INFORMATION SCIENCES **37**, 227–256 (1985)

A Formal Model of Diagnostic Inference. I. Problem Formulation and Decomposition

JAMES A. REGGIA

Departments of Computer Science and Neurology, University of Maryland, College Park, Maryland 20742

DANA S. NAU

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

and

PEARL Y. WANG

Department of Mathematical Sciences, George Mason University, Fairfax, Virginia 22030

ABSTRACT

This paper, which is Part I of a two-part series, introduces a new model of diagnostic problem solving based on a generalization of the set-covering problem. The model formalizes the concepts of (1) whether or not a set of one or more disorders is sufficient to explain a set of occurring manifestations, (2) what a solution is for a diagnostic problem, and (3) how to generate all of the alternative explanations in a problem's solution. In addition, conditions for decomposing a diagnostic problem into independent subproblems are stated and proven. This model is of interest because it captures several intuitively plausible features of human diagnostic inference, it directly addresses the issue of multiple simultaneous causative disorders, it can serve as a theoretical basis for expert systems for diagnostic problem solving, and it provides a conceptual framework within which to view some recent AI work on diagnostic problem solving in general. In Part II, the concepts developed in this paper will be used to present algorithms for diagnostic problem solving.

INTRODUCTION

A diagnostic problem is a problem in which one is given a set of manifestations (symptoms, signs, laboratory 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

52 Vanderbilt Ave., New York, NY 10017

0020-0255/85/$03.30

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 computer models of this reasoning process.

This paper, which is Part I of a two-part series, describes a new model of diagnostic inference that supports a descriptive, object-oriented knowledge representation. This model is called the GSC model, for "generalized set covering," because it is based on a generalization of the set covering problem in mathematics [6]. The GSC model is intended to serve as the theoretical basis for a number of diagnostic expert systems. In addition, it captures several plausible features of human diagnostic inference, it directly addresses the issues of diagnostic context and of multiple simultaneous causative disorders, and it provides a conceptual framework within which to view recent work on diagnostic problem solving in general.

The GSC model answers a criticism sometimes directed at AI expert systems that explicitly model human diagnostic reasoning: that these systems are "ad hoc" (e.g., see [2]) because of their domain-specific nature and absence of a theoretical foundation. We will illustrate in Part II that generalized set covering can provide a theoretical foundation for some of these expert systems. The GSC model thus represents an effort to bring mathematical rigor to an area of AI where it has previously been relatively lacking, and it is an attempt to create an abstraction of expert system implementations in the sense that Nilsson has advocated [16].

Part I (the current paper) begins with a brief review of empirial studies of diagnostic problem solving and diagnostic expert systems. It then presents the simplest GSC formulation of diagnostic problems, formalizing the concepts of explanation, solution, and solution generators. Conditions for decomposing a diagnostic problem into independent subproblems are stated and proven. Part II discusses formalized algorithms for solving diagnostic problems cast within the basic GSC model, contains proofs of their correctness, and shows how the GSC model can be applied to analyze some existing expert systems involving diagnostic inference. More advanced versions of the GSC model are introduced, and ongoing research is outlined.

DIAGNOSTIC REASONING AND DIAGNOSTIC EXPERT SYSTEMS

Diagnostic reasoning has received a great deal of attention over the last few years by cognitive psychologists, AI researchers interested in modeling the underlying thought processes, and educators interested in improving the training of diagnosticians (e.g., [1, 3, 7, 9, 11, 17, 23, 24, 25, 29]). This section first briefly reviews past empirical studies of the diagnostic reasoning process, and then

discusses how models of this process have been implemented as expert systems. Our goal is to provide the reader with sufficient background so that the relationship of the theoretical GSC model to diagnostic reasoning can be appreciated, and to relate the GSC model to previous expert system work in AI.

EMPIRICAL STUDIES OF DIAGNOSTIC REASONING

While a variety of experimental designs have been used in empirical studies of diagnostic reasoning, perhaps the most common and most fruitful approach has been the use of simulated diagnostic problems. A problem-solving session is typically recorded or videotaped, and is followed by "debriefing" of the diagnostician ("Why did you ask this? What were you thinking of here? ..."). Based on these studies, diagnostic reasoning is generally accepted to be a sequential hypothesize-and-test (hypotheticodeductive) process during which the diagnostician conceptually constructs a "model" of the underlying causative disorders (see Figure 1) [1, 7, 9, 11, 17, 23, 25]. This model, or *hypothesis*, is based largely on what manifestations are known to be present (e.g., a patient's symptoms). It postulates the presence of one or more disorders that could explain the given manifestations. Each cycle of the hypotheticodeductive inference process can be viewed as consisting of three phases: disorder evocation, hypothesis evaluation, and question generation. In reality, these three phases overlap extensively.

Disorder evocation is the retrieval from long-term memory of causative disorders as the diagnostician detects a new manifestation in the information available about a problem. This evoking of potential causes for the manifestations begins very early in the diagnostic process and draws on the diagnostician's memory or "knowledge base" of causal associations between disorders and their manifestations. Ideally, the diagnostician's knowledge base includes the set of all possible causative disorders for each manifestation, and the set of all possible

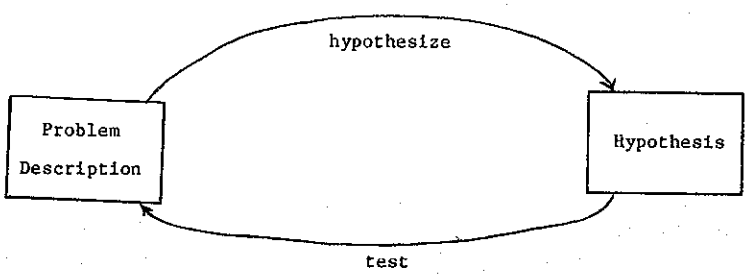

Fig. 1. Hypotheticodeductive inference process.

manifestations for each disorder [29]. Usually a single manifestation (rather than combinations of manifestations) is responsible for evoking new disorders for incorporation into the hypothesis [9].

The second phase of the hypotheticodeductive cycle, hypothesis evaluation, involves the incorporation of the new manifestation into the hypothesis. This may require attributing the manifestation to some disorder already assumed to be present, or adding new disorders evoked by the manifestation to the hypothesis. The diagnostician's hypothesis may at times be relatively complex. Not only may it contain a great deal of uncertainty about which of several diagnoses 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." Teleologically, the parsimony goal can be viewed either as an attempt to focus the reasoning process and therefore restrict searching, as a reflection of the limitations of human short-term memory, or as a "common sense" heuristic that is correct most of the time.

It is important to appreciate the sequential nature of diagnostic reasoning. As the diagnostician gradually learns information about a problem, his or her hypothesis 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 H1.

"Heart attack, *or* pulmonary embolus, *or* ...".

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

HYPOTHESIS H2.

"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. Note that at this point, the hypothesis contains both uncertainty (indicated by *or*) and the presumption that multiple simultaneous disorders are present (indicated by *and*).

Another aspect of hypothesis evaluation is the ranking of the likelihood of competing disorders. The term "competing disorders" refers to hypothesized alternatives which can account for the same or similar manifestations, such as heart attack and pulmonary embolus in Hypothesis H1. Perhaps surprisingly, human diagnosticians appear to use only a three-point weighting scheme to rank competing disorders: a particular finding may be "positive, noncontributory or negative with respect to a particular hypothesis" [7]. At the end of a problem-solving session, diagnosticians are thus able to rank competing disorders only in a very coarse fashion (e.g., *d* is definitely present, *d* is very likely to be present, *d* may be present, *d* is possible but improbable). Most of this ranking can be accounted for by either of two rules: (a) weighting based on counting the number of positive findings, or (b) weighting based on counting the number of positive findings minus the number of expected findings found to be absent [7].

The third phase of the hypotheticodeductive cycle is question generation, and it represents the "test" in "hypothesize-and-test" (see Figure 1). The word "question" here is being used in a very general sense to indicate not only verbal questions, but also any type of information-gathering activity (e.g., measuring a voltage in a malfunctioning circuit). Investigators studying human diagnostic problem solving typically divide such questions into two categories: protocol-driven and hypothesis-driven. Protocol-driven questions are those that a diagnostician generally asks as a routine during any diagnostic session (e.g., "review of systems" in medicine), and these will not be of further concern in this paper. In contrast, hypothesis-driven questions seek information that is specifically needed to modify the evolving hypothesis. Investigators who observe diagnosticians often attribute each hypothesis-driven question to a specific problem-solving strategy: attempting to confirm a hypothesis, attempting to eliminate a hypothesis, or attempting to discriminate between two or more hypotheses.

Many aspects of the diagnostic reasoning process are incompletely understood at the present time. For example, it is unclear how a diagnostician reasons about multiple simultaneous disorders. In such situations the manifestations must be attributed to appropriate disorders and competing disorders must be ranked in the context of other disorders assumed to be present. It is also unclear exactly how diagnosticians decide to terminate the diagnostic process because a "solution" has been reached.

DIAGNOSTIC EXPERT SYSTEMS

A great variety of approaches has been taken in representing and processing knowledge in expert systems for diagnostic problem solving [20]. Table 1 lists three prominent examples of such methods. In systems using statistical pattern

TABLE 1

Some Methods Used in Diagnostic Expert Systems

Method	Examples	Theoretical basis
Statistical pattern classification	[5, 31]	Probability Theory
Rule-based deduction	[10, 22, 26, 30]	First-order predicate calculus
Frame-based abduction	[12, 21, 27]	?

classification, the knowledge base typically consists of tables of probabilities, and the inference mechanism involves the calculation of posterior probabilities of disorders using formulas such as Bayes's theorem. Expert systems of this type have clearly achieved expert-level performance, at times outperforming human diagnosticians [5, 31]. It is important to what follows to note that expert systems using statistical pattern classification have a strong theoretical foundation in probability theory.

Diagnostic expert systems using rule-based deduction typically have a knowledge base consisting of conditional rules and an inference mechanism based on modus ponens or resolution (the latter in expert systems written in PROLOG). As with statistical pattern classification, diagnostic expert systems of this type have clearly been demonstrated to exhibit an expert level of performance in empirical testing [10, 22, 30]. They also have a strong theoretical foundation in first-order predicate calculus.

The third approach to building diagnostic expert systems listed in Table 1 is frame-based abduction. Unlike deductive rule-based systems, whose inferences might in their simplest form be characterized by modus ponens:

Given fact "A" and rule "$A \rightarrow B$", infer "B",

expert systems of this type inherently involve *abductive* inference of the form

Given fact "B" and association "$A \rightarrow B$", infer "plausible A".

Although the "$\rightarrow$" in the deductive syllogism refers to logical implication, in the abductive syllogism as used in diagnostic problem solving it refers to a causal association between A and B: "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 not only aim for a high level of performance, but they also are often explicit attempts to model the underlying reasoning of the

diagnostician. Information in the knowledge base is generally represented in an object-oriented fashion, and a sequential hypothesize-and-test inference process is used (e.g., [12, 21, 27]). The key point here is that frame-based abductive expert systems, in marked contrast to statistical pattern classification and rule-based deduction, do *not* have a readily identifiable well-developed theoretical foundation. It is in large part due to this absence of a theoretical basis that AI research on abductive expert systems is sometimes dismissed as "ad hoc" (e.g., [2]). The basic GSC model described in this paper (and Part II) represents a first attempt to fill this gap.

PROBLEM FORMULATION AND "EXPLANATION"

In many real-world domains, the knowledge base used in diagnostic problem solving is very large, complex, and "ill-structured" (e.g., in medicine [17, 18]). One approach to formalizing the structure and use of this knowledge in expert systems would be to examine the most general case possible. While this is obviously the ultimate goal, as a starting point it introduces a great amount of complexity and detail that obscures the central ideas of the theory being developed here. Thus, in this and the subsequent companion paper, we initially present the simplest meaningful version of the GSC model and discuss its implications. Once this basic formulation is understood, its basic assumptions are reexamined in Part II to explain why more general formulations are necessary. Our efforts to extend the GSC model to progressively more general formulations will be explained at that time.

In the basic form of the GSC model discussed here the underlying knowledge for a diagnostic problem is organized as pictured in Figure 2(a). 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,

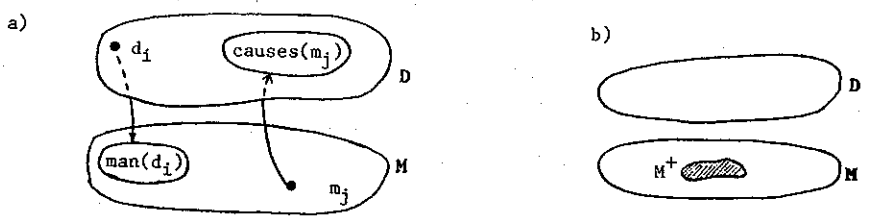

Fig. 2. Organization of diagnostic knowledge (a) and problems (b).

examination findings, and abnormal laboratory results that can be caused by diseases in **D**. We will assume that $\mathbf{D} \cap \mathbf{M} = \varnothing$, that the individual disorders are independent of one another, and that the individual disorders are of relatively equal likelihood in general. We will reexamine these assumptions in Part II and discuss what happens when they do not hold.

Within this basic framework, we make the following definition.[1]

DEFINITION. A *diagnostic problem P* is a 4-tuple $\langle \mathbf{D}, \mathbf{M}, \mathbf{C}, M^+ \rangle$ where

$\mathbf{D} = \{d_1, \ldots, d_n\}$ is a finite set of disorders;
$\mathbf{M} = \{m_1, \ldots, m_k\}$ is a finite set of manifestations;
$\mathbf{C} \subseteq \mathbf{D} \times \mathbf{M}$ is a relation with domain(**C**) = **D** and range(**C**) = **M**; and $M^+ \subseteq \mathbf{M}$ is a distinguished subset of **M**.

In this definition we are using the notation domain(**C**) = $\{d | \langle d, m \rangle \in \mathbf{C}$ for some $m\}$ and range(**C**) = $\{m | \langle d, m \rangle \in \mathbf{C}$ for some $d\}$, implying that $\mathbf{C} = \varnothing$ only if **M** and **D** are empty.

The relation **C** captures the intuitive notion of causation, where $\langle d_i, m_j \rangle \in \mathbf{C}$ means "d_i can cause m_j." Note that $\langle d_i, m_j \rangle \in \mathbf{C}$ does not imply that m_j always occurs when d_i is present, but only that m_j may occur. For example, in medicine 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. Whereas **D**, **M**, and **C** are general knowledge about a class of diagnostic problems, M^+ represents the manifestations occurring in a specific case (i.e., M^+ is those manifestations which are "present"). M^+ is illustrated in Figure 2(b).

Given the above problem formulation, we define the following sets:

DEFINITION.

$$\text{man}(d_i) = \{m_j | \langle d_i, m_j \rangle \in \mathbf{C}\} \qquad \forall d_i \in \mathbf{D},$$

and

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

These sets are depicted in Figure 2(a), 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, medical textbooks frequently have descriptions of diseases which include, among

[1] The notation used in this paper is summarized in Appendix A.

other facts, the set $\text{man}(d_i)$ for each disease d_i. Physicians often refer to the "differential diagnosis" of a symptom, which corresponds to the set $\text{causes}(m_j)$. Clearly, if $\text{man}(d_i)$ is known for every disorder d_i, or if $\text{causes}(m_j)$ is known for every manifestation m_j, then the causal relation C is completely determined, and vice versa. We will use $\text{man}(D) = \cup_{d_i \in D} \text{man}(d_i)$ and $\text{causes}(M) = \cup_{m_j \in M} \text{causes}(m_j)$ to indicate all possible manifestations of a set of disorders D and all possible causes of any manifestation in M, respectively. We will also use the following abbreviations:

$$\text{man}^+(d_i) = \text{man}(d_i) \cap M^+,$$

$$\text{man}^+(D) = \text{man}(D) \cap M^+.$$

Thus, $\text{man}^+(d_i)$ represents that subset of M^+ which could be caused by d_i.

Two lemmas immediately follow from these definitions, and we will repeatedly use them both explicitly and implicitly in subsequent proofs.

LEMMA 1.1. *Let* $P = \langle \mathbf{D}, \mathbf{M}, \mathbf{C}, M^+ \rangle$ *be a diagnostic problem, and let* $d_i \in \mathbf{D}$, $m_j \in \mathbf{M}$, $D \subseteq \mathbf{D}$, $D' \subseteq \mathbf{D}$, *and* $M \subseteq \mathbf{M}$. *Then*

(a) $\text{man}(d_i) \neq \varnothing$,
(b) $d_i \in \text{causes}(\text{man}(d_i))$,
(c) $D \subseteq \text{causes}(\text{man}(D))$,
(d) $\mathbf{D} = \text{causes}(\mathbf{M})$,
(e) $d_i \in \text{causes}(m_j)$ *if and only if* $m_j \in \text{man}(d_i)$,
(f) $\text{man}(D) - \text{man}(D') \subseteq \text{man}(D - D')$.

(a′) $\text{causes}(m_j) \neq \varnothing$,
(b′) $m_j \in \text{man}(\text{causes}(m_j))$,
(c′) $M \subseteq \text{man}(\text{causes}(M))$,
(d′) $\mathbf{M} = \text{man}(\mathbf{D})$,

Proof. All proofs are given in Appendix B.

LEMMA 1.2. *If* $P = \langle \mathbf{D}, \mathbf{M}, \mathbf{C}, M^+ \rangle$ *is a diagnostic problem with* $D \subseteq \mathbf{D}$ *and* $M \subseteq \mathbf{M}$, *then*

$$D \cap \text{causes}(M) = \varnothing \quad \textit{iff} \quad M \cap \text{man}(D) = \varnothing.$$

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 explanation.

DEFINITION. For any diagnostic problem P, $E \subseteq \mathbf{D}$ is an *explanation* for M^+ if

(i) $M^+ \subseteq \text{man}(E)$, or in words: E *covers* M^+; and
(ii) E is parsimonious.

This definition captures 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 (the "coverage goal" described in the Introduction). However, that is not enough; part (ii) specifies that E must also be one of the most "parsimonious" sets to do so. Part (ii) reflects what was referred to earlier as the parsimony goal or Occam's razor: the simplest explanation is the preferable one. This principle or heuristic is generally accepted as valid by human diagnosticians, but raises the issue of how one should go about formalizing the notion of parsimony or simplicity. One possibility is to replace (ii) in the definition of an explanation with

(ii′) $|E| \leqslant |D|$ for any other cover D of M^+.

In this case simplicity has been equated with minimal cardinality, and an explanation with a minimal cover [6].

While this is a reasonable concept of explanation in some situations, in others it is inadequate. For example, in medicine there are situations where two very common disorders might be a more plausible explanation for a patient's symptoms than a single very rare disorder. We will examine other notions of parsimony in Part II, but until then we will use (ii′), or minimal cardinality, as our criteria for a cover to be an explanation.

EXAMPLE. Let $P_1 = \langle \mathbf{D}, \mathbf{M}, \mathbf{C}, M^+ \rangle$, where $\mathbf{D} = \{d_1, d_2, \ldots, d_9\}$, $\mathbf{M} = \{m_1, \ldots, m_6\}$, and

$$\begin{aligned}\mathbf{C} = \{&\langle d_1, m_1\rangle, \langle d_1, m_4\rangle, \quad \langle d_2, m_1\rangle, \langle d_2, m_3\rangle, \langle d_2, m_4\rangle,\\ &\langle d_3, m_1\rangle, \langle d_3, m_3\rangle, \qquad \langle d_4, m_1\rangle, \langle d_4, m_6\rangle,\\ &\langle d_5, m_2\rangle, \langle d_5, m_3\rangle, \langle d_5, m_4\rangle, \quad \langle d_6, m_2\rangle, \langle d_6, m_3\rangle,\\ &\langle d_7, m_2\rangle, \langle d_7, m_5\rangle, \qquad \langle d_8, m_4\rangle, \langle d_8, m_5\rangle, \langle d_8, m_6\rangle,\\ &\langle d_9, m_2\rangle, \langle d_9, m_5\rangle\}.\end{aligned}$$

Thus, for example, $\text{man}(d_5) = \{m_2, m_3, m_4\}$ and $\text{causes}(m_3) = \{d_2, d_3, d_5, d_6\}$. Suppose that a specific case has manifestations $M^+ = \{m_1, m_4, m_5\}$. Note that no single disorder can cover (account for) all of M^+, but that some pairs of disorders do cover M^+. For instance, if $E = \{d_1, d_8\}$ then $M^+ \subseteq \text{man}(E)$. Since there are no covers for M^+ of smaller cardinality than E, it follows that E is an explanation for M^+. E is illustrated graphically in Figure 3.

We proceed to characterize further the concept of explanation.

LEMMA 1.3 (Explanation existence). *There exists at least one explanation for M^+ for any diagnostic problem.*

LEMMA 1.4. *If E is an explanation for M^+, then $|E| \leqslant |M^+|$.*

LEMMA 1.5. *$E = \varnothing$ is the only explanation for $M^+ = \varnothing$.*

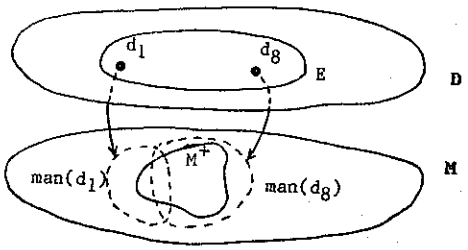

Fig. 3. Pictorial representation of an explanation E for M^+ in P_1, where M^+ is not contained in man(d_i) for any d_i.

LEMMA 1.6. *If E is an explanation for M^+, then $E \subseteq \text{causes}(M^+)$.*

Note that in contrast to Lemma 1.6, if D is a cover for M^+ then it is not necessarily true that $D \subseteq \text{causes}(M^+)$.

LEMMA 1.7. *If E_1 and E_2 are two explanations for M^+, then $|E_1| = |E_2|$.*

In view of Lemma 1.7, all explanations for a particular M^+ have the same cardinality so the following definition is meaningful:

DEFINITION. The *order* of a diagnostic problem P, designated order(P), is the cardinality of an explanation for M^+.

THEOREM 1.8 (Competing-disorders theorem). *Let E be an explanation for M^+, and let $\text{man}^+(d_1) \subseteq \text{man}^+(d_2)$ for some $d_1, d_2 \in \mathbf{D}$. Then*

(a) *d_1 and d_2 are not both in E, and*
(b) *if $d_1 \in E$, then $E' = (E - \{d_1\}) \cup \{d_2\}$ is also an explanation for M^+.*

The term "competing disorders" here refers to the fact that d_1 and d_2 can generally be viewed as alternatives to one another in forming an explanation. Thus, d_1 and d_2 might be grouped together in a disjunction such as was illustrated in Hypothesis H1 in the Introduction. As will be explained in Part II, a variant of the competing-disorders theorem was discovered by others on an intuitive basis and used as a central component of a computer model of diagnostic reasoning which they developed.

SOLUTIONS AND SOLUTION GENERATORS

In many diagnostic problems, such as those occurring in medicine, the diagnostician is generally interested in knowing all possible explanations for a

set of manifestations rather than a single explanation. This leads to the following definition:

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

Note that by Lemma 1.5, if $M^+ = \varnothing$ then Sol(P) = $\{\varnothing\}$.

EXAMPLE. The solution to problem P_1 in the previous example is the set of all explanations for $M^+ = \{m_1, m_4, m_5\}$:

$$\begin{aligned}\text{Sol}(P_1) = \{&\{d_1, d_7\}, \{d_1, d_8\}, \{d_1, d_9\},\\ &\{d_2, d_7\}, \{d_2, d_8\}, \{d_2, d_9\},\\ &\{d_3, d_8\}, \{d_4, d_8\}\}.\end{aligned}$$

Rather than representing Sol(P) as an explicit list of all possible explanations for M^+, it is advantageous to represent the disorders involved by "generators."

DEFINITION. Let $g_1, g_2, \ldots, g_n$ be nonempty pairwise disjoint subsets of **D**. Then $G_I = \{g_1, g_2, \ldots, g_n\}$ is a *generator*. The *class generated by* G_I, designated $[G_I]$, is defined to be $[G_I] = \{\{d_1, d_2, \ldots, d_n\} \mid d_i \in g_i, 1 \leqslant i \leqslant n\}$.

Note that if $G_I = \varnothing$ then $[G_I] = \{\varnothing\}$. The use of the numerical subscript I here is a notational convenience used to distinguish a single generator G_I from a generator set $G = \{G_1, G_2, \ldots, G_N\}$ as defined later.

A generator can be interpreted from two perspectives. By definition, a generator G_I is simply a set of $n \geqslant 0$ sets. However, $[G_I]$ is analogous to a Cartesian set product, the difference being that $[G_I]$ consists of unordered sets rather than ordered tuples. In particular, $[G_I]$ contains precisely all possible sets of n disorders which can be formed by taking one disorder from each $g_i \in G_I$. Thus, we can think of a generator as a structure similar to Hypothesis H2 in the Introduction. Each g_i corresponds to a disjunction (e.g., "bronchitis *or* asthma *or*..."), while G_I as a whole corresponds to a conjunction of such disjunctions (e.g., Hypothesis H2). To remind the reader of this special purpose of G_I, we will subsequently write $G_I = (g_1, g_2, \ldots, g_n)$ rather than $G_I = \{g_1, g_2, \ldots, g_n\}$.

DEFINITION. $G = \{G_1, G_2, \ldots, G_N\}$ is a *generator set* if each $G_I \in G$ is a generator and $[G_I] \cap [G_J] = \varnothing$ if $I \neq J$. The *class generated by* G is $[G] = [G_1] \cup [G_2] \cup \cdots \cup [G_N]$.

Even though an individual generator can itself be viewed as a set, we will use the term "generator set" to mean a set of generators as defined above. Note that a "generator set" is more than just a set of generators: the classes generated by each generator involved must be disjoint.

An immediate consequence of the above definitions is that if $G=\varnothing$ is a generator set, then $[G]=\varnothing$. Also, using our earlier convention, if $G=\{\varnothing\}$ then $[G]=\{\varnothing\}$. We will primarily be interested in a generator set which generates Sol(P) exactly, and henceforth will use the phrases "G is a generator set for Sol(P)" or "G represents Sol(P)" to mean that $[G]=\text{Sol}(P)$. In this context, each generator is a collection of sets of "competing" disorders that implicitly represents a set of explanations in Sol(P) and can be used to generate them.

EXAMPLE. Consider problem P_1 as presented in the previous example. Let $G_1=(\{d_1,d_2\},\{d_7,d_8,d_9\})$ and $G_2=(\{d_3,d_4\},\{d_8\})$. Then, $[G_2]=\{\{d_3,d_8\},\{d_4,d_8\}\}$, so $[G_2]\subseteq\text{Sol}(P_1)$. $[G_1]$ is the other six explanations in Sol(P_1). Since $[G_1]\cap[G_2]=\varnothing$, $G=\{G_1,G_2\}$ is a generator set for Sol(P_1).

There are at least three advantages to representing Sol(P) by a generator set. First, this is usually a more compact form of the explanations in the solution. Second, generators are a very convenient representation for developing algorithms to process explanations sequentially (see Part II). Finally, and perhaps most important, generators are closer to the way the human diagnostician organizes the possibilities during problem solving (e.g., the "differential diagnosis" in medicine). The following lemmas relate generators to diagnostic problems and are easily derivable from the preceding definitions.

LEMMA 1.9 (Generator-set existence). *For any diagnostic problem P, there exists a nonempty generator set G such that* $[G]=\text{Sol}(P)$.

LEMMA 1.10. *If P is a diagnostic problem with* $\text{order}(P)\leqslant 1$, *then it is possible to find a single generator G_I such that* $[G_I]=\text{Sol}(P)$.

Of course, if $\text{order}(P)>1$, then more than one generator may be necessary to completely and exactly produce Sol(P).

LEMMA 1.11. *Let $G_I=(g_1,g_2,\ldots,g_n)$ and $H_J=(h_1,h_2,\ldots,h_n)$ be two generators.*

(a) $[G_I]\cap[H_J]\neq\varnothing$ *if and only if there exists a relabeling of the members of G_I and H_J such that* $g_i\cap h_i\neq\varnothing\ \forall i,\ 1\leqslant i\leqslant n$.

(b) $[G_I]\subseteq[H_J]$ *if and only if there exists a relabeling of the members of G_I and H_J such that* $g_i\subseteq h_i,\ \forall i,\ 1\leqslant i\leqslant n$.

COROLLARY 1.12. *For two generators G_I and H_J,* $[G_I]=[H_J]$ *if and only if* $G_I=H_J$.

Because generators are central concepts in algorithmically solving diagnostic problems, it proves useful to define two operations that manipulate generators and establish the basic properties of these operations. The first operation is that of generator composition.

DEFINITION. Let $G_I = (g_1, g_2, \ldots, g_n)$ and $H_J = (h_1, h_2, \ldots, h_m)$ be two generators where $g_i \cap h_j = \varnothing \ \forall i, j$. Then the *composition* of G_I and H_J is

$$G_I \cdot H_J = (g_1, g_2, \ldots, g_n, h_1, h_2, \ldots, h_m).$$

For example, if $G_1 = (\{d_1, d_2\}, \{d_3, d_4, d_5\})$ and $G_2 = (\{d_6, d_7\})$, then $G_1 \cdot G_2 = (\{d_1, d_2\}, \{d_3, d_4, d_5\}, \{d_6, d_7\})$, effectively appending together G_1 and G_2. Generator composition will be used in the next section to construct the solution to a diagnostic problem that has been "decomposed" into smaller subproblems from the solutions to those subproblems. Note that generator composition is commutative and associative, and that $\varnothing$ is the identity element. In addition:

LEMMA 1.13. *Let G_I and H_J be two generators containing no disorders in common. Then* $[G_I \cdot H_J] = \{D_1 \cup D_2 | D_1 \in [G_I], D_2 \in [H_J]\}$.

LEMMA 1.14. *Let* $G_I = (g_1, g_2, \ldots, g_n)$, $H_J = (h_1, h_2, \ldots, h_m)$, *and* $F_K = (f_1, f_2, \ldots, f_p)$ *be generators such that* $g_i \cap f_j = \varnothing$ *and* $h_k \cap f_j = \varnothing \ \forall i, j, k$. *Then*

$$[G_I \cdot F_K] \cap [H_J \cdot F_K] = \{D_1 \cup D_2 | D_1 \in [G_I] \cap [H_J], D_2 \in [F_K]\}.$$

The second operation that will prove useful involves the "division" of generators by a set of disorders. Generator division will be used in Part II for modeling the sequential updating of a diagnostic hypothesis during problem solving.

DEFINITION. Let $G_I = (g_1, g_2, \ldots, g_n)$ be a generator, and let $D \subseteq \mathbf{D}$ with $D \neq \varnothing$. Then $H^I = \{H_{Ik} | H_{Ik}$ is a generator$\}$ is a *division of* G_I *by* D if $\forall k$, $1 \leqslant k \leqslant n$, we have $H_{Ik} = (h_{k1}, h_{k2}, \ldots, h_{kn})$, where

$$h_{kj} = \begin{cases} g_j - D & \text{if } j < k, \\ g_j \cap D & \text{if } j = k, \\ g_j & \text{if } j > k. \end{cases}$$

The qualification "H_{Ik} is a generator" in this definition eliminates from H^I entities of the form $H_{Ik} = (\ldots, \varnothing, \ldots)$, i.e., any H_{Ik} which contains $h_{kj} = \varnothing$ and is therefore by definition not a generator. Note that if $G_I = \varnothing$, then $n = 0$ and there is no k such that $1 \leqslant k \leqslant n$, so $H^I = \varnothing$. The concept of generator division seems quite strange when first seen, but we will give it an intuitive meaning shortly and will demonstrate its usefulness in Part II.

LEMMA 1.15. *Let H^I be a division of G_I by D. Then H^I is a generator set.*

LEMMA 1.16. *If H^I is a division of G_I by D, then $[H^I] \subseteq [G_I]$.*

EXAMPLE. If $G_I = (g_1, g_2, g_3)$ where $g_1 = \{d_1, d_2, d_3\}$, $g_2 = \{d_4, d_5, d_6\}$, and $g_3 = \{d_7, d_8\}$, and if $D = \{d_1, d_2, d_5\}$, then

$$H_{I1} = (\{d_1, d_2\}, \{d_4, d_5, d_6\}, \{d_7, d_8\}),$$

$$H_{I2} = (\{d_3\}, \{d_5\}, \{d_7, d_8\}),$$

and H_{I3} is not a generator, because $\varnothing \in H_{I3}$. Thus, $H^I = \{H_{I1}, H_{I2}\}$ is a division of G_I by D. Note that by considering the members of G_I in a different order, other divisions of G_I by D may be obtained. For example, by "reordering" G_I as (g_2, g_1, g_3) we obtain $K^I = \{K_{I1}, K_{I2}\}$ as a division of G_I by D, where $K_{I1} = (\{d_5\}, \{d_1, d_2, d_3\}, \{d_7, d_8\})$ and $K_{I2} = (\{d_4, d_6\}, \{d_1, d_2\}, \{d_7, d_8\})$.

Since there may be more than one division of G_I by D, we will adopt the following notation:

DEFINITION. G_I/D is the set of all divisions of G_I by D:

$$G_I/D = \{H^J | H^J \text{ is a division of } G_I \text{ by } D\}.$$

Thus, $H^I \in G_I/D$ indicates "H^I is a division of G_I by D." While a division of G_I by D does not necessarily produce a unique generator set, all generator sets in G_I/D are "the same" in that they generate the same class:

LEMMA 1.17. *Let G_I be a generator and D a nonempty subset of* **D**. *For every $H^I \in G_I/D$, $[H^I] = \{E \in [G_I] | E \cap D \neq \varnothing\}$.*

Lemma 1.17 is illustrated in the example above, where $[H^I] = [K^I]$ even though $H^I \neq K^I$. Lemma 1.17 also provides an intuitive meaning for generator division: if G_I represents the solution of some diagnostic problem, then after division of G_I by D we have H^I, where $[H^I]$ is the set of all explanations in $[G_I]$ which have an element in common with D. Other results that will be of use are the following:

LEMMA 1.18. *Let $G_I = (g_1, g_2, \ldots, g_n)$ be a generator and $D \subseteq$* **D** *where $D \neq \varnothing$.*

(a) *If $g_i \cap D = \varnothing$ $\forall i$, $1 \leqslant i \leqslant n$, then $G_I/D = \{\varnothing\}$.*
(b) *If $g_i \subseteq D$ for some $g_i \in G_I$, then $\{G_I\} \in G_I/D$.*

We now extend the definition of division by D in a natural fashion to generator sets and establish analogous results.

DEFINITION. Let $G = \{G_1, G_2, \ldots, G_N\}$ be a generator set and $D \subseteq \mathbf{D}$, where $D \neq \emptyset$. H is a *division of G by D* if

$$H = H^1 \cup H^2 \cup \cdots \cup H^N,$$

where $H^I \in G_I/D \ \forall I, 1 \leqslant I \leqslant N$. In other words, H is a division of G by D if H is the union of divisions of each of the G_I in G by D.

LEMMA 1.19. *Let $G = \{G_1, G_2, \ldots, G_N\}$ be a generator set, let $D \subseteq \mathbf{D}$ be nonempty, and let H be a division of G by D. Then H is a generator set.*

Since each H^I used in forming H is not necessarily unique, neither is H. Thus we let G/D represent the set of all divisions of G by D. Regardless of the form H takes, however, the class it generates is always the same:

LEMMA 1.20. *Let G be a generator set and D a nonempty subset of* **D**. *For every $H \in G/D$, $[H] = \{E \in [G] | E \cap D \neq \emptyset\}$.*

Finally, as with division of a single generator, the following holds as an immediate corollary of Lemma 1.20:

COROLLARY 1.21. *Let G be a generator set and D a nonempty subset of* **D**. *If $H \in G/D$ then $[H] \subseteq [G]$.*

PROBLEM DECOMPOSITION

Problem decomposition refers to the splitting or "decomposing" of a problem into independent subproblems [15]. When such subproblems are easier to solve than the original problem, and when the solutions to the subproblems can be easily "composed" to provide a solution to the original problem, decomposition provides a powerful tool for use in problem solving. For diagnostic problems as defined in this paper, decomposition is an important issue for the following reason.

THEOREM 1.22. *Finding the solution to a diagnostic problem is* NP *hard.*

The implication of this theorem is that in the general case, the effort required to construct Sol(P) can be expected to increase exponentially as order(P) increases. Since this potential for "combinatorial explosion" exists, it is im-

portant to address the question of when a diagnostic problem can be conveniently decomposed into smaller, independent subproblems, so that the sum of the efforts required to solve the subproblems will be less than the effort required to solve the original problem.

We present two examples of situations where diagnostic problems can be reduced to smaller subproblems. The first of these conditions is relatively weak in that it takes a problem P with $\text{order}(P) = N$ and breaks it into two subproblems with $\text{order}(P_1) = 1$ and $\text{order}(P_2) = N - 1$. It is based on the concept of pathognomonic manifestations.

DEFINITION. A manifestation $m \in \mathbf{M}$ is *pathognomonic for* d if $\text{causes}(m) = \{d\}$.

The concept of a pathognomonic manifestation is a familiar one in several fields. For example, physicians use the same name for such findings that uniquely characterize a disease (e.g., Kayser-Fleischer rings are a pathognomonic sign for hepatolenticular degeneration). Also, the concept of an essential prime implicant in Boolean minimization problems is analogous to a disorder which is the sole cause of a pathognomonic manifestation. When present, pathognomonic manifestations can be used to simplify solving diagnostic problems as shown below.

LEMMA 1.23. *Let* $P = \langle \mathbf{D}, \mathbf{M}, \mathbf{C}, M^+ \rangle$ *be a diagnostic problem where* $m \in M^+$ *is a pathognomonic manifestation for* $d \in \mathbf{D}$. *Then*

(a) *every cover for* M^+ *contains* d,

(b) D *covers* $M^+ - \text{man}(d)$ *if and only if* $D \cup \{d\}$ *covers* M^+, *and*

(c) D *is an explanation for* $M^+ - \text{man}(d)$ *if and only if* $D \cup \{d\}$ *is an explanation for* M^+.

The following decomposition theorem immediately follows:

THEOREM 1.24 (Decomposition using pathognomonic manifestations). *Let* $P = \langle \mathbf{D}, \mathbf{M}, \mathbf{C}, M^+ \rangle$ *be a diagnostic problem where* $m \in M^+$ *is a pathognomonic manifestation for* $d \in D$, *and let* $H = \{H_1, H_2, \ldots, H_n\}$ *be a generator set for* $\text{Sol}(P')$, *where* $P' = \{\mathbf{D}, \mathbf{M}, \mathbf{C}, M^+ - \text{man}(d)\}$. *Then* $G = \{G_1, G_2, \ldots, G_n\}$, *where* $G_I = H_I \cdot \{\{d\}\}$, *is a generator set for* $\text{Sol}(P)$.

The second example of a situation where a diagnostic problem can be conveniently decomposed into subproblems is based on a concept of "connectedness" of manifestations. This approach is more powerful than pathognomonic manifestations in that the order of subproblems may not be just one less than the order of the original problem. Connectedness is defined recursively

in the following definition:

DEFINITION. Manifestations $m_a \in \mathbf{M}$ and $m_b \in \mathbf{M}$ are *connected* to each other if either

(1) $\text{causes}(m_a) \cap \text{causes}(m_b) \neq \varnothing$, or
(2) there exist $\{m_1, m_2, \ldots, m_n\} \subseteq \mathbf{M}$ such that $m_a = m_1$, $m_b = m_n$, and each m_j is connected to m_{j+1} $\forall j, 1 \leqslant j \leqslant n-1$.

Otherwise, m_a and m_b are *unconnected* to each other.

This concept can be extended to sets of manifestations and pairs of sets of manifestations in a natural way. For example, $M \subseteq \mathbf{M}$ is a connected set if $\forall m_a, m_b \in M$, m_a is connected to m_b. However, the key concept we will need is the following:

DEFINITION. Two sets of manifestations M_1 and M_2 are *unconnected* to each other if $\forall m_1 \in M_1$, $\forall m_2 \in M_2$, $\text{causes}(m_1) \cap \text{causes}(m_2) = \varnothing$.

Using the concept of unconnectedness, we can immediately show the following.

LEMMA 1.25. *Let* $P = \langle \mathbf{D}, \mathbf{M}, \mathbf{C}, M^+ \rangle$ *be a diagnostic problem with* $M^+ = M_1^+ \cup M_2^+$ *where* M_1^+ *and* M_2^+ *are unconnected to each other. Then*

(a) $\text{causes}(M_1^+) \cap \text{causes}(M_2^+) = \varnothing$; *and*
(b) *E is an explanation for* M^+ *if and only if there are explanations* E_1 *and* E_2 *for* M_1^+ *and* M_2^+ *such that* $E = E_1 \cup E_2$.

THEOREM 1.26 (Decomposition using unconnected manifestations). *Let* $P = \langle \mathbf{D}, \mathbf{M}, \mathbf{C}, M^+ \rangle$ *be a diagnostic problem, and let* $M^+ = M_1^+ \cup M_2^+$, *where* M_1^+ *and* M_2^+ *are unconnected to each other. If* $G = \{G_1, G_2, \ldots, G_Q\}$ *is a generator set for* $\text{Sol}(P_1)$ *where* $P_1 = \langle \mathbf{D}, \mathbf{M}, \mathbf{C}, M_1^+ \rangle$ *and* $H = \{H_1, H_2, \ldots, H_R\}$ *is a generator set for* $\text{Sol}(P_2)$ *where* $P_2 = \langle \mathbf{D}, \mathbf{M}, \mathbf{C}, M_2^+ \rangle$, *then*

$$K = \{G_I \cdot H_J | 1 \leqslant I \leqslant Q, \quad 1 \leqslant J \leqslant R\}$$

is a generator set for $\text{Sol}(P)$.

This theorem establishes that if M^+ can be partitioned into N unconnected subsets, then the original diagnostic problem can be partitioned into N independent subproblems. The generator set for the solution to the original problem is then easily constructed by composing the generators for the solutions to the subproblems.

DISCUSSION

This paper has proposed a new model of diagnostic problems and their solutions that is based on a generalization of the set-covering problem in mathematics. The basic form of the GSC model presented in this paper conforms to much of what is known from past empirical studies of how human diagnosticians conceptually organize the information they use in diagnostic problem solving (e.g., see [1, 3, 7, 9, 11, 17, 23, 24, 25, 29]). For example, in medicine the GSC model captures a number of concepts that are intuitively familiar to the physician. The causal relationship between diseases and manifestations is represented as a relation **C**, which is then used to define the sets man(d_i) and causes(m_j). As we have explained elsewhere [21], the set man(d_i) is analogous to the descriptive information about a disorder that is found in medical textbooks, and provides a convenient organization for knowledge representation in expert systems. The set causes(m_j) corresponds to the well-known notion of a "differential diagnosis" for a single manifestation. Manifestations where $|\text{causes}(m_j)| = 1$ represent what a physician would refer to as "pathognomonic" manifestations.

The concept of "explanation" defined as a minimal cover in the basic GSC model also embodies intuitively reasonable ideas from the human diagnostician's viewpoint. Not only does an explanation represent a set of disorders that can cause a set of manifestations, but it also reflects the generally recognized principle of Occam's razor. For example, in medicine the solution to a diagnostic problem, especially when organized as a generator containing sets of "competing disorders," represents a generalized "differential diagnosis" for a patient's manifestations. Generators provide a plausible approach to diagnostic problems where multiple, simultaneous causative disorders may be present, a situation which is well recognized to be difficult to represent in other models [20]. Finally, decomposition of a diagnostic problem into subproblems based on manifestation unconnectedness seems very reasonable from the point of view that such subproblems are intuitively unrelated.

It is of interest to contrast the basic form of the GSC model of diagnostic problems presented here with the classical set-covering problem and to emphasize in what ways it is a generalization. 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 K of S from F is a subfamily $K \subseteq F$ such that $\cup(K) = S$. A cover K is called minimum if its cardinality is as small as possible [6].

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^+, which we have designated man$^+$(d_i). A minimum cover K is analogous to the idea of an explanation E. However,

diagnostic problems as defined in this paper generalize the usual formulation of set-covering problems in two ways. First, in the case where $M^+ \neq \mathbf{M}$ an explanation E need only satisfy $M^+ \subseteq \text{man}(E)$ rather than $M^+ = \text{man}(E)$, as would be required in the set covering problem. This first difference represents a partial rather than exact match of causative disorders to manifestations. Furthermore, any set-covering problem can be formulated as a diagnostic problem where $\mathbf{M} = M^+$ (see the proof of Theorem 1.22). Second, we have been concerned with finding *all* minimal covers rather than a single minimal cover as is done in the set covering problem. This latter difference reflects the diagnostician's desire to know what all of the possibilities are rather than just a single possibility. As described in Part II, the basic form of the GSC model also embodies a third generalization: algorithmic solution in a sequential fashion where only part of M^+ is known initially and the rest of M^+ must be "discovered" through question generation at problem-solving time.

Set covering has previously found application in such diverse areas as optimizing logic circuits, scheduling, assembly-line balancing, and information retrieval (see [4] for a review). However, to our knowledge, the analogy between the set-covering problem and general diagnostic reasoning has not previously been examined in detail. Some related biomedical work has been done for assignment of HLA specificities to antisera (see [13] and [28]), but this was more limited in scope and did not address many of the issues involved in general diagnostic problem solving.

APPENDIX A. TABLE OF SYMBOLS

Symbol	Meaning
$\in$	Element of
$\subseteq$	Subset of
$\leqslant$	Less than or equal
$\varnothing$	Empty set
$\forall$	For all
$\cup$	Set union
$\cap$	Set intersection
$-$	Set difference
$\lvert\ \rvert$	Set cardinality
$\mathbf{M}$	Universe of manifestations (effects)
$\mathbf{D}$	Universe of disorders (causes)
$\mathbf{C}$	Causative relationship
$\{\ \}$	Set
$\langle\ \rangle$	Ordered tuple
$(\ )$	Generator
$[\ ]$	Class generated by a generator

APPENDIX B. PROOFS

LEMMA. 1.1. (a): *Since* $\text{domain}(\mathbf{C}) = \mathbf{D}$ *by the definition of a diagnostic problem,* $\forall d_i \in \mathbf{D}$ *there exists* m_j *such that* $\langle d_i, m_j \rangle \in \mathbf{C}$, *so* $\text{man}(d_i) \neq \varnothing$. ($a'$) *is analogous.*

(b): *By* (a) *there exists* $m_j \in \text{man}(d_i)$ *such that* $\langle d_i, m_j \rangle \in \mathbf{C}$. *Thus* $d_i \in \text{causes}(m_j) \subseteq \cup_{m_k \in \text{man}(d_i)} \text{causes}(m_k) = \text{causes}(\text{man}(d_i))$.

(b′) *is analogous.*

(c), (c′): *Immediate from* (b), (b′).

(d): $\forall d_i \in \mathbf{D}$, $d_i \in \text{causes}(\text{man}(d_i)) \subseteq \text{causes}(\mathbf{M})$ *by* (*b*). *Thus,* $\mathbf{D} \subseteq \text{causes}(\mathbf{M})$. *If* $d_i \in \text{causes}(\mathbf{M})$, *then* $d_i \in \{d_j | \langle d_j, m_k \rangle \in \mathbf{C}\} \subseteq \mathbf{D}$. *Thus,* $\text{causes}(\mathbf{M}) \subseteq \mathbf{D}$. (d′) *is analogous.*

(e): $d_i \in \text{causes}(m_j)$ *iff* $\langle d_i, m_j \rangle \in \mathbf{C}$ *iff* $m_j \in \text{man}(d_i)$, *all by definition of* $\mathbf{C}$.

(f): *If* $m \in \text{man}(D) - \text{man}(D')$, *then there is a* $d \in D$ *such that* $m \in \text{man}(d)$. *Since* $m \notin \text{man}(D) - \text{man}(D')$, *it follows that* $d \notin D'$, *so* $d \in D - D'$. *Thus* $m \in \text{man}(d) \subseteq \text{man}(D - D')$. *Since this is true* $\forall m \in \text{man}(D) - \text{man}(D')$, *it follows that* $\text{man}(D) - \text{man}(D') \subseteq \text{man}(D - D')$.

LEMMA 1.2. *Let* $D \cap \text{causes}(M) = \varnothing$. *Then* $\forall m_j \in M$, $\text{causes}(m_j) \cap D = \varnothing$, *so* $\forall d_i \in D$, $\langle d_i, m_j \rangle \notin C$. *Thus,* $\forall m_j \in M$, $m_j \notin \text{man}(D)$, *and therefore* $M \cap \text{man}(D) = \varnothing$. *The converse is proven in an analogous fashion.*

LEMMA 1.3. $\mathbf{D}$ *covers any* M^+, *so there always exists a cover for* M^+. *Selection of one of the smallest covers* $D \subseteq \mathbf{D}$ *will therefore always provide an explanation for* M^+.

LEMMA 1.4. *Let* $D \subseteq \mathbf{D}$ *be constructed by selecting any one* d_i *from the guaranteed nonempty* $\text{causes}(m_i)$ *for each* $m_i \in M^+$. *Then* $|D| \leqslant |M^+|$. *D covers* M^+, *because* $M^+ \subseteq \text{man}(D)$. *Since E is an explanation for* M^+, $|E| \leqslant |D|$, *so* $|E| \leqslant |M^+|$.

LEMMA 1.5. $\varnothing \subseteq \text{man}(\varnothing)$, *and* $\varnothing$ *is clearly minimal. There are no other covers of cardinality zero.*

LEMMA 1.6. *Suppose there is a* $d \in E$ *such that* $d \notin \text{causes}(M^+)$. *Then* $\{d\} \cap \text{causes}(M^+) = \varnothing$, *so by Lemma* 1.2,

$$M^+ \cap \text{man}(d) = \varnothing. \tag{6A}$$

Since E is an explanation,

$$M^+ \subseteq \text{man}(E)$$

so,

$$M^+ - \text{man}(d) \subseteq \text{man}(E) - \text{man}(d),$$

so

$$M^+ \subseteq \text{man}(E) - \text{man}(d) \qquad \text{(by [6A])},$$

so

$$M^+ \subseteq \text{man}(E - \{d\}) \qquad \text{[by Lemma 1.1 (f)]}.$$

But $|E - \{d\}| < |E|$, *contradicting the minimality of* E.

LEMMA. 1.7. *Otherwise the minimality of one of these explanations is contradicted.*

THEOREM 1.8. (a): *Suppose that both* d_1 *and* d_2 *are in* E. *Since* $\text{man}^+(d_1) \subseteq \text{man}^+(d_2)$, $\text{man}^+(E - \{d_1\}) = \text{man}^+(E) = M^+$, *making* $E - \{d_1\}$ *a cover for* M^+. *This contradicts the minimality of* E.

(b): E' *covers* M^+, *because*

$$\begin{aligned} M^+ &= \text{man}^+(E) = \text{man}^+(E - \{d_1\}) \cup \text{man}^+(d_1) \\ &\subseteq \text{man}^+(E - \{d_1\}) \cup \text{man}^+(d_2) \\ &= \text{man}^+\left[(E - \{d_1\}) \cup \{d_2\}\right] \\ &= \text{man}^+(E'). \end{aligned}$$

Also, $|E'| = |E|$, so E' is minimal, making E' an explanation for M^+.

LEMMA 1.9. *By Lemma* 1.3, $\text{Sol}(P) \neq \varnothing$. *If* $\text{Sol}(P) = \{\varnothing\}$, $G = \{\varnothing\}$ *is a nonempty generator set for* $\text{Sol}(P)$. *If* $\text{Sol}(P) \neq \{\varnothing\}$, *then* $\text{Sol}(P) = \{E_1, E_2, \ldots, E_n\}$, *where (by Lemma* 1.7), $|E_1| = |E_2| = \cdots = |E_n| = N$ *for some* $N \geq 1$. *For each* $E_i = \{d_{i1}, d_{i2}, \ldots, d_{iN}\}$, *construct a corresponding generator* $G_i = (\{d_{i1}\}, \{d_{i2}\}, \ldots, \{d_{iN}\})$ *and let* $G = \{G_1, G_2, \ldots, G_n\}$. *Since* $[G_i] = \{E_i\}$ $\forall i$, *it follows that* $[G_j] \cap [G_k] = \varnothing$ *if* $j \neq k$, *and that* $[G] = \text{Sol}(P)$. *Hence,* G *is a nonempty generator set for* $\text{Sol}(P)$.

LEMMA 1.10. *If* $\text{order}(P) = 0$, *then* $\text{Sol}(P) = \{\varnothing\}$ *and* $G_I = \varnothing$ *is a single generator such that* $[G_I] = \text{Sol}(P)$. *If* $\text{order}(P) = 1$, *then* $\text{Sol}(P) = \{E_1, E_2, \ldots, E_n\}$, *where* $n \geq 1$ *and* $|E_i| = 1$ $\forall i$, $1 \leq i \leq n$. *Without loss of generality, let* $E_i = \{d_i\}$ $\forall i$. *Then* $G_I = (\{d_1, d_2, \ldots, d_n\})$ *is a single generator such that* $[G_I] = \text{Sol}(P)$.

LEMMA 1.11. (a): *Suppose* $[G_I] \cap [H_J] \neq \varnothing$. *Then there exists some* $D = \{d_1, d_2, \ldots, d_n\}$ *such that* $D \in [G_I] \cap [H_J]$. *Since* $D \in [G_I]$, *relabel the elements of* G_I *so* $d_i \in g_i$ $\forall i$, $1 \leq i \leq n$. *Similarly, since* $D \in [H_J]$, *relabel the elements of*

H_J so $d_i \in h_i$ $\forall i, 1 \leqslant i \leqslant n$. Then $g_i \cap h_i \neq \varnothing$ $\forall i$. Conversely, suppose $g_i \cap h_i \neq \varnothing$ $\forall i$, $1 \leqslant i \leqslant n$. Let $D = \{d_1, d_2, \ldots, d_n\}$ be such that $d_i \in g_i \cap h_i$ $\forall i$. Then $D \in [G_I]$ and $D \in [H_J]$, so $[G_I] \cap [H_J] \neq \varnothing$.

(b): *Suppose $[G_I] \subseteq [H_J]$. For $1 \leqslant i \leqslant n$, let $d_i \in g_i$, whence $\{d_1, d_2, \ldots, d_n\} \in [G_I] \subseteq [H_J]$. Then $\forall d_i$, there is a single $h_j \in H_J$ such that $d_i \in h_j$, so relabel the elements of H_J such that $d_i \in h_i$. But then $\forall d_i' \in g_i$, $\{d_1, d_2, \ldots, d_{i-1}, d_i', d_{i+1}, \ldots, d_n\} \in [G_I] \subseteq [H_J]$, so $d_i' \in h_i$. Thus $g_i \subseteq h_i$ $\forall i$, $1 \leqslant i \leqslant n$.*

Conversely, suppose there is a relabeling of G_I and H_J such that $\forall i$, $g_i \subseteq h_i$. Then $[G_I] = \{\{d_1, d_2, \ldots, d_n\} | d_i \in g_i\} \subseteq \{\{d_1, d_2, \ldots, d_n\} | d_i \in h_j\} = [H_J]$, so $[G_I] \subseteq [H_J]$.

COROLLARY 1.12. *Immediate from Lemma* 1.11(b).

LEMMA 1.13.

$$
\begin{aligned}
[G_I \cdot H_J] &= [(g_1, g_2, \ldots, g_n, h_1, h_2, \ldots, h_m)] \\
&= \{\{d_1, \ldots, d_n, d_{n+1}, \ldots, d_{n+m}\} | d_i \in g_i, 1 \leqslant i \leqslant n;\ d_i \in h_{i-n}, \\
&\qquad\qquad n+1 \leqslant i \leqslant n+m\} \\
&= \{\{d_1, \ldots, d_n\} \cup \{d_{n+1}, \ldots, d_{n+m}\} | \ldots\} \\
&= \{D_1 \cup D_2 | D_1 \in [G_I],\ D_2 \in [H_J]\}.
\end{aligned}
$$

LEMMA 1.14. *By Lemma* 1.13,

$$
\begin{aligned}
[G_I \cdot F_K] \cap [H_J \cdot F_K] &= \{D_1 \cup D_2 | D_1 \in [G_I],\ D_2 \in [F_K]\} \\
&\qquad \cap \{D_3 \cup D_4 | D_3 \in [H_J],\ D_4 \in [F_K]\} \\
&= \{D_1' \cup D_2' | D_1' \in [G_I] \cap [H_J],\ D_2' \in [F_K]\}.
\end{aligned}
$$

LEMMA 1.15. *By the definition of H^I, if $H_{Ik} \in H^I$ then H_{Ik} is a generator. We therefore only need to show that if $H_{Ii} = (h_{i1}, h_{i2}, \ldots, h_{in}) \in H^I$ and $H_{Ij} = (h_{j1}, h_{j2}, \ldots, h_{jn}) \in H^I$ where $i \neq j$, then $[H_{Ii}] \cap [H_{Ij}] = \varnothing$. Without loss of generality, assume $i < j$. From Lemma* 1.11(*a*) *it suffices to show that $h_{ii} \cap h_{jr} = \varnothing$ $\forall r$. For the case where $r = i$, we have $h_{ii} = g_i \cap D$ and $h_{ji} = g_i - D$, so $h_{ii} \cap h_{ji} = \varnothing$. For the case where $r \neq i$, we have $g_i \cap g_r = \varnothing$, and thus since $h_{ii} \subseteq g_i$ and $h_{ir} \subseteq g_r$, we have $h_{ii} \cap h_{jr} = \varnothing$.*

LEMMA 1.16. *If* $H^I = \emptyset$, *then* $[H^I] = \emptyset \subseteq [G_I]$ *for any* G_I. *If* $H^I \neq \emptyset$, *consider any generator* $H_{Ik} \in H^I$. *By the definition of division of a generator,* $h_{kj} \subseteq g_j$ $\forall j$, $1 \leqslant j \leqslant n$. *Thus, by Lemma* 1.11(b), $[H_{Ik}] \subseteq [G_I]$, *so by the definition of the class generated by a generator set,* $[H^I] \subseteq [G_I]$.

LEMMA 1.17. *We show below that for any* $H^I \in G_I/D$, $[H^I]$ *and* $\{E \in [G_I] \mid E \cap D \neq \emptyset\}$ *are subsets of each other and therefore are equal.*

Let $E \in [H^I]$. *By Lemma* 1.16, $E \in [G_I]$. *Also* $E \in [H_{Ik}]$ *for some* $H_{Ik} \in H^I$, *so* $E = \{d_1, d_2, \ldots, d_n\}$ *with* $d_i \in h_{ki}$ $\forall i$. *Then* $d_k \in h_{kk} = g_k \cap D \neq \emptyset$, *so* $d_k \in D$, *whence* $E \cap D \neq \emptyset$. *Thus* $[H^I] \subseteq \{E \in [G_I] \mid E \cap D \neq \emptyset\}$.

Conversely, let $E \in [G_I]$ *such that* $E \cap D \neq \emptyset$. *Then* $E = \{d_1, d_2, \ldots, d_n\}$ *with* $d_i \in g_i$ $\forall i$, *and* E *contains at least one* $d_i \in D$. *Let* k *be the smallest* i *such that* $d_i \in D$. *Then from the definition of* H_{Ik},

$\forall i < k$, $h_{ki} = g_i - D = g_i \neq \emptyset$, *so* $d_i \in h_{ki}$,
for $i = k$, $h_{ki} = g_k \cap D \neq \emptyset$ *because* $d_k \in h_{kk}$,
and $\forall i > k$, $h_{ki} = g_k \neq \emptyset$, *so* $d_i \in h_{ki}$.

Thus, $D \in [(h_{k1}\ h_{k2} \cdots h_{kn})] = [H_{Ik}] \subseteq [H^I]$. *Since this is true* $\forall E \in [G_I]$ *such that* $E \cap D \neq \emptyset$,

$$\{E \in [G_I] \mid E \cap D \neq \emptyset\} \subseteq [H^I].$$

LEMMA 1.18. (a): *Let* $H^I \in G_I/D$, *and suppose there exists a generator* $H_{Ik} \in H^I$. *Then* $h_{kk} \in H_{Ik}$ *is given by* $h_{kk} = g_k \cap D = \emptyset$, *contradicting the fact that* H_{Ik} *is a generator. Therefore* $H^I = \emptyset$, *so* $G_I/D = \{\emptyset\}$.

(b): *Relabel* G_I *as* $G_I' = (g_1', g_2', \ldots, g_n')$ *where* $g_1' = g_i$, $g_j' = g_{j-1}$ *for* $2 \leqslant j \leqslant i$, *and* $g_j' = g_j$ *for* $j > i$. *Let* $H^I \in G_I/D$ *be constructed according to the relabelling* G_I' *of* G_I. *Then*

$$H_{I1} = (g_1' \cap D, g_2', \ldots, g_n') = (g_i \cap D, g_1, g_2, \ldots, g_{i-1}, g_{i+1}, \ldots, g_n)$$
$$= (g_i, g_1, g_2, \ldots, g_{i-1}, g_{i+1}, \ldots, g_n) = G_I.$$

Also for $k \neq 1$,

$$H_{Ik} = (g_1' - D, \ldots) = (g_i - D, \ldots) = (\emptyset, \ldots),$$

so $H_{Ik} \notin H^I$. *Thus,* $H^I = \{G_I\}$, *so* $\{G_I\} \in G_I/D$.

LEMMA 1.19. *Since* $H = \cup H^I$ *where each* H^I *is a generator set by Lemma* 1.15, H^I *contains only generators. To show* H *is a generator set, we therefore only need to show that the classes generated by its generators are distinct. Let* $H_1 \in H^I$ *and* $H_2 \in H^J$ *be two generators in* H. *If* $I = J$, *then* $[H_1] \cap [H_2] = \emptyset$ *if* $H_1 \neq H_2$

by Lemma 1.15. *Suppose* $I \neq J$. *Then* $[H_1] \subseteq [H^I]$ *by the definition of* $[H^I]$, *so* $[H_1] \subseteq [G_I]$ *by Lemma* 1.18. *Analogously,* $[H_2] \subseteq [H^J] \subseteq [G_J]$. *Since* G *is a generator set,* $[G_I] \cap [G_J] = \varnothing$. *Hence,* $[H_1] \cap [H_2] = \varnothing$.

LEMMA 1.20.

$E' \in [H]$ *iff* $E' \in [H^I]$ *for some* I, *where* $H^I \in G_I/D$ *and* $G_I \in G$,
iff $E' \in \{E \in [G_I] | E \cap D \neq \varnothing\}$ *for some* I (*by Lemma* 1.17),
iff $E' \in \{E \in [G] | E \cap D \neq \varnothing\}$.

Thus, $[H] = \{E \in [G] | E \cap D \neq \varnothing\}$.

COROLLARY 1.21. *Immediate from Lemma* 1.20.

THEOREM 1.22. *The long-studied problem of finding an exact, single set cover is known to be* NP *complete* [8]. *We will therefore reduce the set covering problem to a diagnostic problem by showing that there is a function* f *operating in polynomial time such that* $f(S, F)$ *is a diagnostic problem having a cover of size* $\leqslant n$ *if and only if* S *is a set and* F *is a family of subsets of* S *with a set cover* $K \subseteq F$ *of size* $\leqslant n$.

Given S *and* F, *let* $f(S, F) = \langle \mathbf{D}, \mathbf{M}, \mathbf{C}, M^+ \rangle$, *where* $\mathbf{D} = F - \{\varnothing\}$, $\mathbf{M} = S$, $M^+ = S$, *and* $\mathbf{C} = \{\langle d, m \rangle | d \in \mathbf{D}, \ m \in \mathbf{M}, \text{ and } m \in d\}$. *Clearly* $P = \langle \mathbf{D}, \mathbf{M}, \mathbf{C}, M^+ \rangle$ *can be constructed in polynomial time from* S *and* F.

Suppose S *has a set cover* $K \subseteq F$ *of size* $\leqslant n$. *Then*

$$\begin{aligned} \text{domain}(\mathbf{C}) &= \{d \in \mathbf{D} | \langle d, m \rangle \in \mathbf{C} \text{ for some } m \in \mathbf{M}\} \\ &= \{d \in F - \{\varnothing\} | m \in d \text{ for some } m \in S\} \\ &= F - \{\varnothing\}, \end{aligned}$$

since every nonempty member of F *contains at least one element of* S;

$$\begin{aligned} \text{range}(\mathbf{C}) &= \{m \in \mathbf{M} | \langle d, m \rangle \in \mathbf{C} \text{ for some } d \in \mathbf{D}\} \\ &= \{m \in S | m \in d \text{ for some } d \in F - \{\varnothing\}\} \\ &= S, \text{ since } F \text{ contains a set cover for } S; \end{aligned}$$

and $\mathrm{M}^+ = \mathrm{S} \subseteq \mathrm{S} = \mathbf{M}$.

Thus, P *is a diagnostic problem. In addition,* $K - \{\varnothing\} \subseteq F - \{\varnothing\} = \mathbf{D}$, *and* $\forall s \in S$ *there exists* $K_i \in K - \{\varnothing\}$ *such that* $s \in K_i$, *so* $\langle K_i, s \rangle \in \mathbf{C}$. *Thus,* $\forall s \in M^+$ *there is a* $K_i \in K - \{\varnothing\}$ *such that* $\langle K_i, s \rangle \in \mathbf{C}$, *so* $K - \{\varnothing\}$ *is a* P*-cover in the diagnostic problem* P *for* M^+, *and* $|K - \{\varnothing\}| \leqslant n$.

Conversely, suppose $P = \langle \mathbf{D}, \mathbf{M}, \mathbf{C}, M^+ \rangle$ *is a diagnostic problem having a cover* E *for* M^+ *of size* $\leqslant n$. *Then* $E \subseteq \mathbf{D} = F - \{\varnothing\} \subseteq F$, *and* $\forall m \in M^+$ *there exists* $d \in E$ *such that* $\langle d, m \rangle \in \mathbf{C}$, *whence* $m \in d$. *Thus* $\forall m \in S$ *there exists* $d \in E$ *such that* $m \in d$. *Thus,* E *is a set cover for* S *of size* $\leqslant n$.

LEMMA 1.23. (a): *Since* m *is pathognomonic for* d, $\text{causes}(m) = \{d\}$. *By Lemma* 1.1(e), d *is the only disorder such that* $m \in \text{man}(d)$. *Let* D *be a cover for* M^+. *Then* $m \in M^+ \subseteq \text{man}(D)$, *so* $d \in D$.

(b) *Suppose* D *is a cover for* $M^+ - \text{man}(d)$. *Then* $M^+ - \text{man}(d) \subseteq \text{man}(D)$, *so* $M^+ \subseteq [M^+ - \text{man}(d)] \cup \text{man}(d) \subseteq \text{man}(D) \cup \text{man}(d) = \text{man}(D \cup \{d\})$, *making* $D \cup \{d\}$ *a cover for* M^+. *Conversely, suppose* $D \cup \{d\}$ *covers* M^+. *Then* $M^+ \subseteq \text{man}(D \cup \{d\}) = \text{man}(D) \cup \text{man}(d)$. *Therefore,* $M^+ - \text{man}(d) \subseteq [\text{man}(D) \cup \text{man}(d)] - \text{man}(d) = \text{man}(D) - \text{man}(d) \subseteq \text{man}(D)$, *so* D *covers* $M^+ - \text{man}(d)$.

(c): *Suppose* $D \cup \{d\}$ *is an explanation for* M^+. *Then by Lemma* 1.23(b), D *covers* $M^+ - \text{man}(d)$. *If* D *is* not *an explanation for* $M^+ - \text{man}(d)$, *then there is a smaller cover* E *for* $M^+ - \text{man}(d)$, *so by Lemma* 1.23(b), $E \cup \{d\}$ *covers* M^+. *But* $|E \cup \{d\}| < |D \cup \{d\}|$, *contradicting the assumption that* $D \cup \{d\}$ *is an explanation for* M^+.

Conversely, suppose D *is an explanation for* $M^+ - \text{man}(d)$ *but that* $D \cup \{d\}$ *is* not *an explanation for* M^+. *By Lemma* 1.23(b), $D \cup \{d\}$ *covers* M^+, *so there is a smaller cover* E *for* M^+. *From Lemma* 1.23(a) *we have* $d \in E$. *Thus, by Lemma* 1.23(b), $E - \{d\}$ *is a cover for* $M^+ - \text{man}(d)$. *But then* $|E - \{d\}| < |D|$, *contradicting the assumption that* D *is an explanation for* $M^+ - \text{man}(d)$. *Thus,* $D \cup \{d\}$ *must be an explanation for* M^+.

THEOREM 1.24. *Let* $H_I = (h_{I1}, h_{I2}, \ldots, h_{Ik})$ *for* $I = 1, 2, \ldots, n$, *where* $k = \text{order}(P)$. *Then*

$$\begin{aligned}
\text{Sol}(P) &= \{D \cup \{d\} \mid D \in \text{Sol}(P')\} \qquad \text{by Lemma 1.23c} \\
&= \left\{ D \cup \{d\} \mid D \in \bigcup_{I=1}^{n} [H_I] \right\} \\
&= \left\{ D \cup \{d\} \mid D \in \bigcup_{I=1}^{n} \{\{d_{I1}, d_{I2}, \ldots, d_{Ik}\} \mid d_{Ij} \in h_{Ij}\} \right\} \\
&= \bigcup_{I=1}^{n} \{\{d_{I1}, d_{I2}, \ldots, d_{Ik}, d\} \mid d_{Ij} \in h_{Ij}\} \} \\
&= \bigcup_{I=1}^{n} [(h_{I1}, h_{I2}, \ldots, h_{Ik}, \{d\})] \\
&= \bigcup_{I=1}^{n} [H_I \cdot (\{d\})] = \bigcup_{I=1}^{n} [G_I] = [G].
\end{aligned}$$

Also, if $G_I \in G$ *and* $G_J \in G$ *with* $I \neq J$, *then* $[G_I] \cap [G_J] = \varnothing$, *since* H *is a generator set.*

LEMMA 1.25. (a): *Immediate from definitions of* causes(M) *and set unconnectedness.*

(b): "*If*": *Assume* E_1 *and* E_2 *are explanations for* M_1^+ *and* M_2^+ *respectively, and let* $E = E_1 \cup E_2$. E *covers* M^+, *because*:

$$M^+ = M_1^+ \cup M_2^+ \subseteq \text{man}(E_1) \cup \text{man}(E_2) = \text{man}(E).$$

It remains to show that for any other cover D *for* M^+, $|E| \leqslant |D|$.

Let D *be a cover for* M^+, *and let* $D_1 = D \cap \text{causes}(M_1^+)$, $D_2 = D \cap \text{causes}(M_2^+)$, *and* $D_3 = D - D_1 - D_2$. *Then* $D = D_1 \cup D_2 \cup D_3$. *By Lemma* 1.25(a), $\text{causes}(M_1^+) \cap D_2 = \varnothing$, *so by Lemma* 1.2,

$$M_1^+ \cap \text{man}(D_2) = \varnothing.$$

Similarly, since $\text{causes}(M_1^+) \cap D_3 = \varnothing$, *we again have by Lemma* 1.2

$$M_1^+ \cap \text{man}(D_3) = \varnothing.$$

Thus, since $M_1^+ \subseteq \text{man}(D)$, *it must be because* $M_1^+ \subseteq \text{man}(D_1)$, *so* D_1 *covers* M_1^+. *Thus,*

$$|D_1| \geqslant |E_1|. \tag{25A}$$

Analogously, it can be shown that D_2 covers M_2^+, so

$$|D_2| \geqslant |E_2|. \tag{25B}$$

Thus,

$\|D\| = \|D_1\| + \|D_2\| + \|D_3\|$	*because* $D_1 \cap D_3 = \varnothing$ *and* $D_2 \cap D_3 = \varnothing$ *by their definition, and* $D_1 \cap D_2 = \varnothing$ *by unconnectedness of* M_1^+ *and* M_2^+ *and Lemma* 1.25(a)
$\geqslant \|D_1\| + \|D_2\|$	*because* $\|D_3\| \geqslant 0$
$\geqslant \|E_1\| + \|E_2\|$	*by* (25A) *and* (25B) *above*
$= \|E\|$	*because* $E_1 \cap E_2 = \varnothing$ *by unconnectedness of* M_1^+ *and* M_2^+ *and Lemma* 1.25(a).

Thus, E *is an explanation for* M^+.

"Only if": Assume E is an explanation for $M^+ = M_1^+ \cup M_2^+$. *E can be written* $E = E_1 \cup E_2 \cup E_3$, *where* $E_1 = E \cap \text{causes}(M_1^+)$, $E_2 = E \cap \text{causes}(M_2^+)$, *and* $E_3 = E - E_1 - E_2$. *As in the immediately preceeding paragraphs, it follows that* E_1 *covers* M_1^+, E_2 *covers* M_2^+, *and* $E_1 \cap E_2 = \emptyset$ *by the unconnectedness of* M_1^+ *and* M_2^+. *Furthermore,* $E_3 = \emptyset$, *for otherwise* $E_1 \cup E_2$ *alone would be a smaller cover of* M^+, *contradicting the minimality of E. It remains to show that both* E_1 *and* E_2 *are minimal covers for* M_1^+ *and* M_2^+. *Suppose they are not. Then at least one (say* E_1*) is not minimal, so there exists a proper subset of* E_1 *called* E_1' *which is a cover for* M_1^+ *with* $|E_1'| < |E_1|$. *But then* $E_1' \cup E_2$ *covers* M^+, *and since* $E_1' \subsetneq E_1$, $E_1' \cap E_2 = \emptyset$, *so* $|E_1' \cup E_2| = |E_1'| + |E_2| < |E_1| + |E_2| = |E|$, *contradicting the minimality of E. Thus,* E_1 *and* E_2 *are explanations for* M_1^+ *and* M_2^+ *such that* $E = E_1 \cup E_2$.

THEOREM 1.26. *We first show that K is a generator set. Since composition of generators results in generators by definition, it is only necessary to prove that the classes generated by the generators in K are disjoint.*

By Lemma 1.13, $\forall G_I, G_J \in G$ *and* $\forall H_S, H_T \in H$, *we have* $[G_I \cdot H_S] = \{D_i \cup D_s | D_i \in [G_I],\ D_s \in [H_S]\}$ *and* $[G_J \cdot H_T] = \{D_j \cup D_t | D_j \in [G_J],\ D_t \in [H_T]\}$. *For each* $D_1 \cup D_2 \in [G_I \cdot H_S]$ *where* $D_1 \in [G_I]$ *and* $D_2 \in [H_s]$, *consider any* $D_3 \cup D_4 \in [G_J \cdot H_T]$, *where* $D_3 \in [G_J]$ *and* $D_4 \in [H_T]$. *Suppose* $I \neq J$. *Then* $D_1 \neq D_3$ *because G is a generator set, so there is a* $d_1 \in D_1 - D_3$, *because* $|D_1| = |D_3|$. *But* $D_1 \cap D_4 = \emptyset$ *by the unconnectedness of* M_1^+ *and* M_2^+, *so* $d_1 \notin D_4$, *and thus* $d_1 \notin D_3 \cup D_4$. *Since this is true* $\forall D_3 \cup D_4 \in [G_J \cdot H_T]$, *it follows that* $D_1 \cup D_2 \notin [G_J \cdot H_T]$. *Since this is true* $\forall D_1 \cup D_2 \in [G_I \cdot H_S]$, *it follows that no member of* $[G_I \cdot H_S]$ *is in* $[G_J \cdot H_T]$. *Similarly, it can be shown that no member of* $[G_J \cdot H_T]$ *is in* $[G_I \cdot H_S]$. *Thus, when* $I \neq J$, $[G_I \cdot H_S] \cap [G_J \cdot H_T] = \emptyset$. *Analogous results hold if* $S \neq T$. *Hence, K is a generator set.*

It remains to show that $[K] = \text{Sol}(P)$. *Let* $D \in [K]$. *By the definition of K, there exist* $G_I \in G$ *and* $H_J \in H$ *such that* $D \in [G_I \cdot H_J]$. *Hence, by Lemma* 1.13, $D = D_1 \cup D_2$, *where* $D_1 \in [G_I]$ *and* $D_2 \in [H_J]$. *By the premises of this theorem* D_1 *is an explanation for* M_1^+ *and* D_2 *is an explanation for* M_2^+. *By Lemma* 1.25(b), *it follows that D is an explanation for* M^+. *Thus,* $D \in \text{Sol}(P)$, *so* $[K] \subseteq \text{Sol}(P)$.

Let $D \in \text{Sol}(P)$. *Then, by Lemma* 1.25(b), *there exist* D_1 *and* D_2 *such that* D_1 *is an explanation for* M_1^+ *and* D_2 *is an explanation for* M_2^+, *and* $D = D_1 \cup D_2$. *Thus, there exist* $G_I \in G$ *and* $H_J \in H$ *such that* $D_1 \in [G_I]$ *and* $D_2 \in [H_J]$, *whence* $D = D_1 \cup D_2 \in [G_I \cdot H_J]$ *by Lemma* 1.13. *Thus,* $D \in [K]$, *so* $\text{Sol}(P) \subseteq [K]$.

By the preceding two paragraphs we therefore have $\text{Sol}(P) = [K]$.

This work is supported by grants to the University of Maryland from NIH (P50-NS-16332) and an NSF Award (DCR-8451430) with matching industrial support from Software A&E and AT&T Information Systems. Dr. Nau is the

recipient of an NSF Presidential Young Investigator Award (matching funds from General Motors, IBM, and Martin Marietta). Technical assistance in preparing this manuscript was provided by Ms. Gloria Kimbles.

REFERENCES

1. H. Barrows, The diagnostic (problem solving) skill of the neurologist, *Arch. Neurol.* 26:273–277 (1972).
2. 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. Pattern Anal. and Machine Intelligence* 2:148–160 (1980).
3. W. Card, The diagnostic process, in *Medical Computing* (Abrams, Ed.), American Elsevier, 1970.
4. N. Christofides and S. Korman, A computational survey of methods for the set covering problem, *Management Sci.* 21:591–599 (1975).
5. F. deDombal, Computer assisted diagnosis of abdominal pain, in *Advances in Medical Computing* (J. Rose and J. Mitchell, Eds.), Churchill-Livingston, New York, 1975, pp. 10–19.
6. J. Edwards, Coverings and packings in a family of sets, *Bull. Amer. Math. Soc.* 68:494–499 (1962).
7. A. Elstein, L. Shulman, and S. Sprafka, *Medical Problem Solving — An Analysis of Clinical Reasoning*, Harvard U.P., 1978.
8. R. Karp, Reducibility among combinatorial problems, in *Complexity of Computer Computations* (R. Miller and J. Thatcher, Eds.), Plenum, New York, 1972, pp. 85–103.
9. J. Kassirer and G. Gorry, Clinical problem solving—a behavioral analysis, *Ann. Int. Med.* 89:245–255 (1978).
10. L. Kingsland, G. Sharp, R. Capps, et al., Testing of a Criteria-Based Consultant System in Rheumatology, in *Proceedings of MEDINFO-83* (J. van Bemmel, M. Ball, and O. Wigertz, Eds.), North Holland, 1983, pp. 514–517.
11. D. Leaper et al., Clinical diagnostic process—an analysis, *British Med. J.* 3:569–574 (1973).
12. R. Miller, H. Pople, and J. Myers, INTERNIST-1, An experimental computer-based diagnostic consultant for general internal medicine, *New England J. Med.* 307:468–476 (1982).
13. D. Nau, G. Markowsky, M. Woodbury, and D. Amos, A mathematical analysis of human leukocyte antigen serology, *Math. Biosci*, 40:243–270 (1978).
14. D. Nau, J. Reggia, and P. Wang, Knowledge-based problem solving without production rules, in *Proceedings of the 1983 Trends and Applications Conference*, IEEE Computer Society Press, 1983, pp. 105–108.
15. N. Nilsson, *Principles of Artificial Intelligence*, Tioga, 1980, p. 37.
16. N. Nilsson, The interplay between experimental and theoretical methods in artificial intelligence, *Cognition and Brain Theory* 4:69–74 (1980).
17. S. Pauker, G. Gorry, J. Kassirer, and W. Schwartz, Towards the simulation of clinical cognition, *Amer. J. Med.* 60:981–996 (1976).
18. H. Pople, Heuristic methods for imposing structure on ill-structured problems: The structuring of medical diagnostics, in *Artificial Intelligence in Medicine* (P. Szolovits, Ed.), Westview Press, Boulder, Colo., 1982, pp. 119–190.

19. J. Reggia, A production rule system for neurological localization, in *Proceedings of the Second Annual Symposium on Computer Applications in Medical Care*, IEEE Press, 1978, pp. 254–260.
20. J. Reggia, Computer-assisted medical decision making, in *Applications of Computers in Medicine* (M. Schwartz, Ed.), IEEE Press, 1982, pp. 198–213.
21. J. Reggia, D. S. Nau, and P. Y. Wang, Diagnostic expert systems based on a set covering model, *Internat. J. Man-Machine Stud.* 19:437–460 (1983).
22. J. Reggia, D. Tabb, T. Price, et al., Computer-aided assessment of transient ischemic attacks: A clinical evaluation, *Arch. Neurol.*, 1984.
23. W. Rouse, Human problem solving performance in a fault diagnosis task, *IEEE Trans. Systems Man Cybernet.* 8:258–271 (1978).
24. W. Rouse, Problem solving performance of maintenance trainees in a fault diagnosis task, *Human Factors* 21:195–203 (1979).
25. A. Rubin, The role of hypotheses in medical diagnosis, in *Fourth International Joint Conference on Artificial Intelligence*, 1975, pp. 856–862.
26. E. Shortliffe, *Computer-Based Medical Consultations: MYCIN*, Elsevier, New York, 1976.
27. H. Shubin and J. Ulrich, IDT: An intelligent diagnostic tool, in *Proceedings of the National Conference on Artificial Intelligence*, AAAI, 1982, pp. 290–295.
28. M. Woodbury, E. Ciftan, and D. Amos, HLA serum screening based on an heuristic solution of the set cover problem, *Comput. Programs Biomed.* 9:263–273 (1979).
29. P. Wortman, Representation and strategy in diagnostic problem solving, *Human Factors* 8:48–53 (1966).
30. V. Yu, B. Buchanan, E. Shortliffe, et al., Evaluating the performance of a computer-based consultant, *Comput. Programs Biomed.* 9:95–102 (1979).
31. R. Zagoria and J. Reggia, Transferability of medical decision support systems based on Bayesian classification, *Med. Decision Making* 3:501–510 (1983).

Received 8 April 1985; revised 29 April 1985