structural identity of argument and explanation, as the intuition that it was intended to support, *viz.*, the *epistemic identity* of explanation with the evidential sense of argument.

Ultimately, the weakness of Hempel's reply to the asymmetry objection must be understood in conjunction with the decline of Logical Positivism, and the gradual acceptance of the view that philosophical inquiry is not limited to the formal analysis of concepts. Most of the traditional criticisms of the DN model of explanation are now understood to indicate its failure (and the failure of Logical Positivism generally) to connect adequately to the ordinary and scientific meaning of this term. This means that even if we accept the structural identity thesis- and given Hempel's quite narrow concerns, I see little reason not to- we are only accepting that explanations and evidential arguments have a common element, not that this element defines their respective contributions to rational inquiry.

My point here is easily lost. To summarize: Hempel's theory of explanation was an attempt to show that explanations are actually arguments in *both* senses of the term. The view was appealing, not the least because it provided a substantially a priori basis for evaluating explanations. Although the DN model is now widely rejected in the philosophy of science, it is still the only theory of explanation that someone outside the philosophy of science is likely to have much acquaintance with. To the extent that it remains influential, the DN model discourages us from granting explanation independent status in theory of rational inquiry.

Reason 3: The Relational Analysis

Although the asymmetry criticism undermines the attempt to identify explanation with the evidential sense of argument, it does not necessarily support the idea that explanation makes an independent contribution to rational inquiry. In fact, in some ways it is inimical to it. This is because in accepting explanatory asymmetry, we now need to provide an analysis of explanation that accounts for it, and the conventional ways of doing this do not lend much plausibility to this idea.

One simple approach is to claim that explanatory asymmetry reflects the metaphysical asymmetry of causation. I.e., if all explanation is causal in nature, then explanatory asymmetry can be understood by appeal to the temporal ordering of causal relations and the apparently a priori truth that a later event can not be the cause of an earlier one. This approach allows for the identification of explanation with a certain subset of evidential arguments, namely those that contain causal laws, but it is fraught with difficulties. For our purposes it will suffice to point out that it is utterly stipulative to insist that explanation, in either scientific or ordinary contexts, is essentially tied to causation. In the scientific context, many explanations have a purely mathematical or reductive nature, involving no diachronic relationship whatsoever. In ordinary contexts, the glaring counterexample is folk psychological explanations, which resist straightforwardly causal analysis.⁶

The popular alternative to the causal analysis is to locate asymmetry in the pragmatics of communication. The basic idea here involves recognizing that explanation is essentially a communicative process in which questions posed by one individual ("why-questions" according to the best known view [van Fraassen 1980]) are supplied with answers by another. All questions are interest driven, but explanation-seeking questions seem especially so. For not only do individual interests generate the question, but to a surprising extent they seem to determine what counts as an acceptable answer.

To take a simple example, a person who asks why a certain traffic accident occurred may reject as non explanatory a DN argument (demonstrating, say, that whenever two objects are traveling so as to reach the same point at the same time, a collision will result) while accepting an unrigorous non deductive explanation in terms of brake failure or personal negligence.

According to the communicative account of explanation asymmetry is, as Hempel claimed, a function of interests. It is just a fact, and not a necessary one, that individuals whose interests are satisfied by an explanation that runs in one direction, will not be satisfied with an explanation that runs in the opposite direction. The interest-relativity of explanation has been stressed repeatedly in the last several decades so that today even the most stalwart causal realists (e.g., Salmon 1992, p.33-39) seem inclined to concede it. Indeed, in many ways this is the solution that Hempel himself, urged. The difference is that he, and logical positivists generally, did not regard the pragmatics of explanations as having any philosophical interest.

Although it is no longer feasible to reject pragmatic issues as non philosophical, I am inclined to agree that the pragmatics of explanatory communication are of secondary interest to the theory of rational inquiry. The communicative approach may provide conditions under which understanding is *transferred* from one individual to another, but it has nothing to do with its creation. But the fact is that the pragmatics of explanation do not have to be understood in communicative terms. Beside being a form of communication, explanation can also be understood as a process of *discovery* in which individuals (or groups) attempt to explain things to themselves. The essential difference between the two is this: when explainer and explainee are engaged in communication, the explainee ordinarily assumes that the explainer has access to the required information. Explanations may be rejected as irrelevant to the interests which generated the questions, but the explainee typically does not question the information itself.⁷ In the context of discovery, however, the epistemological situation is reversed. Here, explainer and explainee are one; explanations do not appear as pre-established truths, but as hypotheses produced by the inquirers themselves. Ideally, in this context there is no doubt as to the explainer's interests, so the task is not to comprehend the hypothesis nor its relevance to the question at issue, but to determine whether the hypothesis itself should be accepted.

The discovery model of explanation provides the obvious framework for analyzing the process by which individuals or scientific communities extend our knowledge and understanding of the world. The process itself is perfectly familiar to scientists and non scientists alike, and it is remarkable that this way of looking at the

concept has gone largely unexplored in philosophy.⁸ The most obvious reason for this is that discovery processes have been taboo in the philosophy of science ever since philosophers officially distinguished the context of discovery from the context of justification. As Karl Popper put it:

..the initial stage, the act of conceiving or inventing a theory, seems ... neither to call for logical analysis nor to be susceptible of it...There is no such thing as a logical method of having new ideas or a logical reconstruction of this process. (Popper 1959, p. 31-32)

The distinction between justification and discovery was challenged forcefully in the work of naturalists like Kuhn (1962), Quine (1961) and their respective disciples, who (for different reasons) rejected the distinction between philosophical and psychological inquiry, and who (again, for different reasons) advocated a holistic account of rationality in which the logical possibility of subjecting *individual* theories to disconfirming evidence does not exist. But while today many, if not most, analytic philosophers claim to be in essential agreement with the main elements of the naturalistic critique, most are deeply reluctant to identify with the postmodern anti-foundationalism that these views have spawned. The result is that, for the most part, philosophical inquiry takes place as before, and the taboo on discovery is still widely observed.

Perhaps the most recalcitrant of all intuitions in this context is the one that tells us that rationality is fundamentally a matter of grasping logical, epistemic, or metaphysical *relations*. This is not a false assumption, it is just built into the analytical method. To analyze a concept just is to show how it is related to other concepts; to grasp a concept just is to grasp these relations. But, while conceptual analysis is an indispensable tool of inquiry, it is not self-sufficient. Underlying the concepts and inferential relations are the phenomena themselves, and no amount of a priori analysis will tell us whether the concept will yield an adequate grasp of the phenomena.

With particular reference to explanation, then, my point is not that it is seriously wrong-headed to analyze explanation in terms of logical, causal, or even communicative relations. Rather, it is that the relational approach can not by itself reveal how explanation contributes to the process of rational inquiry. Now, it is interesting to note that with respect to the concept of inductive *argument* this point has been appreciated for some time. Quine (1969) and others made the case for conceiving of inductive inference, not in terms of an a priori relation between premises and conclusion, but as a cognitive process driven by environmental input and constrained by the pragmatic requirements of the cognitive system in question. This view has spawned (or at least dovetailed nicely with) an entire cognitive science research program dedicated to overcoming alleged

inherent shortcomings of the purely relational accounts of induction. Characterizing this program, Holland et al write:

Although we have no doubt that considerations of probability theory and statistical inference are often highly relevant to induction, we do not believe that the primarily formal and syntactic approach to induction can succeed. The formal approach resolutely ignores the kinds of events about which the person is trying to make inferences as well as the goals that the inferences serve. In our view, it is because of its blindness to environmental and cognitive realities that the Carnapian approach to inductive reasoning has given rise to numerous paradoxes....[These] are not mere problems to be solved by some refined syntactic account of induction but rather are symptoms of the fundamental inadequacy of such accounts. (Holland et al 1989, p. 6)⁹

Within this research program the concept of explanation has been put to work as well. Cognitive scientists make of the concept of 'explanation-based generalization' in order to account for the way cognitive systems revise their beliefs in accord with new information. To borrow an example from Steven Pinker (1997, p.127), when I learn that chimpanzees will eat onions, this may or may not cause me to update my probability that gorillas like onions depending on whether I assume that eating habits of gorillas and chimpanzees have the same explanation. If I know only that they are both primates, I may be inclined to make this assumption. On the other hand, if I know that chimpanzees are omnivores and gorillas are herbivores with distinctly finicky eating habits, I will not.

This way of thinking about explanation is not new to philosophy. Philosophers like Gilbert Harman (1986) and Keith Lehrer (1992), for example, have attempted to develop coherentist alternatives to traditional foundationalist epistemologies on the basis of exactly the same approach to explanation. Indeed, most analytic philosophers will recognize in "explanation-based generalization" the inductive principle known as "inference to the best explanation" (IBE), which is just the contemporary term for what C. S. Peirce more succinctly termed "abduction".

Although this approach to explanation is a step in the right direction, it is not exactly what I am advocating. The reason is that the concept of explanation is actually *still* being treated in fundamentally relational terms. In explanation-based generalization, for example, the inductive processes of a cognitive system are understood by reference to its having grasped an irreducibly explanatory relationship. Although this may suffice as the practical basis for building cognitive systems that think like we do, it doesn't tell us any more than traditional theories precisely what this relationship amounts to. It is actually because of the opacity (or vacuity) of this relationship that contemporary

epistemologists have remained deeply skeptical of attempts to ground inductive inference in the concept of explanation.

Although naturalistic epistemologists and cognitive scientists eschew the purely a priori syntactic approach to rationality developed under Logical Positivism, they have not necessarily freed themselves from its influence. While they reject the idea that inductive argument can be understood in purely relational terms, their understanding of the process of induction nevertheless remains grounded in an appeal to irreducibly explanatory relations. If Hempel incorrectly reduced explanation to argument, then cognitive science has incorrectly reduced (some forms of) argument to explanation. The middle way I am suggesting is to recognize that, while relational talk with respect to both argument and explanation is indispensable, a full account of rational inquiry will recognize them as distinct processes contributing to the same general goal.

Fix the Vocabulary

So far we have been trying to account for the general failure of philosophy to study explanation in a significant way, and we have traced this to the fact that explanation both exposes and challenges our attachment to the a priori. If this is correct, then the obvious choice for philosophers who maintain a principled opposition to the naturalistic turn in philosophy is to stop resisting explanation and simply reject it outright. This is not the option I recommend, but, from the point of view of clarity, it would still have to be counted as an advance.

In the remainder of this essay I want to make some recommendations relevant to the option I do recommend, namely of fully incorporating explanation into the philosophical study of rational inquiry.

Earlier we noted that, as distinct from the context of formal logic, the general study of reasoning requires essential reference to intentions. Traditionally, that is, an argument is defined as a set of premises and a conclusion in which the premises are *intended* to support the conclusion. We also pointed out that the concept of support that interests us in the study of reasoning is not adequately captured by the concept of entailment employed in formal logic. This is illustrated by the fact that, while formally a proposition entails itself, it does not, in any sense relevant to the aim of reasoning, *support* itself.

One simple way to capture this difference between entailment and support is to say that, regarding a given conclusion, informal logic is concerned only with those premises that can also function as *reasons*. Minimally, this requires that we be able to

attribute a certain kind of intention to the speaker. But the aims of informal logic can not be satisfied by knowing simply that a person intends to give a reason in support of a conclusion. We need to know more. Specifically, we need to know whether the support should be understood as *causal* or *evidential* in nature.¹⁰ This is just the point we discussed in connection with explanatory asymmetry: If you understand that I intend "It stinks in here," to be taken as a reason for "There's a skunk outside," you will judge the reasonableness of this assertion very differently depending on whether you take me to be saying that the stench is the cause or the evidence of the skunk's proximity.

In philosophy, especially the philosophy of mind, it is customary to distinguish between reasons and causes. (Crudely, reasons move the mind; causes move the body.) This, however, is quite contrary to ordinary usage, where the term "reason" is used generically. Whether one is concerned with the cause of something being the case, or evidence for thinking that it actually is the case, it is, in either case, appropriate to ask for a reason.

In an effort to avoid the equivocal use of 'argument' discussed above, I propose that we simply refrain from using the term in the technical sense employed in formal logic and coin a different technical term for describing the generic relation of reason to conclusion. I suggest the term "rationale". This will allow us to restrict our use of the term 'argument' and 'explanation' to evidential and causal relationships respectively. Philosophers will, of course, find this an awkward transition, but to the unindoctrinated it should be quite intuitive.

These suggestions are depicted in Figure 1. This realignment of the vocabulary enhances our ability to describe the difference between argument and explanation. Rather than having to say that an explanation is a certain kind of argument (formally defined) which is nevertheless distinct from another kind of argument (evidentially defined), we can simply say that arguments and explanations are just two kinds of rationales, which can be distinguished by the function of their respective reasons; viz., in an argument the reason is presented as evidence for a belief, in an explanation the reason is presented as a cause of an accepted fact.

A Different Sense of Support

When we distinguish argument and explanation in this way the world of informal reasoning can start to look very different. Consider the following examples, which I have culled from a handout I use in an introductory logic class.

1. Linda must love lemon drops. She eats them all the time. I guess that's why she has so many cavities.
2. The car won't start again. I guess the battery is dead. That's what the problem was last time.
3. Mammograms are probably not worthwhile for women under fifty. Young women have denser breasts, which makes their x-rays very hard to read.
4. People in California love their privacy. Just look at how many people drive to work by themselves when it would be so much cheaper to carpool. That's why mass transit will never work in this state.

These examples are a bit spare with words; otherwise, I take them to be realistic representations of ordinary reasoning. Each of them, however, invokes what I call (somewhat awkwardly) an 'argument-explanation pattern', and they would be impossible to analyze accurately in terms of the evidential sense of argument alone. In the first example we see two distinct rationales: an argument for the belief that Linda loves lemon drops coupled with an explanation of the fact that Linda has cavities. In the second example we see an explanation of the fact that the car won't start coupled with an argument for the belief that the battery is dead. (The third and fourth examples I leave as exercises.)

Notice that in each of these examples, at least one proposition occurs in both the argument and the explanation. This is an important fact, but it can not be comprehended by an analytical model that fails to distinguish between these two types of rationale. If the relevant intentions to argue and explain can not be distinguished, then the reasoning contained in (1) can only be represented as a single rationale.

Argument-explanation patterns are familiar to everyone, but I think they are more than that. I think that arguments given in the absence of explanations, and explanations given in the absence of arguments are ordinarily regarded as *incomplete*. Arguments and explanations are different kinds of rationale which *support* each other in a sense that is simply not captured by the concept of logical entailment. Consider two observations in support of this claim.

The first is an appeal to introspection. What is the first question I ask once I have accepted an argument for a certain conclusion? If, for example, I am provided with compelling evidence that the neighbor just mowed *my* lawn, the first question that occurs to me is "Why?" It is not that without an explanation of the event I remain unconvinced of the conclusion, it is that without an explanation my understanding of the event is fundamentally incomplete. The explanation supports the argument, not in the sense of entailing it, but in the sense of completing our *understanding* of what has occurred.

Secondly, consider the structure of the typical magazine or newspaper editorial. Logic students are usually taught to analyze these as arguments for some kind of normative conclusion. What is missed on this approach, however, is that in addition to an argument for the view the author endorses there is almost invariably an *explanation* of the view that is being rejected. So, for example, an editorial in which the author argues that the United States should not support NATO expansion will normally be supplemented with an explanation for why the United States *is* supporting NATO expansion. Without such an explanation we feel, correctly I claim, that the author's case is incomplete. (This point may be initially difficult to appreciate because the explanations in question are usually psychological in nature. Psychological explanations of people's beliefs are easily confused with arguments in support of these beliefs.)

Of course, anyone who remains committed to the idea that the fundamental aim of rational inquiry is to establish knowledge through argument will be unmoved by observations such as these. One can always respond, for example, that the understanding we seek subsequent to knowledge is really just more knowledge; until an explanatory hypothesis has been independently established through argument it lacks the power to support anything at all. But the blade cuts both way: one could also say that the knowledge we seek subsequent to understanding is really just more understanding; until a justified belief has been adequately explained it lacks the power to support anything at all. These views are equally plausible and, I think, equally procrustean. The relationship between argument and explanation is simply not that of master to slave; rather, it is properly understood as a cooperative, mutually supportive relationship between equal parties.

The Dynamics of Argument and Explanation

In the context of logic it is useful to speak of argument and explanation in relational terms. However, as I indicated earlier, in order to contribute to a scientific theory of rational inquiry it is necessary to think of them as mutually supportive cognitive processes. We conclude this essay with some thoughts on how this is to be done.

We noted earlier that in naturalistic epistemology 'inference to the best explanation' has been popular as a candidate for the basic principle of inductive rationality, and that this idea has been developed in cognitive science as 'explanation-based generalization'. The rough idea here is simply that the explanatory power of a hypothesis is a reason for believing it to be true. The principle is appealing because it feels intuitively plausible and seems to have broad application. Consider three quick

examples. (1) IBE is employed in a popular argument for scientific realism in which it is claimed that the literal truth of scientific theories is the best explanation of their success (Richard Boyd in Leplin 1984). (2) IBE has been used to explicate subcognitive perceptual processes; here it is claimed that a perceptual object is inferred as the best explanation of sensory data (Rock 1983). (3) In ordinary problem-solving contexts, it seems intuitively correct to say that a person infers a hypothesis because it is explanatory. For example, I infer that my car battery is dead because it is the best explanation of the fact that the car will not start.

Despite its intuitive appeal, however, IBE is not supported by any well known theory of explanation. This is simply because no such theory conceives of explanation as an inductive principle. According to the DN model and its various successors, a hypothesis must be inductively confirmed *before* it can count as having any explanatory power at all. From this point of view, talk of *inferring* a hypothesis on the basis of its explanatory power is perfectly circular.

Of course we might just eschew theories of explanation altogether and agree to use the term to signify a primitive relation, but even with this generous concession inference to the best explanation is not viable in conventional terms. The main problem has been pointed out most effectively by Bas van Fraassen:

Inference to the Best Explanation is not what it pretends to be, if it pretends to fulfill the ideal of induction. As such its purport is to be a rule to form warranted new beliefs on the basis of the evidence, the evidence alone, in a purely objective manner. It purports to do this on the basis of an evaluation of hypotheses with respect to how well they explain the evidence, where explanation again is an objective relation between hypothesis and evidence alone....It can not be *that*, for it is a rule that only selects the best among the historically given hypotheses. We can watch no contest of the theories we have so painfully struggled to formulate, with those no one has proposed. So our selection may well be the best of a bad lot. To believe is *at least* to consider more likely to be true, than not. So to believe the best explanation requires more than an evaluation of the given hypothesis. It requires a step beyond the comparative judgment that this hypothesis is better than its actual rivals... [It] requires a prior belief that the truth already be more likely to be found in [the hypothesis] than not. (1989, p.142-143)

An even-handed consideration of IBE is well beyond the scope of this essay, so I will simply state here that van Fraassen's arguments (which themselves extend well beyond the relatively simple point made above) seem to me to discredit IBE completely as a principle of inductive inference. However, I am not interested in it as such. Considered as a cognitive process, we do not propose IBE as something that will "fulfill

the ideal of induction" but rather as part of an attempt to explain how cognitive systems do *in fact* select among the "historically given hypotheses."

We can begin to appreciate the cognitive significance of IBE as soon as we stop expecting it to function as an a priori inductive principle. It is as a *discovery* mechanism that IBE seems to have the most intuitive plausibility. The fact is that 'inference to the best explanation' is a bit of a misnomer. We do not (rationally)*infer* a hypothesis on the basis of its explanatory power; rather we *propose* it on this basis, and then subsequently submit it to the rigors of argument. This seems perfectly obvious in ordinary contexts. For example, I propose the hypothesis that my car battery is dead as the best explanation *I can think of* for the fact that the car won't start. It is a mistake, however, to infer it on this basis. Rather, I should actually accept the hypothesis only after the battery itself has been tested.

This can also be appreciated by reconsidering the hypothetico-deductive model of theory confirmation. This model is well known to be defective as an a priori account of inductive confirmation. However, it remains intuitive as an abstract description of the process of empirical inquiry: we generate and propose hypotheses on the basis of their explanatory power, but we infer them only after they have suffered a future trial by argument. This interpretation of explanation has always been available to us. The only difficulty is that it requires us to break the Popperian taboo on discovery processes.

Argument and explanation are distinct processes that create distinct products. Explanation generates causal hypotheses that, if true, would give us knowledge. Arguments create truths that, if they satisfy explanatory requirements, would give us understanding. These processes are connected and driven by a feedback relationship. The output of one process provides the input for the other. This relationship is depicted in Figure 2.

I just referred to the "satisfaction of explanatory requirements". This, of course, is intolerably vague as it stands. Indeed, the real work of developing an adequate account of the explanatory part of the process of inquiry is to determine with precision what these requirements are. In my view, these could vary with the type of cognitive system being investigated (or designed), however it seems reasonable to suggest that predictive control of the environment will be among the standard desiderata. This, anyway, I assume without argument in the following.

The interactive relation between the processes of argument and explanation may be described as follows. Explanation is a process that is triggered by a certain kind of input, viz., a surprising fact, a salient feature of our environment that we have somehow failed to predict. (E.g., the car won't start). One of our aims is to improve our

understanding of the environment so that events of the sort that we failed to predict in the past will not continue to surprise us in the future. Explaining a fact involves the formation of a causal hypothesis (The battery is dead.) This possible cause is the output of the explanatory process. But for any given fact there will always be a number of possible causes. Hence, the process of explanation will be useful as a way of gaining predictive control over our environment only if it is supported by another process whose function is to determine which, if any, of the possible causes should be accepted.

So the output of the process of explanation, a possible cause, is received as the input to the process of argument, as a possible belief. The process of argument involves the collection of evidence (the battery is tested) that will either convert this possible belief into an accepted fact, or dispose of it. (This, of course, is extremely simplified. Both acceptance and rejection can be a matter of degree and all such conclusions remain defeasible.) Significantly, however, this newly collected evidence may itself have the status of a surprising fact (the battery is practically new), for which possible causes may be sought (the alternator is overcharging the battery) and themselves subjected to the process of argument . In short, the interaction of argument and explanation is the basic mechanism for the growth of human knowledge and understanding.

The interpretation above begins with a predictive failure, but this is a property of the description, not the process, which is cyclical and can be entered into at any point. Alternatively, one might begin with a possible belief that suggests itself outside the context of any explanatory process, though it too will have the capacity to trigger the explanatory process as evidence is gathered in support of it.

It is important for the reader to understand that I do not regard this sketch of the dynamics of argument and explanation as even the beginning of a general theory of inquiry. I actually doubt very seriously that terms like 'argument' and 'explanation' will occupy in any part of such a theory. As with other theoretical sciences, such terms are very likely to be eliminated in favor of those that allow for precise modeling of the phenomena. (Of course, anyone conversant in the literature of cognitive science knows that this process is already well underway.) But good theoretical science must also connect with clear ordinary-language descriptions of the phenomena, and it is this that I believe has been lacking. If explanation is excluded from our basic conception of the reasoning process, neither our scientific understanding of inquiry, nor our ordinary understanding of it, can be adequate.

Conclusion

The aim of this paper has been twofold: I have attempted to *explain* why the concept of explanation has not been given a significant treatment within the general theory of rational inquiry. I have also *argued* that the process of rational inquiry may be profitably understood on the assumption that explanation and argument make equally fundamental contributions to it. Reflexive consistency is essential in this context, so I hope it is clear that the structure of this paper itself exemplifies both the cognitive interdependence and the logical independence of these two aspects of the reasoning process.

¹ I would like to thank two anonymous reviewers for their helpful criticisms of an earlier draft of this essay.

² It may reasonably be objected here that considerable work in explanation has been done outside the philosophy of science. The original deductive-nomological (D-N) model of explanation (Hempel and Oppenheim 1948) was conceived within the philosophy of science. However, the DN model is now widely rejected and it has become more common to analyze explanation on an erotetic model in which an explanation is seen as an answer to a why-question. In support of this view, philosophers of science (e.g., van Fraassen 1980) have drawn heavily on the work of linguists and formal logicians (e.g., Belnap and Steele 1976).

³I do not mean to suggest that our understanding of rational agency is necessarily tied to intentional psychology. The important point here is just that outside the context of formal logic judgments of rationality are fundamentally tied to a causal theory of behavior.

⁴ It's worth noting, however, that confusion might result if such a claim were supported with examples that struck the reader as fundamentally explanatory in nature.

⁵ In fact, there are situations in which it obviously could. For example, if skunk stench also functioned as a skunk attractant, then the presence of skunk stench would explain the future appearance of skunks.

⁶ This does not mean that, for example, the computational theory of mind is not a causal theory, only that the functional explanations it produces do not depend on identifying a type of mental event with any particular type of causal relationship.

⁷This is a defeasible assumption. For example, an explanation that contradicts an explainee's background knowledge would certainly be questioned at this level. Also, a form of explanatory communication obviously does occur between investigators mutually engaged in process of discovery. There is, I think, a subtle but important difference between *the act of explaining* something to someone else, which presumes a certain kind of authority on the part of the explainer, and the *act of suggesting an explanation* to someone else, which clearly does not rest on this assumption.

⁸ This remark will be received critically by philosophers who are both acquainted with the cognitive science literature and unaccustomed to making a hard distinction between the two disciplines. Many philosophers and cognitive scientists are now actively involved in computational modeling of the process of explanation as I describe it here, e.g., Thagard (1989), Holland et al (1986), Churchland (1989).

⁹ The authors here refer to Hempel's raven paradox and Goodman's grue paradox.

¹⁰As an anonymous reviewer pointed out, this restriction of explanation to causal relationships conflicts with an earlier point. In the discussion of the relational analysis of explanation I note that the the causal realist's resolution to the asymmetry problem is to identify explanation with causation. I criticize this gambit on the grounds that not all explanation is causal. Here I am using the term 'cause' in a much broader sense, one that can allow for a varitey of kinds of causal relationship, not all of which would depend on the irreversibility of time.