TOWARD A SCIENCE OF EXPERT SYSTEMS

Eric Horvitz, Stanford University

ABSTRACT

Over the last several years, teams working on expert systems have been exploring formal approaches for belief revision and information acquisition. The formalization of major components of expert systems operation is useful for understanding and characterizing system behavior and for predicting changes with modification. Formalization also facilitates the involvement of investigators in more well-developed disciplines such as statistics. While the use of formal methodologies for diagnostic problem solving is attractive because of the generality, power, and axiomatic basis of inference, the methodologies have been criticized for making inferences that are difficult to understand and explain. I shall focus on the problem of explaining formal reasoning methodologies. The PATHFINDER system for pathology diagnosis is presented as an example of current research on aspects of the use of formal methodologies in expert systems. I will demonstrate that a formal system is amenable to controlled degradation to enhance its explanation capability.

1. INTRODUCTION

It is fitting that there be a focus of discussion on expert systems in a session on computers and medical decision-making. Original ground-breaking research on expert systems was the result of attempts to build systems to reason about complex medical problems [4]. Expert systems research developed within the field of artificial intelligence over a decade ago and is now an established engineering sub-discipline of artificial intelligence. It is the intent of expert system research to develop methodologies for the representation and manipulation of the knowledge of experts in a variety of disciplines.

Artificial intelligence research is still in its youth. As in other new disciplines in which unifying theories have not been developed, much work has focused on non-axiomatic, descriptive models. In this paper, I would like to briefly introduce the descriptive and formal approaches to research in artificial intelligence in general. I will stress the usefulness of reasoning methodologies that follow from a set of well-characterized axioms. I will then introduce current problems with the use of formal systems. One frequent criticism of formal reasoning strategies is that they are difficult to understand and explain. I will focus on the problem of explanation in expert systems that use formal methods for reasoning under uncertainty. In this regard, I will present research on the PATHFINDER expert system for pathology diagnosis as an example of research on aspects of the use of formal methodologies in expert systems. In answer to some complaints about the rigidity and unnatural nature of formal systems, I shall describe how a formal system is amenable to controlled degradation so that it can perform more descriptively.

2. AXIOMATIC AND DESCRIPTIVE APPROACHES

Science has been marked by an ongoing attempt to explain observed patterns and relationships with models that provide reasonable explanations and predictability. Useful theories tend to simplify phenomena through explaining complexity with a relatively small number of empirically or intuitively justifiable properties or axioms.

Unfortunately, theories based on a set of justifiable axioms often do not exist; when a theory is enumerated, it is often not obviously optimal, unique, or desirable. Throughout the history of science, when useful axiomatic theories have not been available, scientists have resorted to *descriptive* models. Such models summarize complex behavior by *describing* phenomenology without resorting to fundamental axioms. They capture the behavior of systems, often through the postulation of relations that may be inconsistent with one another or with other accepted knowledge. As an example, before Newton constructed the theory of universal gravitation and Kepler developed equations describing the motion of objects orbiting in gravitational fields, astronomers often depended on epicycle machines. These machines could approximately *describe* the movement of heavenly bodies, as viewed from the earth, with a complex tangle of gears and chains. They did not explain the movement of heavenly bodies with a consistent theory of fundamental relationships.

2.1 Descriptive Expert Systems Research

Much of expert systems research can be characterized as either axiomatic or descriptive. The descriptive expert system approach centers on the design and empirical evaluation of algorithms that mimic aspects of human behavior. Descriptive expert systems research is not hindered by the lack of a formal axiomatic basis; it is the intent of the research to discover useful strategies for representing and manipulating expert knowledge regardless of the availability or acceptability of a set of self-consistent desiderata. Investigators in the descriptive school of research view exploration of the sufficiency of informal models of human problem solving as a more direct approach to difficult problems. That is, given poor understanding, many expert systems researchers attempt to capture expertise through building and experimenting with descriptive models in the spirit of the epicycle machines of long ago.

As an example of the descriptive approach to expert system design, the Present Illness Program (PIP) [23], developed ten years ago at M.I.T., was an attempt to simulate the cognition of a physician's reasoning about patients presenting with edema (swelling). A central aspect of the design of the system involved an analysis of the behavior of the clinician. Final versions of PIP had descriptive cognitive structures called the *supervisory program*, the *short-term memory*, and *long-term memory* were constructed.

A large category of descriptive systems is based on the *rule-based* methodology [4]. The rule-based expert system methodology is the result of attempts to adapt the use of an automated logical inference methodology, called *production systems* [32, 7], to capture aspects of human expertise. Production systems are comprised of sets of logically interacting inference rules of the form IF E THEN H, where H is a hypothesis and E is evidence having relevance to the hypothesis. In practice, rules of logical inference are used in automated deduction. For example, *modus ponens* and simple rules of *unification* can be applied to a set or *knowledge base* of rules to do proofs that consist of the forward or backward "chaining" of rules.

One of the most prolific early expert systems was MYCIN [31], a rule-based expert system for the diagnosis of bacterial infection. The MYCIN reasoning framework remains one of the most popular expert system methodologies. MYCIN's knowledge is stored as rules that capture the relationships among relevant medical evidence and hypotheses. For example, a rule in MYCIN might be: "*if an organism infecting a patient is gram-positive and grows in clumps then add support to the hypothesis that the organism is staphylococcus.*" It was recognized early on in the MYCIN research that straightforward application of the production rule methodology would be insufficient because of the uncertainty in the relationships between evidence and hypotheses in medicine.

In order to accommodate these non-deterministic relationships, MYCIN uses certainty factors [4]. To each rule, a certainty factor is attached which represents the *change* in belief about a hypothesis given some evidence. Certainty factors range between -1 and 1. Positive numbers correspond to an *increase* in belief in a hypothesis while negative quantities correspond to a *decrease* in belief. An *ad hoc* calculus for evidence combination was presented in the original research [30].

2.2 The Axiomatic Approach

In contrast to the descriptive approach, investigators pursuing the formal axiomatic approach are interested in exploring the adequacy of systems that satisfy desired properties. That is, they design expert systems that are necessarily *consistent* with desired properties. When such a set is deemed *optimal* for reasoning in the context of particular tasks it is termed a *normative* theory for reasoning.

Investigators interested in the formal approach attempt to design expert systems that behave consistently with established theories for *reasoning under uncertainty*. In exploring the automation of reasoning under uncertainty, investigators have focused on the use of theories for the consistent *revision of belief* in the context of previous belief and for controlling *information acquisition*. Examples of axiomatic theories that have been used in expert systems research for belief revision include probability [24], fuzzy logic [39], Dempster-Shafer theory [28], certainty factors [30], and multi-valued logics [13]. Theories used for controlling information acquisition include information theory [29] and decision theory [25, 26].

Alternative formalisms are often based on clear sets of properties. An expert system engineer can base an expert system on a set of properties that is viewed to be a particularly intuitive or desired. For example, a set of simple properties about continuous measures of belief can be shown to necessitate the use of probability theory to manage the consistent assignment of belief [6, 36, 20]. Agreement with the properties *necessitates* the use of probability theory. A small set of intuitive properties also lies at the foundation of decision theory [37]. Of course, there are differences of opinion among the formalists about the optimality or necessity of particular sets of axioms. For example, there has been ongoing debate in the artificial intelligence community regarding the alternative methodologies for the revision of belief [5, 20].

To date, there have been several attempts to base expert reasoning systems on well-defined formalisms. Three examples are the Acute Renal Failure [15] system, the MEDAS [1] system for emergency medicine, and the PATHFINDER [17] system for lymphoma diagnosis. These systems were designed to be consistent with well-understood formalisms for reasoning.

Both the descriptive and axiomatic approaches have led to the construction of systems that perform at levels rivaling experts in a variety of domains. Given the complexity of problems at hand and the youth of the field, both approaches have been useful in exploring techniques for automated reasoning. In general there has been a healthy interplay between the the descriptive and the axiomatic research; a dynamic research milieu is created by the co-existing approaches.

3. THE BENEFITS OF FORMALIZATION

A worthy fundamental goal of research should be the eventual development of useful *theories*. As in any science, the study of automated reasoning would benefit greatly from attempts to construct theories for representing and manipulating knowledge. Whether an investigator initially chooses to become involved with descriptive or formal research, a fundamental goal should be the construction of a formal science. A strong theoretical basis for components of expert reasoning systems would be extremely useful. While there have already been strides in the application of formal theories to expert systems, greater understanding could facilitate the design, control, and characterization of expert systems.

The subscription to axiomatic bases for components of expert reasoning can be useful in a number of ways. It can assure a system engineer that the behavior of his system will remain consistent with a set of desired properties. Basing a system on a formal theory also ensures that the system will be self-consistent. If an axiomatic theory is not used in building an expert system, it can be quite difficult to maintain self-consistency. The presence of inconsistencies in complex computer systems often leads to unpredictable behavior.

Recent research on the *ad hoc* certainty factor model used for combining evidence in the MYCIN system introduced above has found the original model to be self-inconsistent [16, 18]. Recent work has focused on removing inconsistencies in the model [16]. The consistent reformulation of certainty factors demonstrates that the belief revision theory is a *specialization* of probability in that assumptions of conditional independence are imposed by the methodology. For example, it can be shown that evidence must be conditionally independent given H and its negation [16]. The determination of inconsistency and the detection of constraints were facilitated by the formalization of MYCIN's reasoning strategies.

Formal models can also assist an engineer greatly when a system is modified. A formal system allows for the crisp prediction of changes in system behavior in response to system modifications. It can be quite difficult to predict the impact of modifications on systems for which no underlying theoretical structure is available. Having the ability to control the effect of system modifications is extremely important for the maintenance of systems, for the generalization of specific successes, and for the incremental refinement of techniques. Incremental refinement can be particularly significant in the continuing development of a theoretical framework for automated reasoning.

Most relevant for this conference, formalization can also be crucial for expert systems research to benefit from the participation of investigators in other highly-developed disciplines. Issues surrounding descriptive and axiomatic expert systems research are of special relevance in this regard. For example, expert systems research would benefit if it could attract statisticians to assist in solving difficult problems. Formal descriptions of systems and methodologies are important as they provide conceptual handles necessary for communication with researchers in other fields.

4. PROBLEMS WITH THE FORMAL APPROACH

Two central issues that arise in discussions of the axiomatic approach are problems regarding the pragmatics of engineering and computation, as well as explanation.

4.1 Tractability of Engineering and Computation

More so than for any other reason, researchers in artificial intelligence have looked beyond axiomatic-based techniques for complex domains because of the computational overhead of inference and the requirement for large amounts of knowledge. Formal methodologies are viewed as having an insatiable thirst for data and computer processing [8, 34].

4.2 Explanation

Another significant problem cited with respect to formal methodologies is that it is difficult to explain recommendations to users. The explanation of expert systems has been identified as an significant factor in the acceptance of expert systems [35]. In fact, the transparency of reasoning has been cited as a fundamental feature of expert systems, distinguishing them from numerical programs and other kinds of reasoning systems in artificial intelligence [3]. The important role of reasoning transparency in expert systems has made explanation an artificial intelligence research focus.

It has been said that formal methodologies like probability theory and decision analysis lead to unavoidable losses in comprehensibility to expert system users [8, 34]. The manipulation of the equations of conditional probability or decision trees may indeed be quite difficult to succinctly explain. Such difficulties have provoked some of the ongoing work on techniques for justifying the results of formal reasoning strategies [33, 27, 20]. We shall focus more closely on this problem below.

5. GRACEFUL DEGRADATION OF PERFORMANCE

The concerns about problems with explanation, knowledge acquisition and computational tractability of systems based on formalisms for reasoning under uncertainty are valid. Indeed the methodologies demand large amounts of data and computation. Complaints about the opacity of explanations of recommendations are also justified.

Formal methodologies for reasoning under uncertainty have been put forth as general theories. They have not been designed for use in complex reasoning systems that might be dominated by limitations in computational and engineering resources. An interesting and potentially fruitful area for investigation is the development of strategies for modifying formal methodologies to perform under specified constraints. The process of identifying pressing resource limitations followed by an attempt to reformulate theories (deemed optimal in a world with infinite resources) to perform in constrained environments could be more useful than the outright dismissal of the theories. Such techniques could allow an engineer to gracefully degrade a system's performance to reflect diminishing amounts of available engineering or computational resource.

Theories of belief revision and information acquisition have not traditionally been accompanied by tools that allow a well-defined relaxation of restrictions or requirements. It would be productive to develop such methodologies to generate well-characterized trade-offs such as between the accuracy of a recommendation and computation time. Useful approaches to graceful degradation of various aspects of reasoning behavior would make the disagreement with properties of general parent theories clear. The development of strategies for the controlled degradation of reasoning would allow artificial intelligence researchers to continue to build upon the theoretical achievements of more mature disciplines.

We will now turn to an example of the degradation of expert system performance to satisfy constraints on the *complexity of inference*. As we shall see, degrading an optimal reasoning methodology can serve to enhance the explanation capability in an expert system.

6. EXPLAINING COMPLEX REASONING

I would like to demonstrate an example of the decomposition of a complex reasoning methodology. I hope that it may serve as an example of a category of strategies that can help investigators successfully apply axiomatic models. First I will present an information-optimizing reasoning strategy that makes inferences that are difficult to explain. I will then describe how a less efficient but more explainable strategy could be generated.

6.1 The Complexity of Reasoning Under Uncertainty

We have proposed [19] that a central aspect of the difficulty that investigators have had in explaining expert system recommendations is based on the intrinsic complexity of formal reasoning under uncertainty. As often noted, a fundamental difference between simple deduction and more general reasoning under uncertainty is the inference complexity: within a deductive system, any particular path to a conclusion is considered to be a sufficient proof; in contrast, reasoning under uncertainty usually entails the consideration of all paths [5]. Formal theories of belief revision and information acquisition generally involve the parallel consideration of a greater number of propositions than simple logical deduction problems. For example, probabilistic reasoning systems calculate the values of single conditional probabilities to summarize many steps of inference. This complex summarization process, so central in probabilistic inference, has been seen as a problem in expert system understandability [8].

What is the fundamental basis for problems with complexity? Cognitive psychology results can lend insight to this question. Problems associated with the comprehension of complex problems such as the operation of complex reasoning strategies have been a longtime research focus within cognitive psychology [2]. Classic research in this field has demonstrated severe limitations in the ability of humans to consider more than a handful of concepts in the short term [21]. In fact, studies [38] have discovered that humans cannot retain and reason about more than two concepts in an environment with distractions. Such results underscore the need for managing the complexity of expert systems inference.

For humans to successfully understand, plan, prove, and design in environments that are informationally complex, they must devise schemes for decomposing large unwieldy problems into smaller, interrelated sub-problems. I will present our work on the enhancement of explanation through the decomposition of complex formal reasoning. Before presenting the work, I must first describe the hypothetico-deductive architecture of PATHFINDER.

7. THE PATHFINDER PROJECT

PATHFINDER [17] is a hypothetico-deductive expert system for the diagnosis of lymph node pathology based upon the appearance of microscopic features in lymph node tissue. Disease manifestations in lymph node pathology are microscopic *features*. Features are each subdivided into a mutually exclusive and exhaustive list of *values*. Features are evaluated by the selection of a value that reflects the status of the feature in the case being reviewed. We say that the assignment of a value to a feature constitutes a *piece of evidence*. The PATHFINDER system reasons about 80 diseases, considering over 500 pieces of evidence.

7.1 The Hypothetico-Deductive Architecture

The PATHFINDER system is based on the hypothetico-deductive architecture. The hypothetico-deductive method (also referred to as the method of *sequential diagnosis* [14]) has been studied in several expert systems research projects including the Acute Renal Failure [15] system, the INTERNIST-1 [22] system for diagnosis within the field of internal medicine, and the MEDAS [1] system for emergency medicine.

Hypothetico-deductive systems are presented with an initial set of evidence. The initial evidence is used to assign a probabilistic or quasi-probabilistic score to each hypothesis and a list of plausible hypotheses is formulated from the scores. Then, questions are selected which can help decrease the number of hypotheses under consideration. After a user replies to requests for new information, a new set of hypotheses is formulated and the entire process is repeated until a single diagnosis is reached.

The question selection strategies are termed hypothesis-directed in that reasoning strategies operate on the current list of hypotheses under consideration to generate recommendations for additional evidence gathering. Investigators in the INTERNIST-1 and PATHFINDER research groups have explored the usefulness of tailoring different reasoning strategies to the current list of diseases under consideration or *differential diagnosis*. For example, the strategy selected to narrow the differential diagnosis may depend upon the number of diseases on the differential, the probability distribution over the differential, or both.

The advice generated by hypothesis-directed strategies is often difficult to explain because of the complexity of their operation. This is especially true if recommendations are the result of inferences based on a large hypothesis list. Hypothesis-directed strategies may consider the relevance of hundreds of hypotheses in a single inference step.

The scoring scheme employed by PATHFINDER is based upon the theory of subjective probability [9]. The subjective probabilities of experts are used to infer the probability that each disease is responsible for the evidence that has been entered into the system. Depending on the number and the distribution of probabilities among diseases on the differential diagnosis, PATHFINDER chooses one of several alternative diagnostic strategies for selecting questions. As in other hypothesis-directed systems, it is the goal of the question selection strategies to suggest the optimal test to be evaluated next in an effort to reduce the uncertainty in the differential diagnosis.

Several PATHFINDER strategies discriminate among large numbers of diseases and features in the generation of advice. I shall not describe all of the hypothesis-directed reasoning strategies used by PATHFINDER. Rather, we will look at issues surrounding the explanation of a particular PATHFINDER hypothesis-directed reasoning strategy termed *entropy-discriminate* and its descendant, *group-discriminate*.

7.2 A Strategy to Minimize Uncertainty

The PATHFINDER *entropy-discriminate* reasoning strategy was originally used to refine differential diagnosis disease lists ranging in size from two to eighty diseases. The strategy makes recommendations about information acquisition by searching for tests that maximize a measure of information contained in the differential diagnosis. Similar information-maximizing strategies have been examined in the MEDAS and Acute Renal Failure systems.

Entropy-discriminate makes use of a measure of information known as *relative-entropy*. In this context, relative entropy is a measure of the additional information provided by a piece of evidence E_i about a differential diagnosis DD. Formally,

$$H(DD,E_i) = \Sigma_j\ p(D_j|E_i)\ \log[p(D_j)/p(D_j|E_i)],$$

where $p(D_j)$ is the probability that disease D_j is present before evidence E_i is known, the *prior* probability of the disease, and $p(D_j|E_i)$ is the probability that disease D_j is present after evidence E_i is known, the *posterior* probability of the disease. For a justification of relative entropy as a measure of information gain, see [29].

As each feature consists of a set of mutually exclusive and exhaustive values, we can denote the possible evidence associated with a particular feature, F, as $E_1..E_n$, where n is the number of mutually exclusive values associated with the feature. Entropy-discriminate selects features which give the highest expected relative entropy

$$\langle H(DD,F_n)\rangle = \Sigma_i\ p(E_i)\ H(DD,E_i),$$

where the quantity is summed over feature values $E_1..E_n$, and $p(E_i)$ is calculated using the expansion rule

$$p(E_i) = \Sigma_j\ p(E_i|D_j)\ p(D_j).$$

In an information-theoretic sense, the questions selected by the entropy-discriminate strategy are *optimal* assuming that the goal of the pathologist is to reduce uncertainty in the differential as much as possible.

7.3 Problems With the Optimal Strategy

Soon after the implementation of entropy-discriminate mode, we discovered that several expert pathologists, including the expert that provided the system's knowledge, often found that selected questions were difficult to understand when the differential contained more than approximately ten diseases. The entropy-discriminate strategy of selecting questions that best discriminate among *all* diseases on a differential diagnosis often seemed to be too complex for experts. This is not surprising in light of the limitations of human short term memory discussed above.

We also had problems explaining the recommendations of entropy-discriminate whenever there were more than two diseases on the differential. Attempts were made to provide textual and graphical explanations for the powerful strategy's recommendations. One such graphical explanation justified questions by listing, for each disease, the feature value that would most favor the disease. Physicians found such complex summarizations to be difficult to understand.

7.4 The Graceful Decomposition of Diagnostic Problem Solving

The observed problems with the entropy-discriminate strategy stimulated our interest in strategies for simplifying and explaining hypothesis-directed reasoning. We discovered that pathologists often manage the complexity of the diagnostic problem-solving task by reasoning about a very small number of disease categories or groups at any one time. Questions that discriminate among natural groups tend to be proposed.

Specifically, the chief expert pathologist on the PATHFINDER team often imposes a simple two-group discrimination structure on the problem-solving task. As opposed to a strategy of discriminating among all the diseases on the differential, the pathologist's discrimination task at any point in reasoning about a case is constrained to only two groups of diseases. As categories of diseases are ruled out, the particular pairs of groups considered become increasingly specific. For example, if there are benign and malignant diseases on a differential diagnosis, the pathology expert often deems most appropriate those questions that best discriminate between the benign and malignant groups rather than questions that might best discriminate among all of the diseases. If all benign diseases have been ruled out, leaving only primary malignancies and metastatic diseases on the differential diagnosis, the pathologist will attempt to discriminate between the primary malignancy and the metastatic categories.

We found that the expert's diagnostic strategy can be described by the traversal of a hierarchy of disease categories. The problem-solving hierarchy (see Fig. 1) is a binary tree of disease groups. The hierarchy can be used to group the differential diagnosis at various levels of refinement.

It is interesting to note that several previous studies of medical reasoning have identified similar problem-solving hierarchies [10, 11, 12] for managing the complexity of a wide-variety of reasoning tasks.

The discovery of this expert reasoning strategy in lymph node pathology suggested the development of a new question-selection strategy that could discriminate among binary groups of diseases instead of individual diseases. It was hoped that design and application of such a strategy would make explanation clear, as the user would only have to consider the relevance of a recommendation to two groups.

Our attempt to naturally constrain the discriminatory focus of the entropy-discriminate strategy led to a new reasoning strategy we named *group-discriminate*. The group-discriminate strategy selects questions based on their ability to discriminate between the most specific pair of disease categories that account for all diseases on the differential.

For a given differential diagnosis, group-discriminate identifies the most specific grouping possible and then selects questions that best discriminate among groups of diseases. More formally, suppose the differential is split

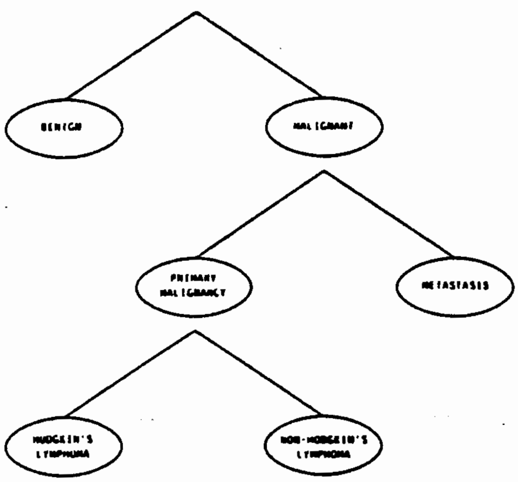

Figure 1: Heuristic problem-solving hierarchy

into two groups, G_1 and G_2, of n_1 and n_2 diseases respectively:

$$G_1 = \{D_{11}, D_{12}, \ldots D_{1n_1}\}$$

$$G_2 = \{D_{21}, D_{22}, \ldots D_{2n_2}\}.$$

As we assume that only one lymph node disease is present in PATHFINDER, we can consider the diseases to be mutually exclusive events. We are interested in the probability that the true diagnosis will be in each group. To calculate this probability we add the probabilities of all the diseases within each group. That is, the probability that a group contains the true diagnosis is

$$p(G_j) = \Sigma_k \, p(D_{j,k}), \quad j = 1, 2.$$

We can also calculate $p(G_j|E_i)$, the probability of the final diagnosis being contained in a group, considering a new piece of evidence E_i. This is

$$p(G_j|E_i) = \Sigma_k \, p(D_{jk}|E_i), \quad j = 1 \text{ or } 2.$$

Therefore, a relative entropy of the grouped differential can be defined. In particular,

$$H_G(DD, E_i) = \Sigma_j \, p(G_j|E_i) \log[p(G_j|E_i)/p(G_j)].$$

This quantity represents the additional information contained in E_i about the grouped differential diagnosis. Group-discriminate selects those features which give the highest expected relative entropy.

Notice that the group-discriminate strategy ignores information concerning the probabilities of diseases within each group. Only the probabilities that the true diagnosis lies within a group is considered in the calculations.

8. DISCUSSION

We integrated the group-discriminate strategy into the PATHFINDER system so that it continues to refine differential diagnosis lists until all diseases remaining on the differential diagnosis are in a category at one of the leaves of the binary problem-solving tree. At this point, other hypothesis-directed strategies are applied to continue pursuing a diagnosis. As the group-discriminate reasoning strategy has a simpler discriminatory focus and more closely follows the decision making protocol of the expert lymph node pathologist than entropy-discriminate, it is quite easy to explain.

Instead of having to present complex summaries explaining how each piece of evidence might impact on belief in the presence of a number of diseases, an explanation of questions generated by group-discriminate must simply demonstrate how possible responses affect the two groups under consideration.

The PATHFINDER system justifies the usefulness of questions selected by group-discriminate with a graphical display. Fig. 2 presents a small portion of a PATHFINDER consultation. At the top of the figure is the differential diagnosis, grouped into benign and malignant categories (at the current level of refinement). Below, several lymph node features recommended by group-discriminate are listed. The group-discriminate strategy has determined that these features can best discriminate between the benign and malignant diseases. In this case, the user requested explanation for the *follicles density* recommendation.

The positions of a set of asterisks in the justification graph at the bottom of the figure are used to indicate the degree to which each group of diseases is favored by each possible feature value. Specifically, the position of an asterisk is a function of the likelihood ratio $p(E_i|G_1)/p(E_i|G_2)$. In the example, the values *separated* and *far apart* strongly support diseases on the differential diagnosis that are in the benign group, while the values *back-to-back* and *closely packed* strongly support the malignant disease hypotheses.

A user can easily ascertain how a question discriminates among two groups of diseases; evidence is either supportive for one group or the other. Even in an environment filled with distractions, the behavior of the strategy is adequately explained by such simple graphs.

Unfortunately, the more explainable group reasoning strategy has some disadvantages. A predictable problem with the use of group-discriminate is that the differential diagnosis refinement process does not always proceed as quickly as it does with the application of the optimal entropy-discriminate. That is, group-discriminate is not as efficient as the more powerful entropy-discriminate; on average, a larger number of evidence-gathering requests will be made by group-discriminate to achieve a similarly refined differential diagnosis. This must be the case as

```
> ask

Discriminating:

 Malignant

    Small cleaved, follicular lymphoma
    Mixed, small cleaved and large cell,
        follicular lymphoma
    Large cell, follicular lymphoma
    Kaposis sarcoma
    Small noncleaved, follicular lymphoma

 Benign

    Florid reactive follicular hyperplasia
    Reactive hyperplasia
    AIDS


I recommend that the following
features be evaluated:

  Status of follicles
  Follicles density
  Subcapsular sinuses
  Medullary sinuses
  Comparison of cytology inside and
       outside the follicles


> justify

 Which feature do you want justified?

> follicles density

The following table elucidates the
discriminating power of this feature.
The position of the asterisk indicates
which of the two groups of diseases is
favored by each value.

 Malignant
         Benign
 |       |
 v       v
*........ back-to-back
*........ closely packed
.......*. separated
........* far apart
```

Figure 2: PATHFINDER consultation

detailed information about the plausibility of individual diseases within each group is discarded in the grouping process.

In general, simplification of an optimal strategy will lead to a less-efficient strategy. Also, given the limits of human cognition identified by research in cognitive psychology, it is not unexpected that a reasoning strategy derived through the constraint or decomposition of a complex problem-solving task may be easier to understand and explain. It seems that for a wide variety of reasoning strategies, there will frequently be an inverse relationship between reasoning understandability and efficiency. In making decisions about alternative reasoning strategies and the clarity of explanation for expert systems, computer scientists may be able to make use of a well-characterized explainability/efficiency trade-off.

9. CONCLUSION

I discussed the usefulness of automated reasoning methodologies that follow from desired fundamental properties and presented an example of the application of a strategy that gracefully degrades complex reasoning of an expert system. The degradation was based in the decomposition of the diagnostic task. The degradation strategy enabled the system to generate transparent justifications for its requests for information, in exchange for a reduction in the optimality of its recommendations.

I believe that continuing research on the pragmatics of applying formal models in the face of severe limitations in data and computation, as well in the abilities of system users will be beneficial. The development and refinement of methodologies for the controlled degradation of reasoning will allow artificial intelligence researchers to build upon the elegant achievements of other disciplines.

Acknowledgements

I am indebted to David Heckerman for many productive conversations. Mr. Heckerman has been an insightful leader of the PATHFINDER Project. I thank Moshe Ben-Bassat, Lawrence Fagan, Ben Grosof, Ted Shortliffe and Peter Szolovits for interesting discussions. I am grateful to Bharat Nathwani and Costa Berard for sharing with me their thoughts on problem solving in pathology. This work was supported in part by the Josiah Macy, Jr. Foundation, the Henry J. Kaiser Family Foundation, the Ford Aerospace Corporation, and the SUMEX-AIM Resource under NIH Grant RR-00785.

References

[1] Ben-Bassat, M., et. al.
Pattern-based Interactive Diagnosis of Multiple Disorders: the MEDAS System.
IEEE Transactions on Pattern Analysis and Machine Intelligence 2:148-160, 1980.

[2] J.S. Bruner, J.J. Goodnow, G.A. Austin.
A study of thinking.
Wiley, 1956.

[3] Buchanan, B. G.
Research on Expert Systems.
In J. Hayes, D. Michie, Y. H. Pao (editors), *Machine Intelligence*, pages 269-299. Ellis Howard Ltd., Chichester, England, 1982.

[4] Buchanan, B. G., and Shortliffe, E. H., eds.
Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project.
Addison-Wesley, Reading, Mass., 1984.

[5] Cheeseman, P.
In defense of probability.
In *Proceedings of the Ninth International Joint Conference on Artificial Intelligence.* IJCAI-85, 1985.

[6] Cox, R.
Probability, frequency and reasonable expectation.
American Journal of Physics 14(1):1-13, January-February, 1946.

[7] Davis, R., Buchanan, B., and Shortliffe, E.
Production Rules as a Representation for a Knowledge-Based Consultation Program.
Artificial Intelligence 8:15-45, 1977.

[8] Davis, R.
Consultation, Knowledge Acquisition, and Instruction.
In P. Szolovits (editor), *Artificial Intelligence In Medicine*, . Westview Press, 1982.

[9] de Finetti, B.
Theory of Probability.
Wiley, New York, 1970.

[10] Elstein, A. S., Loupe, M. J., and Erdman, J. G.
An experimental study of medical diagnostic thinking.
Journal of Structural Learning 2:45-53, 1971.

[11] Elstein, A. S.
Clinical Judgment: Psychological research and medical practice.
Science 194:696, November, 1976.

[12] Elstein, A. S., Shulman, L. S., and Sprafka, S. A.
Medical problem solving: An analysis of clinical reasoning.
Harvard University Press, Cambridge, Mass., 1978.

[13] Gaines, B.R.
Fuzzy and probability uncertainty logics.
Information and Control 38:154-169, 1978.

[14] Gorry, G. A., and Barnett, G. O.
Experience with a Model of Sequential Diagnosis.
Computers and Biomedical Research 1:490-507, 1968.

[15] Gorry, G. A., Kassirer, J. P., Essig, A., and Schwartz, W. B.
Decision Analysis as the Basis for Computer-Aided Management of Acute Renal Failure.
American Journal of Medicine 55:473-484, 1973.

[16] Heckerman, D.E.
Probabilistic Interpretations for MYCIN's Certainty Factors.
In *Uncertainty in Artificial Intelligence*, . North Holland, New York, 1986.

[17] Horvitz, E.J., Heckerman D.E., Nathwani, B.N., and Fagan, L.M.
Diagnostic Strategies in the Hypothesis-Directed PATHFINDER System.
In *Proceedings of the First Conference on Artificial Intelligence Applications*, pages 8. Denver, CO, December, 1984.

[18] Horvitz, E. J., and Heckerman, D. E.
The Inconsistent Use of Measures of Certainty in Artificial Intelligence Research.
In *Uncertainty in Artificial Intelligence*, . North Holland, New York, 1986.
Also available as Technical Report No. KSL-85-57, Knowledge Systems Laboratory, Stanford University.

[19] Horvitz, E.J., Heckerman, D.E., Nathwani, B.N., Fagan, L.M.
The use of a heuristic problem-solving hierarchy to facilitate the explanation of hypothesis-directed reasoning.
In *Proceedings of Medinfo*. Medinfo, October, 1986.
Knowledge Systems Lab Technical Report KSL-86-2, Stanford University, 1986.

[20] Horvitz, E. J., Heckerman, D. E., Langlotz, C. P.
A framework for comparing formalisms for plausible reasoning.
In *Proceedings of the AAAI*. AAAI, Morgan Kaufman, Philadelphia, August, 1986.
Knowledge Systems Lab Technical Report KSL-86-25, Stanford University.

[21] Miller, G.A.
The magical number seven, plus or minus two.
Psychological Review 63:81-97, 1956.

[22] Miller, R. A., Pople, H. E., and Myers, J. D.
INTERNIST-1, An Experimental Computer-Based Diagnostic Consultant for General Internal Medicine.
New England Journal of Medicine 307(8):468-476, 1982.

[23] Pauker, S. G., Gorry, G. A., Kassirer, J. P., Schwartz, W. B.
Toward The Simulation Of Clinical Cognition: Taking A Present Illness by Computer.
American Journal of Medicine 60:981-995, 1976.

[24] Pearl, J.
Fusion, propagation, and structuring in Bayesian networks.
1985.
Presented at the Symposium on Complexity of Approximately Solved Problems, Columbia University, 1985.

[25] Pratt, J. W., Raiffa, H., and Schlaifer, R.
Introduction to Statistical Decision Theory (Preliminary Edition).
McGraw-Hill, New York, 1965.

[26] Raiffa, H.
Decision Analysis: Introductory Lectures on Choice Under Uncertainty.
Addison-Wesley, Reading, Mass., 1968.

[27] Reggia, J.A., Perricone, B.T.
Answer Justification in Medical Decision Support Systems Based on Bayesian Classification.
Comp. Biol. Medicine 15(4):161-167, 1985.

[28] Shafer, G.
A Mathematical Theory of Evidence.
Princeton University Press, 1976.

[29] Shore, J.E.
Relative entropy, probabilisitic inference, and AI.
In *Uncertainty in Artificial Intelligence*, . North Holland, 1986.

[30] Shortliffe, E. H. and Buchanan, B. G.
A model of inexact reasoning in medicine.
Mathematical Biosciences 23:351-379, 1975.

[31] Shortliffe, E. H.
Computer-Based Medical Consultations: MYCIN.
Elsevier/North Holland, New York, 1976.

[32] Simon, H.A.
The Theory of Problem Solving.
Information Processing 71:261-277, 1972.

[33] Spiegelhalter, D.J., and Knill-Jones, R.P.
Statistical and knowledge-based approaches to clinical decision-support systems, with an application in gastroenterology.
J. R. Statist. Soc. A 147:35-77, 1984.

[34] Szolovits, P.
Artificial Intelligence in Medicine.
In P. Szolovits (editor), *Artificial Intelligence In Medicine*, . Westview Press, 1982.