Dana



Approximate Reasoning in Expert Systems
M.M. Gupta, A. Kandel, W. Bandler, J.B. Kiszka (editors)




# A THEORETICAL FOUNDATION FOR ABDUCTIVE EXPERT SYSTEMS[1]

James A. Reggia[2], Dana S. Nau, Yun Peng, and Barry Perricone[3]

Department of Computer Science
University of Maryland
College Park, MD 20742
USA

This chapter describes the GSC model, a new theoretical framework for diagnostic expert systems which supports approximate reasoning. The model captures several intuitively plausible features of diagnostic inference, handles multiple simultaneous causative disorders, supports sequential problem solving and answer justification, and is readily formalized. This chapter summarizes the GSC model, describes its applications, and addresses a number of issues raised by this work.

Key Words: abductive inference, diagnostic expert systems, generalized set covering, cause-effect reasoning, non-monotonic inference

## INTRODUCTION

A diagnostic problem is a problem in which one is given a set of manifestations (symptoms, signs, laboratory test results) and must explain why those findings are present. Problems of this kind are very common: they include diagnosing a patient's signs and symptoms in medicine, determining why a computer program failed, deciding why an automobile will not start, finding the cause of noises in a plumbing system, localizing a fault in an electronic circuit, explaining why a child makes arithmetic mistakes, etc. This ubiquity has prompted several empirical studies of the underlying reasoning processes involved as well as numerous expert systems for diagnostic problem solving [1,4].

This chapter describes a new theoretical basis for diagnostic expert systems which emulate the reasoning processes used by human diagnosticians. Our model, called the GSC model because it generalizes the set covering problem well known in mathematics, captures several intuitively plausible features of diagnostic inference. It also directly addresses such issues as the handling of multiple simultaneous causative disorders, diagnostic ambiguity, non-monotonic inferences, and sequential problem solving in the context of incomplete information.

This chapter is divided into three main sections. The first section briefly summarizes previous studies of diagnostic reasoning and past work on diagnostic expert systems with an emphasis on how the GSC Model fits into this work. The second section introduces the basic concepts of the GSC Model in an intuitive fashion. The third section discusses applications of the model in expert systems, raises a number of more advanced issues, and describes possible future research directions. Although space limitations prevent us from exploring some of the topics presented here in complete detail, the interested reader is referred to other sources for further information when appropriate.

## DIAGNOSTIC PROBLEM SOLVING

During the last decade, a great deal of effort has gone into obtaining a better understanding of the reasoning processes involved in diagnostic problem solving. As an example, medical diagnostic reasoning, which is presumably typical of diagnostic reasoning in general, has been relatively well studied by cognitive psychologists ([2], [3]; see [4] for a review). This empirical work has shown that diagnostic reasoning involves a sequential hypothesize-and-test process during which the physician conceptually constructs a "model" of the patient. This model or hypothesis postulates the presence of one or more diseases that could explain the patient's manifestations. To construct and modify the hypothesis, the physician relies on his medical "knowledge base," which ideally includes the set of all possible causative diseases for each manifestation, and the set of all possible manifestations for each disease.

The reasoning process involved in diagnostic problem solving is sometimes referred to as abductive inference, and it involves many issues which characterize non-monotonic reasoning in general (see [5]). In particular, as explained below, diagnostic problem solving involves a great deal of "approximate reasoning" (by which we mean reasoning involving several alternative possibilities), and thus the diagnostician's hypothesis may at times be relatively complex. Not only may it contain a great deal of uncertainty about which of several disorders account for a certain manifestation, but it might also presume the simultaneous presence of multiple disorders. The empirical evidence suggests that the hypothesis can best be viewed as a resolution of two conflicting goals:

Coverage Goal: the goal of explaining all of the manifestations that are present;

Parsimony Goal: the goal of minimizing the complexity of the explanation.

The second goal is sometimes referred to as "Occam's Razor."

It is important to appreciate both the sequential and approximate nature of diagnostic reasoning. As the diagnostician gradually learns information about a problem, his hypothesis repeatedly changes to reflect this new information. For example, if a patient complains of sudden onset of chest pain, the physician's initial hypothesis might be something like

HYPOTHESIS 1:

"heart attack, or pulmonary embolus, or . . .".

As further details became available, some of these disorders considered initially might be eliminated. In the above example, if it was next learned that the patient also had a chronic cough and was a heavy smoker, the hypothesis might change to

HYPOTHESIS 2:

"heart attack, or pulmonary embolus, or . . ."
and
"bronchitis, or asthma, or, . . .",

reflecting the physician's belief that at least two diseases must be present to account for this patient's symptoms. At this point the hypothesis contains both uncertainty or approximation (indicated by "or") and the presumption that multiple simultaneous disorders are present (indicated by "and").

| Expert System Method | Theoretical Basis |
|---|---|
| Statistical Pattern Classification | Probability Theory (e.g., Bayes' Theorem) |
| Rule-Based Deduction | Deductive Logic (e.g., first-order predicate calculus) |
| Description-Based Abduction | ? |

Table 1: Three examples of common methods used for implementing diagnostic expert systems along with the corresponding theoretical basis for each method.

Many methods have been used to build expert systems for diagnostic problem solving in the past [4]. Table 1 (left column) lists three prominent examples of such methods. Expert systems using statistical pattern classification have a knowledge base consisting of tables of probabilities, and they make inferences by calculating the probabilities of various diagnostic possibilities once the occurring manifestations are known (e.g., by using Bayes' Theorem). Although statistically-oriented expert systems have produced impressive performance at times, they face a number of theoretical and practical limitations [4]. Not the least of these limitations is the unavailability of the prerequisite probabilities for most real-world diagnostic problems, making knowledge acquisition a difficult task.

Expert systems using the second method listed in Table 1, rule-based deduction, have a knowledge-base of IF-THEN rules and typically use a deductive approach to making inferences [1]. This approach[4] can be characterized by the simple syllogism of modus ponens:

Given Fact "A" and Rule "A$\longrightarrow$B", infer "B".

While rule-based systems have also produced impressive performance at times and are the most widely used AI approach in expert systems today, they too face significant limitations [4]. Knowledge acquisition is again a difficult problem because many diagnostic cues are context-dependent, and because much domain-specific diagnostic knowledge is arranged descriptively and thus is not available a priori in the form of rules.

Expert systems using the third method listed in Table 1, description-based abduction, have a knowledge base of descriptive information and use an abductive approach to making inferences. This approach can be characterized by the simple syllogism

Given Fact "B" and Rule "A$\longrightarrow$B", infer "plausible A."

Whereas the "$\longrightarrow$" in the deductive syllogism above refers to logical implication, that in the abductive syllogism refers to a causal association between A and B. For example, the latter syllogism could be interpreted as "Disorder A is capable of causing manifestation B, and manifestation B is known to be present, so perhaps disorder A is causing it."

Abductive expert systems make the claim to being cognitive models in the following sense. The knowledge base in these systems is typically organized around data structures that cognitive psychologists believe are models of human memory organization (semantic networks, frames, etc.; see [6] for example). Figure 1 shows a "frame" describing a disorder in a simple abductive expert

system. The descriptive organization of diagnostic knowledge in abductive systems is an important factor in easing the knowledge acquisition problem: it permits one to organize information in a textbook-like fashion often familiar to the domain expert. In addition, the inference method in these systems is a sequential hypothesize-and-test process that models the abductive reasoning of the human diagnostician [2,3]. Abductive diagnostic expert systems (e.g., [7,8]) appear to have great potential, but they are the most experimental approach considered here and face several unresolved issues (How should hypotheses be tested? When should problem solving terminate? See [8].)

```
benzenesulfonic acid
   [description:
      pH = acidic;
      appearance = oily;
      detected on spectometry = carbon, sulfur;
      specific gravity = decreased]
```

Figure 1: A frame-like description of the manifestations that could occur from water pollution by benzenesulfonic acid. Taken from a "toy" expert system built with KMS [9] that diagnoses the cause(s) of a chemical spill into a creek downstream from a manufacturing plant. Benzenesulfonic acid is seen to cause five manifestations: acidic pH, oily appearance, detection of carbon by spectometry, detection of sulfur by spectometry, and a decreased specific gravity.

In the context of the above perspective on diagnostic expert systems, the goals of the GSC model are twofold. First, the GSC model is intended to provide a new inference method for use in abductive expert systems. It is an attractive model in that it supports a descriptive knowledge representation that does not require the elicitation of IF-THEN rules from application experts. In addition, the GSC model is context sensitive and supports "approximate reasoning" (reasoning involving several possible alternatives) in the presence of multiple simultaneous disorders. For example, we will see that hypotheses such as HYPOTHESIS 2 above are readily captured as "generators" in the GSC model.

The second purpose of the GSC model is to serve as a theoretical basis for abductive diagnostic expert systems, and as a formal framework for research on diagnostic reasoning in general. Various researchers (e.g., [10]) have criticized previous work on abductive expert systems as being ad hoc because of a lack of such a theoretical underpinning. As Table 1 (right column) suggests, this is in marked contrast to other types of diagnostic expert systems such as those using statistical pattern classification (based on probability theory) or rule-based deduction (based on deductive logic). While recent work on non-monotonic logics provides some concepts important in diagnostic problem solving, these logics are largely deductive in nature [11,12] and do not directly address the issues of coverage and parsimony described earlier. The GSC model attempts to dissipate this criticism of abductive expert system work by replacing the "?" in Table 1 with a formal theory of abductive inference.

## THE BASIC GSC MODEL

We now turn to describing the "basic" GSC model informally to provide an introduction to the fundamental concepts involved. A formal detailed presentation of the theory is available elsewhere for the interested reader [13,14].

In the basic GSC model the underlying knowledge for a diagnostic problem is organized as pictured in Figure 2a. There are two discrete finite sets which define the scope of diagnostic problems: **D**, representing all possible disorders $d_i$ that can occur, and **M**, representing all possible manifestations $m_j$ that may occur when one or more disorders are present. For example, in medicine, **D** might represent all known diseases (or some relevant subset of all diseases), and **M** would then represent all possible symptoms, examination findings, and abnormal laboratory test results that can be caused by diseases in **D**. We will assume that $\mathbf{D} \cap \mathbf{M} = \emptyset$.

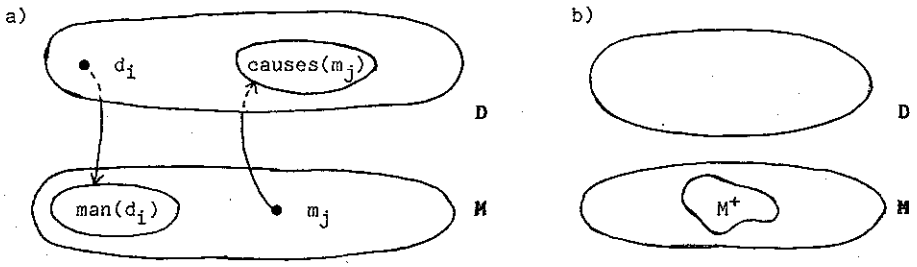


Figure 2: Organization of diagnostic knowledge (a) and problems (b).

To capture the intuitive notion of causation, we assume knowledge of a relation $\mathbf{C} \subseteq \mathbf{D} \times \mathbf{M}$, where $\langle d_i, m_j \rangle \in \mathbf{C}$ represents "$d_i$ can cause $m_j$." Note that $\langle d_i, m_j \rangle \in \bar{\mathbf{C}}$ does not imply that $m_j$ always occurs when $d_i$ is present, but only that $m_j$ may occur. For example, a patient with a heart attack may have chest pain, numbness in the left arm, loss of consciousness, or any of several other symptoms, but none of these symptoms are necessarily present.

Given **D**, **M**, and **C**, the following sets can be defined:

$$man(d_i) = \{m_j \mid \langle d_i, m_j \rangle \in \mathbf{C}\} \qquad \forall\, d_i \in \mathbf{D}, \text{ and}$$
$$causes(m_j) = \{d_i \mid \langle d_i, m_j \rangle \in \mathbf{C}\} \qquad \forall\, m_j \in \mathbf{M}.$$

These sets are depicted in Figure 2a, and represent all possible manifestations caused by $d_i$, and all possible disorders that cause $m_j$, respectively. These concepts are intuitively familiar to the human diagnostician. For example, medial textbooks frequently have descriptions of diseases which include, among other facts, the set $man(d_i)$ for each disease $d_i$. Physicians often refer to the "differential diagnosis" of a symptom, which corresponds to the set $causes(m_j)$. Clearly, if $man(d_i)$ is known for every disorder $d_i$, or if $causes(m_j)$ is known for every manifestation $m_j$, then the causal relation **C** is completely determined. We will use $man(D_I) = \bigcup_{d_i \in D_I} man(d_i)$ and $causes(M_J) = \bigcup_{m_j \in M_J} causes(m_j)$ to indicate all possible manifestations of a set of disorders $D_I$ and all possible causes of any manifestation in $M_J$, respectively.

Finally, there is a distinguished set $M^+ \subseteq \mathbf{M}$ which represents those manifestations which are known to be present (see Figure 2b). Where as D, M, and C are general knowledge about a class of diagnostic problems, $M^+$ represents the manifestations occurring in a specific case.

Using this terminology, we can now make the following definition:

Definition: A diagnostic problem P is a 4-tuple $\langle \mathbf{D},\mathbf{M},\mathbf{C},M^+\rangle$ where these components are as described above.

We will assume in what follows that diagnostic problems are well-formed in the sense that $man(d_i)$ and $causes(m_j)$ are always non-empty sets.

Having characterized a diagnostic problem in these terms, we now turn to defining a solution to a diagnostic problem by first introducing the concept of an explanation.

Definition: For any diagnostic problem P, $E \subseteq D$ is an explanation for $M^+$ if

(i) $M^+ \subseteq man(E)$, or in words: E covers $M^+$; and
(ii) $|E| \leq |D|$ for any other cover D of $M^+$, i.e., E is minimal.

This definition captures what one concept of what one intuitively means by "explaining" the presence of a set of manifestations. Part (i) specifies the reasonable constraint that a set of disorders E must be able to cause all known manifestations $M^+$ in order to be considered an explanation for those manifestations. However, that is not enough: part (ii) specifies that E must also be one of the smallest sets to do so. Part (ii) reflects the Principle of Parsimony or Ockham's Razor described earlier: the simplest explanation is the preferable one. This principle is generally accepted as valid by human diagnosticians. Here, we have equated "simplicity" with minimal cardinality, reflecting an underlying assumption that the occurrence of one disorder $d_i$ is independent of the occurrence of another. (The general nature of "parsimony" obviously involves more than just minimality, an issue to which we will return in the next section. In the meantime, we ask that the reader accept the definition above as a plausible first-approximation to the notion of an explanation.)

With these concepts in mind, we can now define the solution to a diagnostic problem.

Definition: The solution to a diagnostic problem P, designated Sol(P), is the set of all explanations for $M^+$.

The concepts defined above are illustrated in the following example.

Example: Let $P = \langle \mathbf{D},\mathbf{M},\mathbf{C},M^+\rangle$ where $\mathbf{D} = \{d_1,d_2,\ldots,d_9\}$, $M = \{m_1,\ldots,m_6\}$, and $man(d_i)$ are as specified in Table 2. Note that Table 2 implicitly defines the relation C, because $C = \{\langle d_i, m_j\rangle | m_j \in man(d_i) \text{ for some } d_i\}$. Table 2 also implicitly defines $causes(m_j)$ for all $m_j \in \mathbf{M}$. Let $M^+ = \{m_1,m_4,m_5\}$. Note that no single disorder can cover (account for) all of $M^+$, but that some pairs of disorders do cover $M^+$. For instance, if $D = \{d_1,d_7\}$ then $M^+ \subseteq man(D)$. Since there are no covers for $M^+$ of smaller cardinality than D, it follows that D is an explanation for $M^+$. Careful examination of Table 1 should convince the reader that

$$Sol(P) = \{ \{d_1, d_7\}, \{d_1, d_8\}, \{d_1, d_9\}, \{d_2, d_7\}, \{d_2, d_8\}, \{d_2, d_9\}, \{d_3, d_8\}, \{d_4, d_8\} \}$$

is the set of all explanations for $M^+$.

| $d_i$ | man($d_i$) |
|---|---|
| $d_1$ | $m_1$ $m_4$ |
| $d_2$ | $m_1$ $m_3$ $m_4$ |
| $d_3$ | $m_1$ $m_3$ |
| $d_4$ | $m_1$ $m_6$ |
| $d_5$ | $m_2$ $m_3$ $m_4$ |
| $d_6$ | $m_2$ $m_3$ |
| $d_7$ | $m_2$ $m_5$ |
| $d_8$ | $m_4$ $m_5$ $m_6$ |
| $d_9$ | $m_2$ $m_5$ |

Table 2: Knowledge about a class of diagnostic problems. The relation **C** is implicitly defined by this table, as is causes($m_j$) for $1 \leq j \leq 6$.

It is of interest to compare the model of diagnostic problems presented here with the classic set covering problem in mathematics. The set covering problem is typically stated along the following lines:

"For a finite set S of elements and a family F of subsets of S, a cover C of S from F is a subfamily $C \subseteq F$ such that $\bigcup(C) = S$. A cover C is called minimum if its cardinality is as small as possible." [15]

In this definition, S corresponds to $M^+$ and F corresponds to D in the sense that each $d_i \in D$ labels a subset of $M^+$ (the intersection of man($d_i$) with $M^+$). A minimum cover C corresponds roughly to the idea of an explanation E in the GSC model, except man(E) is required only to contain $M^+$ rather than be equal to $M^+$. Thus, even the "basic" GSC model as defined so far is seen to represent two generalizations over the more traditional set covering problem: it does not require an exact covering of $M^+$, and its solution requires finding all covers of $M^+$ rather than a single cover.

Having defined the structure of a diagnostic problem and its solution, we now turn to designing an algorithm that can solve diagnostic problems. The intent of such an algorithm is to model the sequential hypothesize-and-test reasoning process of the human diagnostician. Given an initial set of manifestations, we want our algorithm to construct a tentative hypothesis about the cause of those manifestations, and then to seek further information (e.g., discover additional manifestations) guided by its working hypothesis.

To capture this sequential problem solving paradigm in the context of the GSC Model, the tentative hypothesis at any point during problem solving is defined to be the solution for those manifestations already known to be present, assuming, perhaps falsely, that no additional manifestations will be subsequently discovered.[5] To construct and maintain a tentative hypothesis like this, three simple data structures prove useful:

MANIFS: the set of manifestations known to be present so far;

SCOPE: causes(MANIFS), the set of all disorders $d_i$ for which at least one manifestation is already known to be present; and

FOCUS: the tentative solution for just those manifestations already in MANIFS; FOCUS is represented as a collection of generators.

The term "generator" used here needs further definition. Rather than representing the solution to a diagnostic problem as an explicit list of all possible explanations for $M^+$ or MANIFS, it is advantageous to represent the disorders involved as a collection of explanation generators. An explanation generator is a collection of sets of "competing" disorders that implicitly represent a set of explanations in the solution and can be used to generate them. A generator is analogous to a Cartesian set product, the difference being that the generator produces unordered sets rather than ordered tuples. To illustrate this idea, consider the example diagnostic problem presented earlier (Table 2). Two generators are sufficient to represent the solution to that problem: $\{d_1 , d_2\} \times \{d_7 , d_8 , d_9\}$ and $\{d_3 , d_4\} \times \{d_8\}$. The second generator here implicitly represents the two explanations $\{d_3 , d_8\}$ and $\{d_4 , d_8\}$, while the first generator represents the other six explanations in the solution.

There are at least three advantages to representing the solution to a diagnostic problem as a set of generators. First, this is usually a more compact form of the explanations present in the solution. Second, generators are a very convenient representation for developing algorithms to process explanations sequentially (see below). Finally, and perhaps most important, generators are closer to the way the human diagnostician organizes the possibilities during problem solving (i.e., the "differential diagnosis"). For example, HYPOTHESIS 2, which we used to illustrate the approximate nature of a medical diagnosis involving multiple disorders in a form intuitively plausible to a physician, is a generator.

Using the three data structures MANIFS, SCOPE and FOCUS, a hypothesize-and-test algorithm can be derived to perform diagnostic problem solving in the framework of the GSC model. The FOCUS represents the tentative or working hypothesis at any point during problem solving. The algorithm, described informally, is:

(1) Get the next manifestation $m_j$.
(2) Retrieve causes($m_j$) from the knowledge base.
(3) MANIFS $\leftarrow$ MANIFS $\cup$ $\{m_j\}$.
(4) SCOPE $\leftarrow$ SCOPE $\cup$ causes($m_j$).
(5) Adjust FOCUS to accomodate $m_j$.
(6) Repeat this process until no further manifestations remain.

(The actual algorithm is specified formally in detail along with a proof of its correctness in [14].)

Step 1 in this procedure represents the "test" phase in the "hypothesize-and-test" process, and corresponds to question generation in an expert system. We will ignore for the moment how one selects or discovers the next manifestation, and return to this issue in the next section. For now, we presume that manifestations are discovered one at a time in Step 1, and in Step 2 their corresponding causes are retrieved from the knowledge base.

As each manifestation $m_j$ that is present is discovered, MANIFS is updated simply by adding $m_j$ to it (Step 3). SCOPE is augmented to include any possible causes $d_i$ of $m_j$ which are not already contained in it (derived by taking the union of causes($m_j$) and SCOPE in Step 4). FInally, in Step 5 FOCUS is adjusted to accomodate $m_j$ based on intersecting causes($m_j$) with the sets of disorders in the existing generators. These latter operations are done such that any explanation which can no longer account for the augmented MANIFS (which now includes $m_j$) are eliminated.

The key step in this process is obviously Step 5, the adjustment of the FOCUS or working hypothesis. Perhaps the best way to understand this step is to follow a simple example (see [13] and [14] for formal definition of operations on generators). Recall the abstract knowledge base illustrated in Table 2, and consider the same diagnostic problem $M^+ = \{m_1, m_4, m_5\}$ that was used earlier. The order in which manifestations are discovered to be present during problem solving is not critical: the same solution will ultimately be derived for the same manifestations $M^+$ regardless of the order in which these manifestations are found. We thus suppose in our example that the sequence of events occurring during problem solving were arbitrarily ordered as listed in Table 3. Occurrences during problem solving would be as follows.

| Events in order of their discovery | MANIFS | SCOPE | FOCUS |
|---|---|---|---|
| Initially | $\emptyset$ | $\emptyset$ | $\emptyset$ |
| $m_1$ present | $\{m_1\}$ | $\{d_1\ d_2\ d_3\ d_4\}$ | $\{d_1\ d_2\ d_3\ d_4\}$ |
| $m_2$ absent | " | " | " |
| $m_3$ absent | " | " | " |
| $m_4$ present | $\{m_1\ m_4\}$ | $\{d_1\ d_2\ d_3\ d_4\ d_5\ d_8\}$ | $\{d_1\ d_2\}$ |
| $m_5$ present | $\{m_1\ m_4\ m_5\}$ | $\{d_1\ d_2\ d_3\ d_4\ d_5\ d_7\ d_8\ d_9\}$ | $\{d_1\ d_2\} \times \{d_7\ d_8\ d_9\}$ and $\{d_8\} \times \{d_3\ d_4\}$ |
| $m_6$ absent | " | " | " |

Table 3: Sequential problem solving within the framework of the GSC model.

Initially, MANIFS, SCOPE and FOCUS are all empty. When $m_1$ is discovered to be present, $m_1$ is added to MANIFS, and the new SCOPE is the union of the old SCOPE with causes($m_1$). Since previously there were no generators in the FOCUS, the intersection of causes($m_1$) with them is trivially empty. In such situations a new generator is created, in this case consisting of causes($m_1$). In the terms defined earlier, this generator represents a solution for $M^+ = \{m_1\}$. It tentatively postulates that there are four possible explanations for $M^+$, any one which consists of a single disorder. The FOCUS thus plausibly asserts that "$d_1$ or $d_2$ or $d_3$ or $d_4$ is present."

The absence of $m_2$ and $m_3$ do not change this initial hypothesis. However, when $m_4$ is discovered to be present, MANIFS and SCOPE are augmented appropriately. A new FOCUS is developed, representing the intersection of causes($m_4$) with the single set in the only pre-existing generator in FOCUS. The new generator $\{d_1, d_2\}$ in the FOCUS that results from this intersection operation represents precisely all explanations for the augmented MANIFS. This new FOCUS also illustrates another important point. As information about each possible manifestation becomes available, the FOCUS changes incrementally with a monotonic decrease in the number of explanations it represents (with the exception of situations where the FOCUS becomes empty).

When $m_5$ is noted to be present, MANIFS and SCOPE are again adjusted appropriately. However, in this case the intersection of causes($m_5$) with the single generator in the FOCUS is empty (none of the previous explanations

represented by the old FOCUS can now cover all known manifestations). The occurrence of an empty FOCUS like this again triggers a restructing of the FOCUS: a procedure is called that produces a new set of generators from the now augmented MANIFS and SCOPE. These new generators are based on the fact that the cardinality of any new explanation now contained in the FOCUS must be exactly one greater than the cardinality of its old explanations (for proof, see [13], [14]). Thus, when $m_5$ is found to be present, the new generators represent explanations consisting of two disorders.

Since $m_6$ is absent, the final solution to the problem is given by these same two generators (last line in Table 3). Note that these two generators implicitly represent the eight explanations for $M^+$ that were listed earlier. It is also interesting to note that $d_3$ and $d_4$, eliminated from the FOCUS initially when $m_4$ was found to be present, are once again viable possibilities.

## APPLICATION AND ISSUES

In the interest of clarity, we have so far ignored several aspects of diagnostic problem solving in presenting the GSC model. We now briefly catalog some of these issues, giving an overview of our applications of the model and of a number of more advanced concepts.

### Applications

The GSC model has been used to develop a number of real-world medical expert systems [8]. In addition, it has been applied to a variety of "toy" problems involving non-medical diagnosis (chemical spills, plumbing malfunctions, nuclear reactor monitoring), and even to a number of non-diagnostic tasks (selection of machining operations during process planning, selection of an appropriate statistical test to use for data analysis, designing a model of the growth of biological tree-like structures [21]). The goals of this work have been to provide "proof-of-concept," i.e., to demonstrate that the GSC model works at least reasonably well in the real world, and to uncover its limitations in practice. The paradigm we are using is that theory guides experiment, which in turn guides the subsequent improvement in theory [16].

Knowledge bases for expert systems based on the GSC model consist of a listing of all possible manifestations in **M** and of all disorders in **D** [8,9]. Also included with each disorder $d_i \in \mathbf{D}$ is a textbook-like "description" of $d_i$, analogous to that in Figure 1. Such descriptions specify, among other things, all manifestations which $d_i$ is capable of causing, or $man(d_i)$. Thus the structure of knowledge bases is patterned after that of Table 2, with **C** and $causes(m_j)$ for each $m_j$ being derived from this information. The ability of the GSC Model to support a descriptive knowledge representation, as opposed to requiring a set of rules that specify decision criteria about when a disorder is to be considered present, is one of the model's strengths from the viewpoint of knowledge acquisition.

In addition to using the GSC model as a basis for expert systems, we have applied it as an analytic tool to characterize the behavior of abductive diagnostic expert systems built by others. At times this has produced interesting and unanticipated results. For example, in analyzing the heuristic used to group together competing disorders in INTERNIST-I [7], it was possible to identify situations where this heuristic could lead to a failure to find all of the solution to a diagnostic problem [14]. While the practical implications of this finding remain to be established, it is clearly important to recognize such potential limitations when bulding expert systems for others to use.

Augmenting the Basic Model

Several additional questions arise in applying the GSC model as described above in real-world situations [8]. For example, how a diagnostic expert system should select the next item of information to inquire about in searching for additional manifestations (Step 1 in the informal algorithm described earlier) is not generally obvious. Neither is the issue of when problem solving is complete and can be terminated. There is a tradeoff between maintaining a highly-focused coversation with an expert system user (guided by the current working hypothesis, FOCUS) on the one hand and insuring completeness on the other. In general, we have evolved heuristic approaches to such issues in implemented expert systems [8].

Another question centers on how to rank the alternative explanations in the solution to a diagnostic problem. Since, as we indicated earlier, exact probabilities are not available for most real-world diagnostic associations, expert systems that have been implemented based on the GSC model have adopted a simple symbolic approach to representing probabilities [8]. Among other things, this approach attaches an approximate non-numeric "symbolic probability" to each association $\langle d_i, m_j \rangle \in \mathbf{C}$. These symbolic probabilities provide a coarse, subjective estimate of frequency of causation, prior probability, etc. The five possible symbolic probabilities we have used so far are A = always, H = high likelihood, M = medium likelihood, L = low likelihood, and N = never. This is obviously an arbitrary selection of likelihood estimates but it has been sufficiently robust to produce interesting behavior in our functioning expert systems [8].

Problem Decomposition

Since solving problems in the context of the GSC model is NP hard [14], the task of constructing the solution to a diagnostic problem is potentially computationally expensive as the size of an explanation increases. This difficulty is only academic for some classes of diagnostic problems. For example, it is not uncommon for a patient seen by a physician to have more than one disease simultaneously, but it would be exceedingly rare for someone to have more than 50 diseases simultaneously. However, since the potential for combinatorial explosion exists, it is still important to address the question of when a diagnostic problem can be reduced or decomposed into smaller, independent subproblems.

One example of when this can be done can be explained by introducing the concept of "connected" manifestations. Two manifestations $m_a$ and $m_b$ are said to be connected if either causes($m_a$) and causes($m_b$) have a non-empty intersection, or there exists a finite set of manifestations $\{m_1, m_2, \ldots, m_n\}$ such that $m_1=m_a$, $m_n=m_b$, and each $m_j$ is connected to $m_{j+1}$. All of the manifestations appearing in Table 2, for example, are connected to one another. It can be shown that if $M^+$ can be partitioned into N subsets of connected manifestations, each subset of which contains no manifestation connected to another manifestation in a different subset, then the original diagnostic problem can be partitioned into N independent subproblems. The generators for the solution to the original problem are then easily constructed by appending in an appropriate fashion the generators for the solutions to the subproblems [13].

Furthermore, sequentially constructing and maintaining independent subproblems in this way, each with its own SCOPE, FOCUS and MANIFS, is relatively easy [13]. When a new manifestation $m_i$ is found to be present, the set causes($m_i$) is intersected with the SCOPE of each pre-existing subproblem. When this intersection is non-empty, $m_i$ is said to be related to the corresponding subproblem. There are three possible results of identifying the subproblems to

which $m_i$ is related. First, $m_i$ may not be related to any pre-existing subproblems. In this case, a new subproblem is created, with MANIFS = $\{m_i\}$, SCOPE = causes($m_i$), and FOCUS = a single generator consisting of the single set of competing disorders found in causes($m_i$). This is what always occurs when the first manifestation becomes known, as was illustrated in Table 3. Second, $m_i$ may be related to exactly one subproblem, in which case $m_i$ is assimilated into that subproblem as described earlier and illustrated with $m_4$ and $m_5$ in Table 3. Finally, $m_i$ may be related to multiple existing subproblems. In this situation, these subproblems must be "joined" together to form a new subproblem, and $m_i$ is then assimilated into this new subproblem.

## Answer Justification

"Answer justification" refers to the ability of an expert system to explain how or why it arrived at certain conclusions. In addition to its theoretical importance in AI, potential users of expert systems view answer justification as being of great importance for having confidence in an expert system's performance [17].

In developing the GSC model, no significant thought was initially given to how it might support answer justification. This omission was recently corrected by developing a theoretical rationale for justifying the solution to a diagnostic problem, and by implementing a prototype answer justification program [18]. Our method is based on using the sets of disorders in the generators for the solution to a diagnostic problem to partition $M^+$. This partition then guides the program in explaining why individual disorders are plausible diagnoses for the problem at hand.

The fact that a reasonable approach to answer justification "falls out" of the GSC Model is encouraging. In addition to providing a useful support tool in abductive expert systems, this result avoided what could have been a major setback for the GSC model: failure to find a plausible approach to answer justification in the context of the GSC model would have cast doubt on its validity as a theoretical model of abductive reasoning.

## Conclusion and Future Directions

This chapter has proposed the construction and maintenance of parsimonious set covers ("explanations") as a general model of diagnostic reasoning and described its use as an inference method for diagnostic expert systems. The GSC model is attractive in that it directly handles multiple simultaneous disorders, it can be formalized, it is intuitively plausible, and it is justifiable in terms of past empirical studies of diagnostic reasoning. It also supports answer justification in expert systems.

However, the GSC model is still evolving, both theoretically and in terms of its application in practice. The essential ingredient of the GSC model is the two part constraint on an explanation: coverage and parsimony. In presenting the GSC model in this chapter, we have equated parsimony or simplicity with minimal cardinality. It is clear, however, that there are some situations where a minimal cover would <u>not</u> be the most plausible explanation for a set of manifestations. For example, two very common diseases that can account for all of a patient's symptoms might in some cases be considered a more plausible diagnosis by a physician than one very rare disorder which accounts for all of the symptoms. A minimal cover is therefore to be considered only a first approximation to a plausible explanation, and we are studying alternative formulations of parsimony at the current time, as well as the relationships between rule-based deduction and description-based abduction [19].

Another direction our theoretical research is taking us is the generalization of the GSC Model so that it can be applied to a wider variety of real-world problems [20]. Thus, we are looking at ways to extend the model so that it can capture such things as causal chaining ("A causes B, and B causes C, so A indirectly causes C") and hierarchically-structured diagnostic knowledge. We are continuing to apply the model to a wide range of applications to gain a better understanding of its strengths and weaknesses. The successes we have had so far with this approach to modelling diagnostic inference are clearly sufficient to warrant its further study.

NOTES

[1] Preparation of this paper and the research it describes was supported in part by a grant to the University of Maryland from Software Architecture & Engineering, Inc. and a NSF Presidential Young Investigator Award to Dana Nau.

[2] Also with the Department of Neurology, University of Maryland, Baltimore, MD 21201, USA.

[3] With Software Architecture & Engineering, Inc., Arlington, VA, USA.

[4] We include PROLOG-implemented expert systems in the category of rule-based systems of this sort, for although they use resolution as an inference method, theorems are restricted to Horn clauses.

[5] This "default assumption" and subsequent adjustments when it turns out to be an incorrect assumption illustrate the inherently non-monotonic nature of the abductive inferences made in the GSC model.

REFERENCES

[1] D. Nau. Expert computer systems. Computer, 16, IEEE Press, 1983, 63-85.

[2] A. Elstein, L. Shulman and S. Sprafka. Medical Problem Solving - An Analysis of Clinical Reasoning, Harvard University Press, 1978.

[3] J. Kassirer and G. Gorry: Clinical problem solving - a behavioral analysis, Ann. Int. Med., 89, 1978, 245-255.

[4] J. Reggia. Computer-assisted medical decision making, in Applications of Computers in Medicine, M. Schwartz (ed.), IEEE Press, 1982, 198-213.

[5] R. Reiter. On reasoning by default. Proc. TINLAP II, 1978, 210-218.

[6] J. Anderson. Retrieval of information from long-term memory. Science, 220, 1983, 25-30.

[7] R. Miller, H. Pople and J. Myers. INTERNIST-I, an experimental computer-based diagnostic consultant for general internal medicine. New England Journal of Medicine, 307, 1982, 468-476.

[8] J. Reggia, D. Nau, and P. Wang. Diagnostic expert systems based on a set covering model, Int. J. Man-Machine Studies, 19, 1983, 437-460.

[9] J. Reggia and B. Perricone. The KMS Manual. TR-1136, Department of Computer Science, University of Maryland, 1982.

[10] M. Ben-Bassat, R. Carlson, V. Puri, M. Davenport, J. Schriver, M. Latif, R. Smith, L. Portigal, E. Lipnick, and M. Weil: Pattern-based interactive diagnosis of multiple disorders: The MEDAS system, IEEE Trans. on Pattern Analysis and Machine Intelligence, 2, 1980, 148-160.

[11] D. McDermott and J. Doyle. Non-Monotonic logic I. Artificial Intelligence, 13, 1980, 41-72.

[12] R. Reiter. A logic for default reasoning. Artificial Intelligence, 13, 1980, 81-132.

[13] J. Reggia, D. Nau and P. Wang. A formal model of diagnostic inference. Part I: Problem formulation and decomposition, 1984, submitted for publication.

[14] J. Reggia, D. Nau and P. Wang. A formal model of diagnostic inference. Part II: Algorithmic solution and application, 1984, submitted for publication.

[15] J. Edwards. Coverings and packings in a family of sets, Bull. Am. Math. Soc., 68, 1962, 494-499.

[16] N. Nilsson. The interplay between experimental and theoretical methods in artificial intelligence. Cognition and Brain Theory, 4, 1980, 69-74.

[17] R. Teach and E. Shortliffe. An analysis of physician attitudes regarding computer-based clinical consultation systems. Computers and Biomedical Research, 14, 542-558, 1981.

[18] J. Reggia, B. Perricone, D. Nau and Y. Peng. Answer justification in abductive expert systems for diagnostic problem solving, IEEE Transactions on Biomedical Engineering, 1985, in press.

[19] D. Nau and J. Reggia. Relationships between deductive and abductive inference in knowledge-based diagnostic problem solving, Proc. First Intl. Workshop on Expert Systems, L. Kerschberg (ed), Kiawah Island, SC, Oct. 1984, 500-509.

[20] Y. Peng. A general theoretical model for abductive diagnostic expert systems. TR-1402, Department of Computer Science, University of Maryland, College Park, MD, May 1984.

[21] M. Tagamets and J. Reggia. Abductive Hypothesis Formation and Refinement During Construction of Natural System Models, TR-1463, Department of Computer Science, University of Maryland, Jan. 1985.