Property of
D. Nau



# Answer Justification in Diagnostic Expert Systems—Part II: Supporting Plausible Justifications

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

*Abstract*—**This paper describes how a new method for answer justification in abductive diagnostic expert systems, presented in a preceding companion paper (Part I), can be supported in a domain-independent fashion. Both the issues of explaining why a disorder is included in a differential diagnosis and why it is ranked the way it is relative to its "competitors" are addressed. This approach to answer justification is then compared to previous work on answer justification in medical expert systems.**

## INTRODUCTION

THIS is the second of two companion papers on answer justification in diagnostic expert systems. In Part I, an "abductive" model of diagnostic reasoning was presented. This model is referred to as the generalized set covering (GSC) model because it is based on adopting set covering concepts to simulate diagnostic inference. It was demonstrated that the GSC model can support a plausible form of answer justification.

In this paper, the methods used to support domain-independent answer justification in the context of GSC-directed problem solving are described. The description of these techniques is divided into two sections. In the first of these sections, the issue of why a disorder is or is not included in a differential diagnosis is addressed. In the second of these sections, the issue of why those disorders included in the differential diagnosis are ranked as they are is addressed. An additional final section of this paper then compares the approach to answer justification described here to previous work on this topic.

## CATEGORICAL ASPECTS OF ANSWER JUSTIFICATION

After describing a patient to a diagnostic expert system based on the GSC model, the user is provided with one or more generators representing the "solution" or differential diagnosis (plausible diagnoses under consideration) for that patient. This is illustrated at location (1) in the dizziness conversation of Part I where the user is shown a single generator

$$G_1 = \{d_1\} \times \{d_2, d_3, d_4, d_5, d_6\} \times \{d_7, d_8, d_9, d_{10}, d_{11}\}$$

where $d_1$ = basilar migraine, $d_2$ = ototoxicity secondary to quinine, etc. $G_1$ represents 25 possible explanations in a compact form, such as $\{d_1, d_4, d_{10}\}$ = {basilar migraine, otosclerosis, autonomic neuropathy}, each of which can account for all of the patient's manifestations.

Justifying the "categorical" aspects of a differential diagnosis centers on explaining why a disorder is or is not in the solution to the problem. To understand how this works in the GSC model, it is useful first to consider a single explanation in the solution. The key concept is that the disorders in any given explanation divide $M^+$ into nonempty subsets. Manifestations that lie in a region of $M^+$ which is covered solely by one disorder in the explanation provide a reason for why that disorder must be present: it is necessary to account for those manifestations. (For example, Fig. 3 in Part I shows $M^+$ divided into three regions labeled 2, 3, and 4; manifestations in regions 2 and 4 could be used to justify why $d_1$ and $d_7$ must both be present.)

This concept can be generalized to generators by making the following definition. Let $G_I = g_1 \times g_2 \times \cdots \times g_n$ be a generator in the solution to a diagnostic problem. Then, for some $g_i$ in $G_I$, define

$$\text{common}(g_i) = \bigcap_{d \in gi} \text{man}(d)$$

and let $\text{common}^+(g_i) = \text{common}(g_i) \wedge M^+$. The set $\text{common}^+(g_i)$ represents those present manifestations which can be accounted for by any one of the disorders in $g_i$. Fig. 1 illustrates this concept for a generator $G_J = g_1 \times g_2$, where $\text{common}^+(g_1)$ is indicated by regions 2 and 3.

In the preceding paragraphs, we have represented each generator using the notation $G_I = g_1 \times g_2 \times \cdots \times g_n$, where each $g_i$ is a set of "competing" disorders. Mathematically, a generator $G_I$ is represented as a set $G_I = \{g_1, g_2, \cdots, g_n\}$. The set of all sets of disorders $D$ generated by $G_I$ is designated as $[G_I]$. For a diagnostic problem $P = \langle \boldsymbol{D}, \boldsymbol{M}, \boldsymbol{C}, M^+ \rangle$, let $\text{Sol}(P)$ represent the solution of $P$ as defined in Part I. The following results are readily established.

*Proposition 1:* For a diagnostic problem $P$, let $G_I = \{g_1, g_2, \cdots, g_n\}$ be a generator such that $[G_I] \subseteq \text{Sol}(P)$. Then,

1) $M^+ = \bigcup_{g_i \in G_I} \text{common}^+(g_i)$; and

2) $\forall g_i \in G_I$, $\text{common}^+(g_i) \neq \varnothing$.

*Proof:*

1) This assertion is proved by showing that each side of the equality is a subset of the other.

By the definition of $\text{common}^+(g_i)$, for any $g_i \in G_I$, it holds that $\text{common}^+(g_i) \subseteq M^+$, so $\bigcup_{g_i \in G_I} \text{common}^+(g_i) \subseteq M^+$.

Manuscript received May 18, 1984; revised September 20, 1984. This work was supported in part by the National Institutes of Health under Grant P50 NS 16322-04A1, 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.

0018-9294/85/0400-0268$01.00 © 1985 IEEE

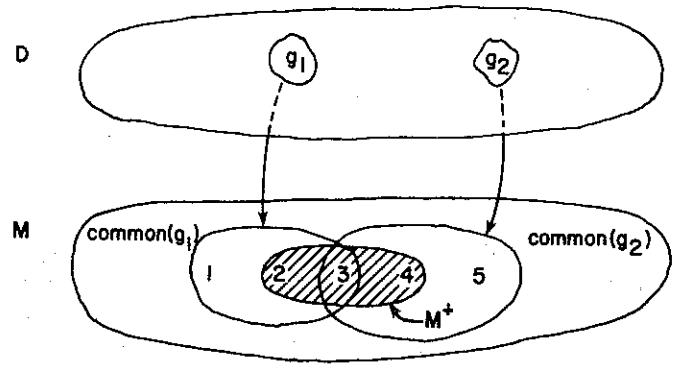


Fig. 1. The sets in a generator $G_I = \{g_1, g_2\}$ divide $M^+$ into common, owned and shared regions (see text for details).

Assume there exists an $m \in M^+$ such that $m \notin \bigcup_{g_i \in G_I} \text{common}^+(g_i)$. Then, $\forall\, g_i \in G_I$, $m \notin \text{common}^+(g_i)$, so it is possible to find a set $D = \{d_i | d_i \in g_i \text{ and } m \notin \text{man}(d_i),\ 1 \leq i \leq n\}$ of $n$ disorders, one from each $g_i$, and none of which cover $m$, so $m \notin \text{man}(D)$. But this contradicts the fact that $D$ is an explanation because $D \in [G_I] \subseteq \text{Sol}(P)$. Thus, our original assumption must be incorrect, and $M^+ \subseteq \bigcup_{g_i \in G_I} \text{common}^+(g_i)$.

2) Suppose $\text{common}^+(g_i) = \varnothing$ for some $g_i \in G_I$. Let $G_I' = G_I - \{g_i\}$, and let $E \in [G_I']$. Since $\text{common}^+(g_i) = \varnothing$, for any $m \in M^+$ there is a $d \in g_i$ such that $m \notin \text{man}(d)$. But since $E \vee \{d\} \in [G_I] \subseteq \text{Sol}(P)$, there must be a $d' \in E$ such that $m \in \text{man}(d')$. Since this holds for every $m \in M^+$, it follows that $M^+ \subseteq \text{man}(E)$. But then for every $D \in [G_I]$, $|E| < |D|$, contradicting the minimality of the explanation $D$. Thus, $\text{common}^+(g_i) \neq \varnothing \forall g_i \in G_I$. •

As illustrated in Fig. 1, it may be that for distinct $g_i$ and $g_j$ in $G_I$ that $\text{common}^+(g_i) \wedge \text{common}^+(g_j) \neq \varnothing$. It is, therefore, convenient to partition $\text{common}^+(g_i)$ into

$$\text{owned}^+(g_i) = \text{common}^+(g_i) - \bigcup_{g_j \neq g_i} \text{common}^+(g_j)$$

and

$$\text{shared}^+(g_i) = \text{common}^+(g_i) - \text{owned}^+(g_i).$$

In Fig. 1, $\text{owned}^+(g_1)$ is indicated by region 2 and $\text{shared}^+(g_1)$ by region 3. While it is possible that $\text{shared}^+(g_i) = \varnothing$ for some $g_i$, in the GSC model it is always the case that the following is true.

*Proposition 2:* For a diagnostic problem $P$, let $G_I = \{g_1, g_2, \cdots, g_n\}$ be a generator such that $[G_I] \subseteq \text{Sol}(P)$. Then, $\forall g_i \in G_I$, $\text{owned}^+(g_i) \neq \varnothing$.

*Proof:* By the definition of $\text{owned}^+(g_i)$, we must show that there is an $m \in \text{common}^+(g_i)$ such that $\forall j \neq i$, $m \notin \text{common}^+(g_j)$.

By 2) above, we know that there is at least one $m \in \text{common}^+(g_i)$. Suppose that $\forall m \in \text{common}^+(g_i)$, there is a $g_j \neq g_i$ such that $m \in \text{common}^+(g_j)$ also. Let $G_I' = G_I - \{g_i\}$ and let $E \in [G_I']$. Then, for every $m' \in M^+$, either of the following holds.

*Case 1:* $m' \in \text{common}^+(g_i)$, so by our supposition above there is a $g_j \in G_I'$ such that $m' \in \text{common}^+(g_j)$.

*Case 2:* $m' \notin \text{common}^+(g_i)$, so by 1) above $m' \in \text{common}^+(g_j)$ for some $g_j \in G_I'$.

Thus, in either case, $m' \in \text{common}^+(g_j)$ for some $g_j \in G_I'$, and so $m' \in \text{man}(E)$. Since this holds for all $m' \in M^+$, $M^+ \subseteq \text{man}(E)$, contradicting the minimality of explanations in $[G_I]$. •

The property that $\text{owned}^+(g_i) \neq \varnothing$ for each $g_i \in G_I$ is essential for the answer justification method described below to work in abductive expert systems such as that illustrated in Part I.

Within this conceptual framework, a categorical justification can be provided for the presence of disorders in a generator $G_I$ forming part of the differential diagnosis by stating the following.

1) That the presence of one of the disorders in $g_i$ is necessary to account for the manifestations in $\text{owned}^+(g_i)$, which, as noted above, is never empty.

2) That any disorders in $g_i$ could account for the manifestations in $\text{shared}^+(g_i)$, although other disorders in $G_I$ not in $g_i$ may also account for these manifestations.

This approach to the categorical aspects of answer justification is illustrated immediately following the points labeled (2) and (3) in the dizziness conversation of Part I. For example, at (3) the dizziness expert system states that one of the listed inner ear diseases (ototoxicity secondary to quinine, etc.) is necessary to account for the patient's impaired hearing and transient rotatory nystagmus (i.e., these two manifestations are "owned" by the listed inner ear diseases, and none is "shared" with other disorders). Clinically, this is a reasonable response to the request for justification.

In implemented expert systems based on the GSC model, the user can also ask for answer justification about a specific diagnosis, such as

justify diagnosis = otosclerosis.

With a suitable change in wording, the same approach described above can explain why "otosclerosis or one or its competitors" (i.e., disorders in the same $g_i$) is necessary in the differential diagnosis.

If the user wants to know why some disorder is *not* in the differential diagnosis, this can be requested by commands such as

justify diagnosis ≠ hypoglycemia.

A rationale for omitting a disorder from the differential diagnosis can be based on either the fact that the disorder is not in a minimal cover [see location (6) in the dizziness conversation in Part I], or that the disorder has been "categorically rejected" (discussed in the next section).

In summary, the theoretical GSC model of diagnostic problem solving is seen readily to support categorical aspects of answer justification. By identifying the manifestations "owned" by a set of competing disorders in a generator/differential diagnosis, clinically plausible reasons for considering the presence of those disorders for a particular patient can be provided.

## Probabilistic Aspects of Answer Justification

The current version of the GSC model as summarized earlier is directed toward categorical aspects of diagnostic problem solving. Probabilistic aspects are addressed in functioning expert systems by superimposing a simple, nonnumeric

weighting scheme on the GSC model. We describe this heuristic scheme here briefly in the context of answer justification, and the interested reader is referred to [10] (in Part I) for further details.

While exact probabilities of diagnostic associations are usually not available in medicine, a great deal of coarse, subjective probabilistic information does exist. Expert systems based on the GSC model represent this useful information as *symbolic probabilities:* $A$ = always, $H$ = high likelihood, $M$ = medium likelihood, $L$ = low likelihood, and $N$ = never. These symbolic probabilities are interpreted in different ways, depending on the context in which they appear. For example, in the description of Cogan's syndrome in the dizziness knowledge base (see Fig. 1, Part I), the "$\langle L \rangle$" following the words "COGAN'S SYNDROME" indicates that this disorder is relatively uncommon, providing a rough approximation to the prior probability $P(d_i)$ used in Bayesian classification ([2] in Part I). The "$\langle L \rangle$" in "TEMPERATURE = ELEVATED $\langle L \rangle$" indicates that Cogan's syndrome only occasionally causes fever, and the "$\langle A \rangle$" in "HYPEREMIC CONJUNCTIVA $\langle A \rangle$" means that when Cogan's syndrome is present, hyperemic conjunctiva (bloodshot eyes) always occur. These latter two symbolic probabilities approximate the conditional probabilities $P(m_j|d_i)$ used in Bayesian classification ([2] in Part I). The user can also indicate uncertainty in describing a patient by using the same set of symbolic probabilities in answering questions generated by an expert system (see [10] in Part I).

During problem solving, the symbolic conditional probabilities $A$ and $N$ are used to determine when any disorder $d_i$ should be *categorically rejected* by the inference mechanism. For example, since the description of Cogan's syndrome (Fig. 1 in Part I) indicates that hyperemic conjunctiva are *always* present (i.e., that $P$(hyperemic conjunctiva|Cogan's syndrome is present) = 1.0), if the user indicates that hyperemic conjunctiva are absent then Cogan's syndrome is immediately discarded from further consideration. In such a situation, the very framework in which the GSC model is functioning is changed (i.e., $\boldsymbol{D}$, the set of all possible disorders, is modified by removing the rejected $d_i$), and the reason for discarding disorder $d_i$ is recorded. This reason can later be retrieved and displayed in response to a "why not $d_i$" request from the user, as is illustrated for Cogan's syndrome at location (7) in the dizziness expert system conversation in Part I. The ability to make categorical rejections as described here is an example of one type of deductive inference that can be made within the framework of the GSC model.

Symbolic probabilities are also used to rank competing disorders in the solution to a diagnostic problem prior to displaying them to the user. Two numerical scores are calculated for each disorder in an explanation. One score, a "setting score," reflects how common the disorder is in the current clinical setting, and is calculated based on problem features which are not manifestations. The second score is a "match score" which reflects how closely the manifestations of the disorder correspond to or "match" those of the patient. These two scores are combined into a single final score which is converted back to a symbolic probability and is then displayed with the differential diagnosis (see location (1) in the conversation in Part I with the dizziness expert system).

The mechanism used to derive setting and match scores in expert systems based on the GSC model is a coarse weighting scheme. The weights involved are the symbolic probabilities associated with the description of each disorder in a knowledge base. The setting score of a disorder $d$ is based on the prior probability of $d$ (e.g., $L$ for Cogan's syndrome) and the setting in which problem solving is occurring. For example, if the description of Cogan's syndrome had included the statement

AGE GT 60 $\langle H \rangle$

indicating that Cogan's syndrome was more likely in the elderly, and the current patient being evaluated by the expert system was 65 year old, then the setting score would be adjusted upward from its initial $L$ level. When statements about the setting, such as that on age immediately above, appear in a description of a disorder $d$, we will refer to them as statements about *nonmanifestations* that are included in $d$'s description.

The match score for a disorder $d$ is calculated based on the conditional symbolic probabilities associated with manifestations listed in $d$'s description (e.g., the $L$ following TEMPERATURE = ELEVATED in the description of Cogan's syndrome). The final score, representing the likelihood of $d$, is based on a normalized product of its setting and match scores. There is a rough correspondence between this approach and Bayesian classification where, for some set of problem features $\boldsymbol{x}$,

$$P(d|\boldsymbol{x}) \propto P(d)P(\boldsymbol{x}|d).$$

Here, $P(d)|\boldsymbol{x})$ corresponds to the final score derived for a disorder $d$, $P(d)$ is a special case of the context-sensitive setting score, and $P(\boldsymbol{x}|d)$ is a special case of the context-sensitive match score ("$\propto$" is read as "is proportional to"). The details of this context-sensitive scoring mechanism as used in expert systems based on the GSC model have been described elsewhere ([10] in Part I).

The relative ranking of two competing disorders in a generator is thus justified by looking at the *differences* between corresponding symbolic probabilities associated with the descriptions of the two disorders. As with the scoring mechanism itself, this analysis involves both the clinical setting and the relevant manifestations. Differences in both the prior probability and the probabilities of nonmanifestations (e.g., the age-related or sex-related risk for a disease) in the descriptions of two disorders form the setting score contribution to justifying the relative ranking of two disorders. Differences in the conditional probabilities of manifestations form the match score contribution to justifying the relative ranking. For example, suppose $d_i$ is ranked higher than $d_j$ in the final differential diagnosis. For any $m \in M^+$, if the symbolic probability of $m$ given $d_i$ is higher than that given $d_j$, then it can be cited as a reason for $d_i$'s higher ranking. Similarly, for manifestation $m'$ *not* in $M^+$, if the symbolic probability of $m'$ given $d_i$ is *lower* than that given $d_j$, $m'$ can also be cited as a reason for $d_i$'s

higher ranking because it is expected to be present "less strongly" when $d_i$ occurs than when $d_j$ occurs.

The specific strategy currently used to justify the relative ranking of disorders is based on the fact that competing disorders, such as the five inner ear diseases in the example conversation earlier

ototoxicity secondary to quinine ⟨$H$⟩
ototoxicity secondary to aminoglycosides ⟨$H$⟩
otosclerosis ⟨$M$⟩
labyrinthine fistula ⟨$L$⟩
meniere's disease ⟨$L$⟩

fall into three groups based on their likelihood: $H$ disorders (most likely), $M$ disorders (possible but less likely), and $L$ disorders (possible, but least likely). For $H$ disorders, a justification of their ranking must explain why they are most likely, so the heuristic justification strategy used cites only factors which favor these disorders. For $M$ disorders, factors which both favor and are against each disorder are cited. Finally, for $L$ disorders, only those factors which are against each disorder are cited to justify why they are ranked lowest. It now remains to be specified, more precisely, what it means for a factor to "favor" or be "against" a disorder.

Factors which *favor* a disorder are those which make the presence of that disorder seem relatively plausible. The following three heuristic criteria are used to identify factors favoring a disorder $d$ relative to its competitors which can be cited in explaining $d$'s ranking.

*1) Prior Probability:* If the prior probability of $d$ is as high as that of all of its competitors and higher than some, then state that $d$ is more common in general than some of its competitors.

*2) Setting:* For each nonmanifestation $S$ in the description of $d$ which

a) has a symbolic probability of $H$ or $A$ specified with it in the description of $d$, and

b) is specified by the user to be present in the current case with a probability of $H$ or $A$, then state that $S$ favors $d$. (This is illustrated in the example conversation on dizziness in Part I immediately following location (4) where "current medications = large amounts of quinine" is cited as a reason that ototoxicity secondary to quinine is ranked highest.)

*3) Present Manifestations:* Let $g_i$ be a set of competing disorders in a generator for the solution to the diagnostic problem and let $d \in g_i$. Then, for each $m \in \text{owned}^+(g_i)$, if the conditional symbolic probability of $m$ given $d$ is as high as that of all its competitors and is higher than some, then state that $m$ is a factor favoring the presence of $d$. (This is also illustrated immediately following location (4) where it is noted that ototoxicity secondary to quinine is more likely to explain the patient's impaired hearing than some of its competitors.)

Analogously, factors which are *against* a disorder are those which make the presence of that disorder seem less likely. The following four criteria are used to identify factors against a disorder $d$ relative to its competitors which can be cited in explaining $d$'s ranking.

*1) Prior Probability:* If the prior probability of $d$ is as low as that of all its competitors and lower than some, then state that $d$ is less common in general than some of its competitors.

*2) Setting:* For each nonmanifestation $S$ in the description of $d$ which

a) has a symbolic probability of $L$ specified with it in the description of $d$, and

b) is specified by the user to be present in the current case with a probability of $H$ or $A$, then state that $S$ is a factor counting against $d$.

*3) Present Manifestations:* Let $g_i$ be a set of competing disorders in a generator for the solution to the diagnostic problem and let $d \in g_i$. Then, for each $m \in \text{owned}^+(g_i)$, if the symbolic probability of $m$ given $d$ is as low as that for all of its competitors and is lower than some, then state that $d$ is less likely to cause $m$ than some of its competitors as a factor counting against $d$. (This is illustrated in the example dizziness conversation at location (5) in Part I where the fact that labyrinthine fistula only occasionally causes impaired hearing is cited as one reason that this disorder is less likely than some of its competitors.)

*4) Absent Manifestations:* For each $m \in \text{man}(d) - M^+$, i.e., for each manifestation in the description of $d$ which is absent, cite the absence of the expected manifestation $m$ as a factor counting against the presence of $d$. (This is illustrated in the example conversation in Part I immediately following location (5) where the absence of tinnitus is given as one of the reasons that labyrinthine fistula is relatively unlikely.)

In summary, expert systems based on our implementation of the GSC model use a simple weighting scheme to rank competing disorders based on the symbolic probabilities in a knowledge base. The structure (generators) imposed on the solution to a diagnostic problem by the GSC model is seen to lend itself readily to justifying the ranking of competing disorders. Criteria have been specified above for identifying the probabilistic factors which count for and against disorders in a differential diagnosis. While these criteria are necessarily heuristic in nature (i.e., not guaranteed to always work), they are at least consistent with empirical studies of how physicians rank competing disorders in a differential diagnosis ([6] in Part I) and with subjective descriptions of the plausible reasons people give for abductive inferences they make [1].

## Discussion

This and the companion paper have described a new method for automated answer justification suitable for use in diagnostic medical expert systems using abductive inference. While other, nonrule-based abductive expert systems exist (e.g., [2]), to the authors' knowledge these systems have not addressed the issue of answer justification in a general and systematic fashion. It is encouraging that the GSC model supports answer justification, in that the GSC model was not originally devised with any conscious attention to providing justifications. The fact that this ability "falls out" of the basic assumptions of the model adds to its attractiveness as a new method for use in medical expert systems.

We conclude this paper by summarizing past research on answer justification in medical expert systems, and by explaining where our approach fits into this earlier work. For our purposes, this previous work can be conveniently characterized as falling into three categories. These categories, sum-

marized below, are distinguished by the "level of machine understanding" at which medical knowledge used for answer justification is represented and processed.

*Precomputed justifications,* or justifications using "canned" text, represent the "shallowest" form in which justification information can be stored and processed. In programs using this approach, the system does not "understand," in any sense, the information it uses or presents to rationalize its actions, but simply retrieves indexed text in appropriate situations. Text retrieval question–answer systems (e.g., [3]), and any program that prints out error messages in response to user actions ("ERROR: YOUR VALUE FOR PATIENT AGE IS OUT OF RANGE"), would fall into this category. A number of AI programs incorporate this ability, which is useful in fairly simple situations, but does not provide for answer justification for questions unanticipated by the system designer. In addition, since in AI systems the knowledge base and textual information may be kept separately, they might be independently changed resulting in justifications inconsistent with inferences actually made by the program.

*Explaining the problem-solving knowledge and activity* used by an expert system provides a somewhat "deeper" approach to answer justification. Answer justification is produced by describing the general inference method that was used (e.g., the procedure followed, calculations performed, or that deduction was used) and/or the specific clinical knowledge that was applied by the program in making decisions. Examples of programs adopting this approach include those which cite a procedurally oriented goal stack [4], state a procedure followed to accomplish a task [5], analyze the probabilities of clinical associations in a statistically oriented knowledge base [6], and maintain a trace of the chain of deductions made during problem solving so that appropriate rules can be produced [7].

This second approach has the advantage that the same clinical knowledge is used for both problem solving and answer justification, assuring that a change to this information is automatically and consistently reflected in both of these activities. However, at times the justifications produced by these systems have appeared quite stilted or even perplexing because they are cast in terms of variables, data structures, or knowledge organizations that do not directly correspond to those familiar to the clinically oriented user. Furthermore, while this approach can explain *what* the expert system has done, it does not really account for *why* the underlying procedures, rules, calculations, etc. which are used are correct.

*Reference to underlying causal mechanisms* is the third and "deepest" method which has been studied. Programs using this approach refer to underlying pathophysiological models to explain why certain actions were taken or inferences made. Examples include expert systems dealing with digitalis therapy [8] and electrolyte disorders [9]. These systems maintain a set of causal associations which are not used directly in problem solving (e.g., "hypercalcemia causes increased cardiac automaticity") as part of their knowledge base and can retrieve them as a rationale for problem solving activities. This approach appears quite promising, but has so far only been applied to relatively small medical domains, is very experimental, and requires potentially large additions to the knowledge base used for problem solving.

At present, it seems reasonable to accept that all of these techniques for answer justification are of value, and that their combination might be the most productive approach to take in future expert systems. The answer justification method described in this paper falls somewhere between the second and third categories above. Although our method is based on citing cause–effect associations to provide a rationale for a diagnosis, these associations are the clinical problem-solving knowledge of the program rather than underlying pathophysiological mechanisms.

Answer justification based on the GSC model is quite attractive in that it provides a relatively intuitive rationale for a differential diagnosis from the viewpoint of the human diagnostician. Citing the causal associations between diseases and the given symptoms to explain why certain diseases are plausible is certainly more satisfying, for example, than citing arbitrary if–then rules. Also, answer justification based on the GSC model does not require supplementing the knowledge base with either free text or pathophysiological models, although either of these approaches could potentially be used to augment justifications.

## References

[1] G. Polya, *Patterns of Plausible Inference.* Princeton, NJ: Princeton University Press, 1954.

[2] R. A. Miller, H. E. Pople, and J. D. Myers, "INTERNIST-I, An experimental computer-based diagnostic consultant for general internal medicine," *New Eng. J. Med.*, vol. 307, pp. 468–476, 1982.

[3] L. Bernstein, E. Siegel, and W. Ford, "The hepatitis knowledge base prototype," in *Proc. 2nd Annu. Symp. Comput. Appl. Med. Care*, Washington, DC, Nov. 1978, pp. 366–367.

[4] T. Winograd, "A procedural model of language understanding," in *Computer Models of Thought and Language*, R. C. Schank and K. M. Colby, Eds. San Francisco, CA: Freeman, 1973, pp. 152–186.

[5] W. Swartout, "A digitalis therapy advisor with explanations," in *Proc. 5th Int. Joint Conf. Artif. Intell.*, 1977, pp. 815–822.

[6] J. A. Reggia and B. T. Perricone, "Answer justification in medical decision support systems based on Bayesian classification," *Comput. Biol. Med.*, to be published.

[7] R. Davis, "Applications of metalevel knowledge to the construction, maintenance and use of large knowledge bases," Memo AIM-283, Stanford AI Laboratory, Stanford, CA, 1976.

[8] W. Swartout, "XPLAIN; A system for creating and explaining expert consulting programs," *Artif. Intell.*, vol. 21, pp. 285–325, 1983.

[9] R. Patil, P. Szolovits, and W. B. Schwartz, "Casual understanding of patient illness in medical diagnosis," in *Proc. 7th Int. Joint Conf. Artif. Intell.*, 1981, pp. 893–899.

**James A. Reggia,** for a photograph and biography, see this issue, p. 267.

**Barry T. Perricone,** for a photograph and biography, see this issue, p. 267.

**Dana S. Nau,** for a photograph and biography, see this issue, p. 267.

**Yun Peng,** for a photograph and biography, see this issue, p. 267.