Answer Justification in Diagnostic Expert Systems—Part I: Abductive Inference and Its Justification

JAMES A. REGGIA, BARRY T. PERRICONE, DANA S. NAU, AND YUN PENG

Abstract—Answer justification refers to the ability of an expert system to explain how or why it arrived at certain conclusions (such as a patient's differential diagnosis or treatment recommendations). In this paper, we describe an "abductive" inference method suitable for use in medical expert systems. We then demonstrate how this method can support a clinically plausible form of answer justification in functioning expert systems. A companion paper (Part II) provides the technical details of how the answer justification method described in this paper is implemented, and compares it to previous answer justification methods developed during the last several years.

Introduction

EXPERT systems are computer programs that attempt to provide expert-level problem-solving abilities in some application area [1]. In medicine, a large number of experimental expert systems have been developed as decision aids for such tasks as diagnosis, treatment, and prediction of drug interactions [2]. For our purposes, we can consider these systems to be composed of two parts: a knowledge base specific to some area of medicine, and an inference mechanism that can interpret patient data by using information stored in the knowledge base. A wide variety of methods has been used to process the knowledge in these programs [2].

Answer justification refers to the ability of an expert system to explain how or why it arrived at certain conclusions (such as a patient's differential diagnosis). There is a great deal of interest today in developing new tools for answer justification. Surveys of physicians' attitudes have revealed that some sort of answer justification ability is very important for physician acceptance [3]. In addition, answer justification abilities are useful for analyzing errors made by an evolving expert system and for teaching problem-solving concepts with completed expert systems. For these and other reasons, artificial intelligence (AI) researchers in medicine have designated answer justification as a key research goal [4] and this topic was adopted as the subject of a recent AI workshop [5].

In this paper, we introduce a new method for answer justification suitable for use in diagnostic medical expert systems. Our approach is distinguished by its basis in an "abductive" (as opposed to deductive) model of diagnostic inference. We briefly review, in the next section, the underlying model of abductive inference on which our answer justification method is based. The subsequent section presents an example of the categorical and probabilistic aspects of this answer justification method. A companion paper (Part II) explains the technical details of our answer justification method and contrasts it with previous studies on answer justification.

Manuscript received May 18, 1984; revised September 20, 1984. This work was supported in part by the National Institutes of Health under NIH Grant P50 NS 16332-04A1, by Software Architecture and Engineering, Inc. (all computer time), and by a National Science Foundation Presidential Young Investigator Award to D. S. Nau.

J. A. Reggia is with the Department of Neurology, University of Maryland Hospital, Baltimore, MD 21201, and the Department of Computer Science, University of Maryland, College Park, MD 20742.

B. T. Perricone was with the Department of Computer Science, University of Maryland, College Park, MD. He is now with Software Architecture and Engineering, Arlington, VA 22209.

D. S. Nau and Y. Peng are with the Department of Computer Science, University of Maryland, College Park, MD 20742.

The GSC Model of Abductive Inference

In AI, the most commonly used approach for building diagnostic expert systems has been to store medical knowledge as a collection of if/then rules and to make inferences about a patient using *deductive* logic [2]. This approach can be characterized by a simple syllogism: given rule "$A \rightarrow B$" and given fact "A," conclude "B" (modus ponens).

In contrast, empirical studies suggest that the physician's diagnostic reasoning is better characterized as an *abductive* inference process (e.g., [6]–[8]). This approach is summarized by a semantically different syllogism: given rule "$A \rightarrow B$" and given fact "B," conclude "perhaps A." Whereas the "$\rightarrow$" in the deductive syllogism refers to logical implication, in the abductive syllogism it refers to a causal association between A and B, e.g., disease A can cause symptom B. Thus, knowing that symptom B is present, it is plausible to postulate that perhaps disease A is causing it. In deductive inference, a fact and appropriate rule lead to an immediate deduction; in abductive inference, they lead to a hypothesis ("perhaps A") which must subsequently be confirmed or refuted in a more global context. For this and other reasons, the diagnostic reasoning process of the physician is often referred to as a sequential hypothesize-and-test process [2]. Abductive reasoning is of much more general interest than just diagnostic problem solving, but until recently very little work has been done on developing abductive logics.

We have been studying a new inference method based on modeling the abductive inference process used by physicians in diagnostic problem solving. This model is referred to as the "generalized set covering" (GSC) model, reflecting the fact that it extends previous work on set covering in mathematics [9]. The GSC model is attractive because, in contrast to deductive inference mechanisms, it supports a descriptive knowledge representation (see Fig. 1), and it intrinsically handles situations where multiple disorders are present simultaneously. We have used the GSC model as the basis for a num-

```
COGAN'S SYNDROME  <L>
  [DESCRIPTION:
          DIZZINESS [TYPE = VERTIGO <H>;
              COURSE = ACUTE AND PERSISTENT <H>,
                       EPISODIC <M>,
                       CHRONIC <L>];
          GENERAL SYMPTOMS =  HEADACHE <M>, NEW ONSET OF COUGH <M>,
                       PERSISTENT SHORTNESS OF BREATH <M>;
          NEUROLOGICAL SYMPTOMS = DIPLOPIA <L>, ...;
          TEMPERATURE = ELEVATED <L>;
          HYPEREMIC CONJUNCTIVA <A>;
          NEUROLOGICAL EXAMINATION = ...                  ]
```

Fig. 1. A "description" of Cogan's syndrome. The letters A for "always," H for "high," M for "medium," L for "low," and N for "never" in angular brackets indicate rough, subjective probabilities for the associated statements. The GSC model supports expert systems which use a knowledge base of "descriptions" such as the one shown here. As explained in the text, if d_i = Cogan's syndrome, then this description specifies, among other things, man(d_i).

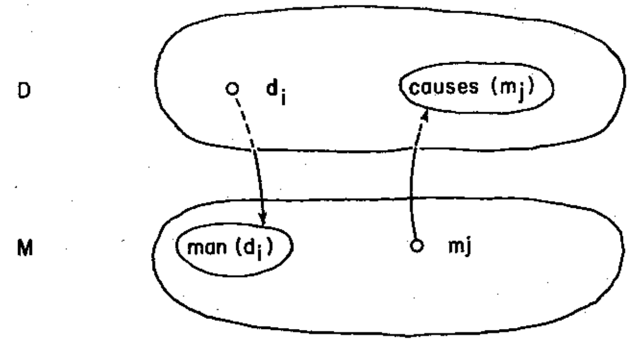


Fig. 2. Organization of diagnostic knowledge.

ber of medical expert systems [10] and we have formalized the underlying concepts mathematically [11]. Our previous work has not addressed how the GSC model might support answer justification abilities. We briefly review the basics of the GSC model below to make the subsequent material on answer justification comprehensible for the reader unfamiliar with our previous work.

In the GSC model, the underlying knowledge for a diagnostic problem is organized as illustrated in Fig. 2. There are two discrete finite sets which define the scope of diagnostic problems: $\boldsymbol{D}$, representing all possible disorders d_i (diseases) that can occur, and $\boldsymbol{M}$, representing all possible manifestations m_j (symptoms and signs) that may occur when one or more disorders are present. A relation $\boldsymbol{C} \subseteq \boldsymbol{D} \times \boldsymbol{M}$ captures the intuitive notion of causation, where $\langle d_i, m_j \rangle \in C$ represents "d_i can cause m_j," the basis of the abductive syllogism described earlier.

With this information, one can designate the set of all possible manifestations of any disorder d_i as

$$\text{man}(d_i) = \{m_j | \langle d_i, m_j \rangle \in \boldsymbol{C}\}$$

and the set of all disorders which can cause any manifestation m_j as

$$\text{causes}(m_j) = \{d_i | \langle d_i, m_j \rangle \in \boldsymbol{C}\}.$$

As shown in Fig. 1, the "description" of a disorder d_i in the knowledge base of an expert system based on the GSC model includes the set man(d_i). This is the basis for the claim that the GSC model supports a descriptive as opposed to rule-based knowledge representation. Furthermore, if man(d_i) is given for each disorder (e.g., in a knowledge base of descriptions such as that in Fig. 1), then causes(m_j) is also implicitly specified for each m_j and represents the differential diagnosis (set of diagnostic possibilities) for that manifestation. For any set D of disorders, man(D) is used to designate $\cup_{d_i \in D}$ man(d_i); and for any set M of manifestations, causes(M) is used to designate $\cup_{m_j \in M}$ causes(m_j).

Finally, there is a distinguished set $M^+ \subseteq \boldsymbol{M}$ which represents those manifestations known to be present for a specific patient. Given the preceding definitions, a diagnostic problem P can thus be designated as a 4-tuple $P = \langle \boldsymbol{D}, \boldsymbol{M}, \boldsymbol{C}, M^+ \rangle$. We will define man$^+(d_i)$ = man$(d_i) \wedge M^+$. Thus, man$^+(d_i)$ is those manifestations of d_i known to be present.

Given M^+, we can define $E \subseteq \boldsymbol{D}$ to be an explanation for M^+ if 1) $M^+ \subseteq$ man(E), or in words: E covers M^+; and 2) $|E| \leq |D'|$ for any other cover D' of M^+, or in words: E is minimal. This definition captures many features of what one means by "explaining" a patient's manifestations. Part 1) specifies the reasonable constraint that a set of disorders E must be able to cause all known manifestations M^+ to be considered to be an explanation for those manifestations. Part 2) specifies that E must be one of the smallest sets to do so. This second constraint reflects the principle of parsimony or *Ockham's Razor:* the simplest explanation is the preferable one. We have thus equated "simplicity" with minimal cardinality, reflecting an underlying assumption that disorders in $\boldsymbol{D}$ are independent of one another.

With these concepts in mind, we define the *solution* to a diagnostic problem to be the set of all explanations for M^+. Rather than represent the solution to a diagnostic problem as a simple list of explanations, it is advantageous (see [10]) to represent these explanations as a set of *generators*. A generator is a collection of sets of "competing" disorders that implicitly represents a set of explanations in the solution to a diagnostic problem and can be used to generate them. A generator is analogous to a Cartesian set product, the difference being that a generator produces explanations which are unordered sets rather than ordered tuples.

As a simple example, suppose that M^+ is as depicted in Fig. 3, and that no single disorder can cover all of M^+. However, pairs of disorders such as $\{d_1, d_7\}$ can cover M^+. Since these are the smallest sets of disorders which can cover M^+, they are explanations. Suppose that there are a total of six explanations for M^+:

$$\{d_1, d_7\} \quad \{d_1, d_8\} \quad \{d_1, d_9\}$$
$$\{d_2, d_7\} \quad \{d_2, d_8\} \quad \{d_2, d_9\}$$

which form the solution. Then, a single generator

$$\{d_1, d_2\} \times \{d_7, d_8, d_9\}$$

can represent this solution. The disorders d_1 and d_2 in the first set of this generator are "competitors" in the sense that they are competing with one another to be part of an explanation for the patient's manifestations; the same holds for d_7, d_8, and d_9.

Expert systems based on the GSC model use a sequential hypothesize-and-test process in solving diagnostic problems. Given one or more manifestations which are present initially, a set of generators representing a *tentative* solution is derived with the assumption that no other manifestations are present.

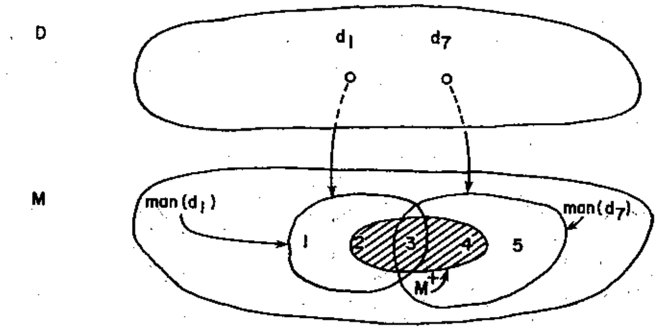


Fig. 3. No single d_i can cover M^+ (shaded region), but explanations such as $\{d_1, d_7\}$ can.

This tentative hypothesis is then used to guide the question generation process. As subsequent questions from the expert system uncover additional manifestations, the hypothesized solution is modified so that it continues to represent all explanations for the patient's manifestations known to be present. These cycles of hypothesizing a solution (set of generators), and then "testing" the hypothesis through further questions to the user, continue until ideally all relevant manifestations have been discovered.

We have previously described a formalized algorithm for performing this sequential diagnostic process, proven its correctness within the GSC model, and established criteria for decomposing diagnostic problems into independent subproblems [11]. Example medical expert systems based on the GSC model have been presented elsewhere, and their development has addressed many practical aspects of implementation such as question generation, criteria for terminating problem solving, and ranking of alternative explanations [10].

An Example of Answer Justification by a GSC-Based Expert System

This section gives an example of a GSC-based medical expert system which supports answer justification abilities. This system provides a differential diagnosis for a patient who complains of dizziness. It will be used to provide concrete examples of the abstract concepts which are subsequently described in Part II.

Dizziness is a very difficult diagnostic problem for the physician because there are numerous potential causes that require knowledge from multiple medical specialties. The knowledge base of the dizziness expert system currently contains descriptions like that in Fig. 1 for about 50 causes of dizziness (see [10]). Such diagnoses include irregular heart rhythms, orthostatic hypotension secondary to drugs (orthostatic hypotension is a fall in blood pressure upon standing up), basilar migraine (painfully dilated blood vessels in the head), inner ear diseases such as ototoxicity from drugs, and hyperventilation (overbreathing associated with anxiety). It is possible that more than one cause of dizziness could be present simultaneously.

In the following, the user's typing is italicized for clarity, and the conversation is annotated with comments enclosed in curly brackets. All other text is generated by the expert system. Ellipses indicate material excised for brevity. Numbers in parentheses in the right margin indicate points in the conversation that will be referenced subsequently in the text. After presenting its differential diagnosis (1), the expert system proceeds to justify the three diseases it hypothesizes are present: basilar migraine (2), an inner ear disorder (3), and a cause of orthostatic hypotension. Along the way [(4) and (5)], it explains why the specific inner ear diseases it lists as possibilities are ranked the way they are in terms of probabilities. Subsequently, it responds to two "why not" propositions from the user [(6) and (7)].

Dizziness Decision Support System

Please describe your patient:

. . .

> {The user describes a 44 year old man with episodic dizziness characterized as "lightheadedness." The patient has also complained of occipital headaches and scintillating scotomas (sparkling spots in front of the eyes) as well as syncope (fainting). The patient's current medications are an aminoglycoside, quinine, and a phenothiazine. Examination reveals orthostatic hypotension (a fall in blood pressure upon standing), homonymous field cut (partial visual loss), rotatory nystagmus (abnormal eye movements), and impaired hearing.}

. . .

OK, based on your description of this patient, this system generates the following differential diagnosis:

basilar migraine ⟨*A*⟩ (1)
and
ototoxicity secondary to quinine ⟨*H*⟩
ototoxicity secondary to aminoglycosides ⟨*H*⟩
otosclerosis ⟨*M*⟩
labyrinthine fistula ⟨*L*⟩
Meniere's disease ⟨*L*⟩
and
orthostatic hypotension secondary to phenothiazines ⟨*H*⟩
idiopathic orthostatic hypotension ⟨*M*⟩
autonomic neuropathy ⟨*L*⟩
multiple system atrophy ⟨*L*⟩.

> {Here, ⟨*A*⟩ means "accepted" or "always," ⟨*H*⟩ indicates high likelihood, ⟨*M*⟩ medium likelihood, and ⟨*L*⟩ low likelihood. This differential diagnosis thus means: "This patient has basilar migraine. In addition, the patient also probably has ototoxicity secondary to the quinine or aminoglycoside he is taking, although he could have otosclerosis or even one of the other unlikely inner ear diseases listed. Finally, the patient also has orthostatic hypotension due to his medications, but this might also be due to one of the other listed causes."}

Ready for command: *justify diagnosis*

> {The user asks that the "value" (i.e., the above differential diagnosis) assigned to the attribute "diagnosis" be justified by the dizziness expert system.}

The value assigned to the attribute 'diagnosis' was inferred by finding all plausible explanations (minimal covers) for the known manifestations.

The value "basilar migraine" is necessary to account for the following manifestations: (2)

general symptoms occurring with the dizziness =
headache,

predominant location = occipital
neurootological symptoms associated with the dizziness = scintillating scotomas
abnormalities on neurological examination = homonymous field cut
duration = transient during dizziness.

This value can also account for the following manifestations but not uniquely:

neurootological symptoms associated with the dizziness = syncope,
dizziness = present,
type = sensation of impending faint,
course = one or more clear-cut episodes,
episode duration = hours,
predominant occurrence = under nonspecific or varying conditions.

{The expert system starts its justification by explaining that basilar migraine is a plausible diagnosis because it explains the patient's transient field cut, occipital headaches, and scintillating scotomas. It also can explain the fainting and dizziness, but so do other disorders postulated to be present.}

One of the following values:

ototoxicity secondary to quinine ⟨*H*⟩ (3)
ototoxicity secondary to aminoglycosides ⟨*H*⟩
otosclerosis ⟨*M*⟩
labyrinthine fistula ⟨*L*⟩
Meniere's disease ⟨*L*⟩

is necessary to account for the following manifestations:

abnormalities on neurological examination = impaired hearing, abnormalities on neurological examination = nystagmus, type = rotatory, duration = transient during dizziness.

{The expert system indicates that it is also necessary to postulate the presence of one of these inner ear disorders to explain the patient's impaired hearing and nystagmus (abnormal eye movements).}

Do you wish to see why these "competing" disorders are ranked the way they are? *yes.*

{The user asks to see an explanation for the relative ordering of the five "competing" inner ear disorders. Whereas the preceding answer justification was concerned with categorically explaining why certain disorders were plausible, the following is concerned with explaining why these plausible "competing" disorders are ranked probabilistically as they are.}

The disorders: ototoxicity secondary to quinine, and ototoxicity secondary to aminoglycosides received the highest ranking. For ototoxicity secondary to quinine, factors which favor this disorder include the following: (4)

current medications = large amounts of quinine.

This disorder is more likely to cause the following present manifestations than some of its competitors: abnormalities on neurological examination = impaired hearing.

For ototoxicity secondary to aminoglycosides, factors which favor this disorder include: · · ·

{For the two highest ranked inner ear disorders, only factors which *favor* (in a relative sense) their presence are cited. These factors include the fact that this patient is taking the relevant medication in each case, and that these causative disorders are more likely to explain the patient's impaired hearing than some of their competitors.}

The disorder otosclerosis received an intermediate ranking. Factors which favor this disorder include: · · · Factors which are against this disorder include: · · ·

{For disorders of intermediate likelihood, such as otosclerosis here, factors that both *favor* (presence of impaired hearing) and are *against* (absence of expected hearing loss, tinnitus, etc.) each disorder are presented.}

The disorders labyrinthine fistula and Meniere's disease received the lowest ranking. For labyrinthine fistula, factors which are against this disorder include: · · ·

This disorder is less likely to cause the following present manifestations than some of its competitors: (5)

abnormalities on neurological examination = impaired hearing.

Expected manifestations which are not present:

neurological symptoms associated with the dizziness =
hearing loss by history ⟨*M*⟩
tinnitus ⟨*M*⟩
abnormalities on neurological examination =
nystagmus ⟨*H*⟩,
type = horizontal ⟨*M*⟩
· · ·

For Meniere's disease, factors which are against this disorder include: · · ·

{For the lowest ranked disorders, only factors which are *against* each disorder are cited. In this example, these factors are predominantly expected manifestations which are absent.}

Ready for command: *justify diagnosis ≠ hypoglycemia.* (6)

{The user wishes to know why hypoglycemia is *not* a reasonable part of the differential diagnosis.}

The value "hypoglycemia" of the attribute diagnosis can explain some of the known manifestations but is not in any minimal set of disorders that explains all manifestations.

Do you wish to see a list of the manifestations it does cover? *no.*

Ready for command: *justify diagnosis ≠ Cogan's syndrome.* (7)

{Again, the user asks "why not," this time for Cogan's Syndrome.}

The value "Cogan's syndrome" of the attribute "diagnosis" was first categorically rejected when the following statement in its description was found to be false: hyperemic conjunctiva = present ⟨*A*⟩.

{Since hyperemic conjunctiva ("bloodshot" eyes) is listed as *always* being present in the description of Cogan's syndrome (see Fig. 1), and since this patient does *not* have hyperemic conjunctiva, a diagnosis of Cogan's syndrome was excluded from further consideration.}

Discussion

From a medical perspective, the justifications provided by the dizziness expert system illustrated above have appeared to us and to other physicians to be quite reasonable and intuitively plausible. This is both because the justifications it provides are in terms of clinical associations familiar to the physician, and because its criteria for considering a specific disorder as a possible diagnosis are based on a model of the criteria used by human diagnosticians (i.e., the GSC model). Thus, the dizziness expert system can make intuitively plausible claims such as "I must include disorder d in my differential diagnosis to account for symptoms x, y, and z."

It should be emphasized that the answer justification method we are describing is domain independent and not restricted to the problem area of dizziness diagnosis. In fact, the dizziness expert system was implemented using a domain-independent knowledge engineering programming environment [10]. Substitution of a different descriptive knowledge base (e.g., for diagnosis of chemical spills) would result in a similar conversation about a different class of diagnostic problems without requiring a change to the underlying programs.

As the reader may have noted, there are two aspects of the answer justification produced by the dizziness expert system: a "categorical" rationale of why certain disorders are included as plausible diagnostic possibilities, and an optional "probabilistic" rationale for ranking competing disorders. We shall follow this pattern in Part II by explaining how the underlying answer justification mechanism works.

REFERENCES

[1] D. S. Nau, "Expert computer systems: A tutorial," *Computer,* vol. 16. New York: IEEE, 1983, pp. 63–85.

[2] J. A. Reggia, "Computer-assisted medical decision making," in *Applications of Computers in Medicine,* M. Schwartz, Ed. New York: IEEE, 1982, pp. 198–213.

[3] R. L. Teach and E. H. Shortliffe, "An analysis of physician attitudes regarding computer-based clinical consultation systems," *Comput. Biomed. Res.*, vol. 14, pp. 542–558, 1981.

[4] H. Pople, 1982 AIM Workshop Rep. (unpublished manuscript).

[5] "Report of the workshop on automated explanation production," *SIGART Newsletter,* vol. 85, pp. 7–13, July 1983.

[6] A. Elstein, L. Shulman, and S. Sprafka, *Medical Problem Solving, An Analysis of Clinical Reasoning.* Cambridge, MA: Harvard University Press, 1978.

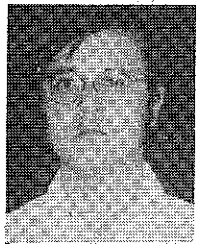
[7] J. Kassirer and G. Gorry, "Clinical problem solving: A behavioral analysis," *Ann. Int. Med.*, vol. 89, pp. 245–255, 1978.

[8] A. Rubin, "The role of hypotheses in medical diagnosis," in *Proc. Int. Joint Conf. Artif. Intell.*, vol. 4, 1975, pp. 856–862.

[9] J. Edwards, "Coverings and packings in a family of sets," *Bull. Amer. Math. Soc.*, vol. 68, pp. 494–499, 1962.

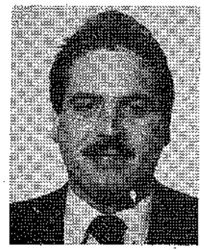
[10] J. A. Reggia, D. S. Nau, and P. Y. Wang, "Diagnostic expert systems based on a set covering model," *Int. J. Man-Machine Studies,* vol. 19, pp. 437–460, 1983.

[11] ——, "A theory of abductive inference in diagnostic expert systems," Dep. Comput. Sci., Univ. Maryland, TR-1338, Dec. 1983, submitted for publication.



James A. Reggia received the M.D. degree from the University of Maryland School of Medicine, Baltimore, in 1975, and the Ph.D. degree in computer science from the University of Maryland, College Park, in 1981.

He is currently jointly appointed as an Associate Professor of Neurology at the University of Maryland Hospital, Baltimore, and Assistant Professor of Computer Science at the University of Maryland, College Park. His research interests include artificial intelligence, computer simulation, decision support systems, and neurolinguistics. He has published many papers in these areas.



Barry T. Perricone received the M.S. degree in computer science from the University of Maryland, College Park, in 1981.

He is currently employed as a Knowledge Engineer at Software Architecture and Engineering, Inc., Arlington, VA. He has coauthored several published papers related to expert system technologies. His current interests include graphics systems, interactive interface design, knowledge-based expert systems, and automated reasoning methodologies.

Mr. Perricone is a member of the Association for Computing Machinery.

Dana S. Nau received the B.S. degree in applied mathematics from the University of Missouri, Rolla, in 1974, and the A.M. and Ph.D. degrees in computer science from Duke University, Durham, NC, in 1976 and 1979, respectively, where he was supported by a National Science Foundation graduate fellowship and a James B. Duke fellowship.

He is an Associate Professor of Computer Science at the University of Maryland, College Park. His previous research experience includes summer appointments at IBM Research in New York, and at the National Bureau of Standards in Maryland. During the latter appointment, he made recommendations for the use of expert computer systems in automated manufacturing. He has published more than 30 papers, mainly on subjects related to artificial intelligence. His current research interests include searching and problem-solving methods in artificial intelligence, expert computer systems, and automated manufacturing.

Dr. Nau is a Recipient of a Presidential Young Investigator Award from the National Science Foundation which will fund his research for the next five years.

Yun Peng was born on April 8, 1946, in Chengdu, Sichuan Province, China. He graduated from the Department of Electric Engineering, Harbin Engineering Institute, China, in August 1970. He received the M.S. degree in computer science from Wayne State University, Detroit, MI, in May 1981. He is currently a Ph.D. candidate in the Department of Computer Science, University of Maryland, College Park.

He worked in data transmission and computer communication for five years at the Electronic Research and Development Institute of the Chinese Ministry of Electronic Industry. His interests include artificial intelligence, logic, and diagnostic expert systems.