Simple Explanations and Reasoning: From Philosophy of Science to Expert Systems*

Daniel Rochowiak
205-895-6217
Johnson Research Center — Philosophy
University of Alabama in Huntsville
Huntsville, Alabama 35899

ABSTRACT

Explanation facilities are an important extension of the rule based paradigm. By using contrast why questions and a more textured notion of reasoning, a robust schema for simple explanations can be developed.

INTRODUCTION

At first glance it seems rather easy to characterize explanation. An explanation is a deductive argument that satisfies the conditions of empirical adequacy. [7] However, behind this apparently simple account can be found a great many issues. One of these concerns the pragmatics of explanation; deductive explanations often fail to explain anything to the person seeking the explanation. I will examine the pragmatic dimension of explanation and indicate how a more 'textured' notion of reasoning can enhance the explanation facilities of the expert system paradigm. The domain of interest will be a general one in which there are objects, states of objects and causal paths between the objects.

Among the facilities commonly found in expert system building tools are the why? and how? explanation facilities. Typically the why? facility is engaged at a prompt and reports the rule that is currently being examined, while the how? facility requests the user to identify a particular parameter for which the system has set a value and reports the rule by which it was set. Such facilities operate in the style of deductive explanation. In both there are conditional claims (laws or rules) together with specified conditions (initial conditions or user entered values) that deductively lead to particular conclusions (the explanandum, or the values of parameters). Deductive explanations have been criticized for attending more to the grounds of an explanation than the particular explanation of a given event. [8] Similarly the how? and why? facilities attend more to the rules of the system than to the events to be explained. The sorts of explanation offered within the expert system should not be confused with the sorts of explanations that might be considered ordinary in other contexts. The explanation facilities might explain, for example, why one would come to think that a particular part failed, but would not explain why the part failed. Such criticism points to the importance of pragmatic considerations in explanation.

VARIETIES OF EXPLANATION

Explanations come in many forms. Scientific explanations are one well studied group of explanations. One form of scientific explanation proceeds from a scientific law, theory or model to a deductive account of a phenomenon. Although Hempel's original formulation of such deductive nomological (DN) explanations has been much criticized, it provides both a good starting point and a base that, with suitable extensions and amendments, can capture a large range of scientific explanations.

The DN model of scientific explanation invites comparison to the notion of backward chaining in expert systems. The explanandun is known and the collection of scientific principles is searched in an effort to find the conditions which, if satisfied, would constitute the explanans. The collection of principles retrieved along with the specified conditions constitute the explanation of the phenomenon as described in the explanandum. Thus, the pattern of such an explanation would be:

<explanandum asserted as true> because <explanans retrieved by backward chaining>.

Within the range of DN explanations, a distinction must be drawn between the epistemic and ontic modes of explanation. The epistemic mode employs the sort of reasoning, captured in rules, that an expert or scientist would use in solving a problem or producing an explanation. The ontic mode concentrates on the scientific principles and laws which, if true, would produce a sound deductive argument with the explanandum as the conclusion. Taking a liberal approach to both scientific explanation and expert systems, it will be assumed that the epistemic mode can be considered to be the locus of explanatory activities. Thus, the explanation produced by the operation of an epistemic system will provide the reasons for asserting that the explanandum is true.

Accepting the epistemic mode as primary suggests that the operation of an expert system itself can be taken as an instance of explanation. However, more is required since what is often at issue is either why certain rules are used or why there are such rules at all. A simple approach to this can be taken by adding the idea that rules themselves are often linked to some backing that explains the rule. Though this is only a small deviation from the epistemic mode of DN explanation, it would provide for richer, more informative, explanations.

Other explanatory patterns require a greater divergence from the DN model.

In some cases the focal point of explanation is why one member of a particular kind behaved in a way that the other members of that kind did not. For example, one might want to explain why two parts of the same kind came to be in different states. Here it seems reasonable to think that an explanation is provided by a list of the property differences between the two instances. The pattern of such an explanation is unlike the DN pattern since the focus is difference and not deduction. The pattern of such an explanation would be:

<differences in instances> because <differences in kinds>.

In other cases the focal point of the explanation is a temporal or causal account of a how an object came to be in a particular state. Such explanations are akin to historical explanations in which one is searching for a significant event that leads to a particular outcome. Such explanations are unlike DN explanations since the focus is temporal and not logical. The pattern of such an explanation would be:

<state of object> because <chain of events>.

The variety of explanations in the context of scientific reasoning strongly suggests that there will be more than one explanation pattern. In turn this suggests that in the pragmatics of explanation attention must be paid to determining what sort of explanation is desired.

WHY QUESTIONS AND THE VARIETIES OF EXPLANATION

One can request an explanation in many different ways. One might ask 'Why does parameter P have value V?' or one might ask 'How is it that the system is now asking this question?' The former although asked as a why question is the province of the how? facility and the latter, though asked as a how question is the province of the why?

facility. For the sake of clarity and convenience I will assume that all requests for explanation can be represented as why questions, and suggest that all appropriate requests for explanation embody contrast why questions, 'Why is it this rather than that?'

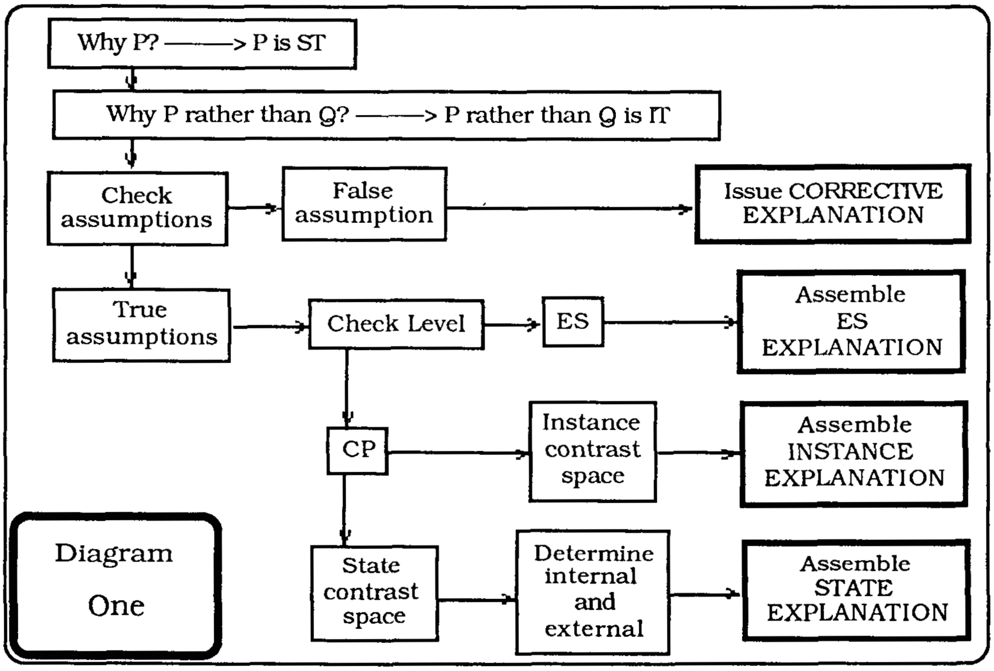
In the simple question 'Why P?' P represents the surface topic (ST) of the question and its assumed that P is true. However, if a context is not invoked, such questions are ambiguous. [10, 1] The question 'Why did part-234 fail?' is ambiguous. At the level of the expert system (ES), it might be a request for an explanation of why the system inferred that part-234 failed rather than asking for a test to be performed on part-234. At the level of the causal process (CP), it might be a request for an explanation of why part-234 rather than part-123 failed, or a request for an explanation of why part-234 failed rather than continued to operate. Thus, why questions should be understood as making reference to some contrast class. 'Why P?' is a specialization of 'Why P rather than Q?' when the contrast class is already understood. In the more general form 'P rather than Q' is the intended topic (IT) of the explanation, and it is assumed that P is true and Q is false.

The contrast class for the IT must contain at least P and Q, but may contain other propositions. The propositions contained in the contrast class may be exclusive, inclusive or unspecified. If they are exclusive, then showing why P is true, will amount to explaining why Q is false. If they are inclusive, a separate account will be needed to show why Q is false. If they are unspecified, then the question should be treated as if it were an ST question.

The particular items that appear in the contrast space can be generated by examining the instances or causal paths. If the IT refers to things that have instances, then the space would include all the instances of that type. For example, if the IT referred to the contrast of part-234 and part-123 and both parts were of the type philosophator, then the contrast space would be composed of all instances of philosophators. Alternatively, if the IT referred to the state of an object, then the space would be composed of all of the paths through the part. For example, if the IT referred to the failure of part-234 rather than its continued operation, then the contrast space would be composed of all the paths through part-234. It should be noted, however, that the state of a device could be determined by inference or by direct measure. If the state is determined by direct measure and cannot be inferred by the rules of the the ES, then the cause of the part's state will be labeled as internal. If the state is either inferred by the rules or is measured but can be inferred from the rules, then it will be labeled as external.

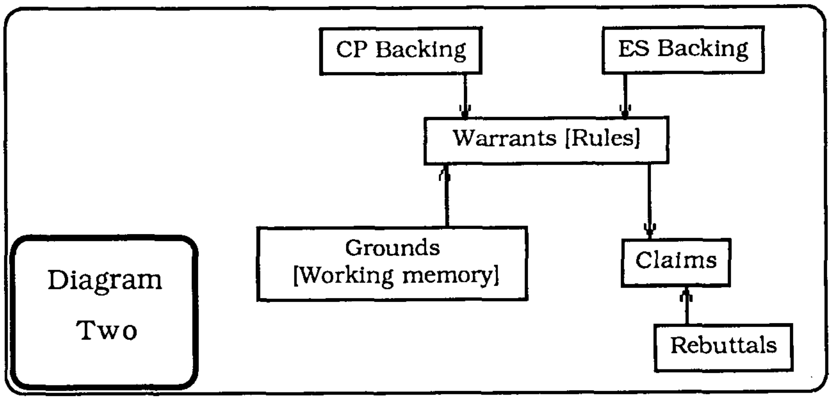
The integration of contrasting why questions with the varieties of explanatory patterns provides an environment in which the user can receive more meaningful explanations. Although the explanations are a bit simple they do respond to the context and provide for multiple explanatory views.

Diagram 1 summarizes the discussion to this point. Simply put the system is designed to have the user refine the ST to an IT. The system then checks for the truth of the assumptions, and if they are false issues a corrective explanation. The user then determines the level of the explanation, ES or CP. If the level is ES, then an ES explanation is assembled. If the level is CP, then the contrast spaces are constructed either by finding the instances of a particular kind of object or by finding all the paths through the object. Appropriate CP explanations are assembled unless the type of contrast is unspecified. If the type of contrast is unspecified, then the system returns for a further clarification of the IT.


Why P? ———> P is ST
Why P rather than Q? ———> P rather than Q is IT
Check assumptions
False assumption
Issue CORRECTIVE EXPLANATION
True assumptions
Check Level
ES
Assemble ES EXPLANATION
CP
Instance contrast space
Assemble INSTANCE EXPLANATION
State contrast space
Determine internal and external
Assemble STATE EXPLANATION
Diagram One

REASONINGS OF GREATER TEXTURE

The final part of the system assembles the explanation. As should be clear from the inclusion of causal paths some modification of the basic representation of knowledge in the system is needed. The pattern of reasoning proposed by Toulmin, Reike and Janik can be understood as providing a pattern that extends the rule based paradigm to provide reasoning of greater 'texture.' [9, 3] In order to use the TRJ model for explanation within the expert system paradigm the basic parts of the model will be interpreted as follows: the grounds are the claims held in working memory, the warrants are the rules, the backings are the support for rules, the rebuttals are a set of rules for alternative outcomes, and the claim is the parameter to be modified. Diagram 2 illustrates this interpretation.



The additional resources of the TRJ model provide an effective way to assemble an explanation. If it is allowed that the rules of the expert system are an operationalized correlate of claims in a model of the system, then one set of backings will provide the details of the model in terms of causal paths. Further, if the expert system allows the creation of instances of types of objects, it should be relatively easy to isolate the type of the object in the consequent of the rule. Moreover, if multiple backings are allowed, another set of backings could establish why a particular rule has been formulated in terms of particular illustrative cases.

The various explanations are assembled using the backings and rebuttals. An ES explanation indicates the rule being used along with its ES backing. Its form is, 'Rule XXX was used to infer P because <backing>'. This basic form is expanded in the case of instance explanation by determining the differences between the conditions of the two objects. Its form is, 'Part-XXX is in state S and part-xxx is not in that state because <differences in properties>'. The state explanations use CP backings to trace through the causal path to a point where an object deviates from its normal state and an internal cause is found. Its form would be 'Part-XXX is in state S because part-xxx entered state s and the path P links Part-XXX to part-xxx'. If the other item of the IT provides an exclusive contrast, then the explanation is finished. If, however, the contrast is inclusive and the other item in IT is in the consequent of the rebuttal, then the form 'and not Q because <rebuttal>' is added.

LINKS TO OTHER MECHANISMS

The proposed simple model of explanation allows for a variety of explanation types, and these types provide links to the efforts of other researchers.

Although Schank examined common sense accounts of explanation, rather than the narrower field of scientific explanation focused upon here, his comments on cognitive understanding are helpful. [5] He suggests that for cognitive understanding, "the program must be able to explain why it came to the conclusions it did, what hypotheses it rejected and why, how previous experiences influenced it to come up with its hypotheses and so on." [p. 15] This notion of cognitive understanding, however, is capable of two related, but distinct, readings. The first reading focuses upon how the program, as a program, came to a conclusion. In this sense the program must be able to explain the steps that it took. This notion seems to be captured in the idea of strategic explanation advanced by Schulman and Hayes-Roth. [6] They consider strategic explanations to be descriptions of the strategic plans and decisions that determine the system's actions. The second reading focuses upon why the program came to the conclusion it did, given the hypotheses (theories) and evidence represented in it. Suthers' examination of the view appropriate to the expert seems to capture this reading. [7] He suggests that experts would expect programs to give summaries of case evaluations in the fields terminology supplemented with accounts of its reasoning and use of evidence.

The two readings of Schank's account of cognitive understanding are complementary and not competitive. The proposed simple model of explanation indicates a way in which the strengths of each can be combined to produce a robust framework. [4] Strategic explanations provide the detail needed to construct corrective and ES explanations, and the views appropriate to the expert indicate the mechanisms required to construct instance and state explanations.

CONCLUSION

A preliminary prototype of a simple explanation system was constructed by Blake Ragsdell (University of Louisville) and Lisa Wurzelbacher (Thomas More College). Although the system, based on the idea of storytelling, did not incorporate all of the principles of simple explanation, it did demonstrate the potential of the approach. The system incorporated a hypertext system, an inference engine, and facilities for constructing contrast type explanations.

The continued development of such a system should prove to be valuable. By extending the resources of the expert system paradigm, the knowledge engineer is not forced to learn a new set of skills, and the domain knowledge already acquired by him is not lost. Further, both the beginning user and the more advanced user can be accommodated. For the beginning user, corrective explanations and ES explanations provide facilities for more clearly understanding the way in which the system is functioning. For the more advanced user, the instance and state explanations allow him to focus on the issues at hand.

The simple model of explanation attempts to exploit and show how the why? and how? facilities of the expert system paradigm can be extended by attending to the pragmatics of explanation and adding 'texture' to the ordinary pattern of reasoning in a rule based system.

* An earlier version of this paper was presented at the AAAI Workshop on Explanation (August 1988). I would like to thank the members of the workshop for helpful comments and criticisms.

References

[1] A. Garfinkle, *Forms of Explanation.* New Haven: Yale University Press.

[2] C. Hempel, *Aspects of Scientific Explanation.* New York: The FreePress.

[3] D. Rochowiak, "Expertise and reasoning with possibility" in Proceedings of the Second NASA Conference on Artificial Intelligence for Space Applications (Huntsville, AL).

[4] D. Rochowiak, "Extensibility and completeness: an essay on scientific reasoning.' *The Journal of Speculative Philosophy,* Vol. 2 No.4.

[5] R. Schank, *Explanation Patterns.* Hillsdale, N.J.: Lawrence Erlbaum.

[6] R. Schulman and B. Hayes-Roth, "Plan-based construction of strategic explanations." Knowledge Systems Laboratory Report No. KSL 88- 23; Stanford University.

[7] D. Suthers. "Providing multiple views of reasoning for explanation." Forthcoming in the proceedings of the International Conference on Intelligent Tutoring Systems.

[8] M. Scriven, "Explanations, predictions, and laws" in *Scientific Explanation, Space and Time* (H. Fiegel and G. Maxwell, eds.). Minneapolis: The University of Minnesota Press.

[9] S. Toulmin, R. Rieke and A. Janik, *An Introduction to Reasoning.* New York: Macmillan.

[10] B. van Fraassen, *The Scientific Image.* Oxford: The Clarendon Press.