PLAUSIBILITY OF DIAGNOSTIC HYPOTHESES: The Nature of Simplicity

Yun Peng and James A. Reggia
Department of Computer Science
University of Maryland
College Park, MD. 20742

Abstract

In general diagnostic problems multiple disorders can occur simultaneously. AI systems have traditionally handled the potential combinatorial explosion of possible hypotheses in such problems by focusing attention on a few "most plausible" ones. This raises the issue of establishing what makes one hypothesis more plausible than others. Typically a hypothesis (a set of disorders) must not only account for the given manifestations, but it must also satisfy some notion of simplicity (or coherency, or parsimony, etc) to be considered. While various criteria for simplicity have been proposed in the past, these have been based on intuitive and subjective grounds. In this paper, we address the issue of if and when several previously-proposed criteria of parsimony are reasonable in the sense that they are guaranteed to at least identify the most probable hypothesis. Hypothesis likelihood is calculated using a recent extension of Bayesian classification theory for multimembership classification in causal diagnostic domains. The significance of this result is that it is now possible to decide objectively *a priori* the appropriateness of different criteria for simplicity in developing an inference method for certain classes of general diagnostic problems.

1. Diagnostic Problem-Solving

During the last decade, a number of artificial intelligent (AI) systems have been developed that use an "abductive" * approach to diagnostic problem-solving [Pople73, 82] [Pauker76] [Reggia81, 83] [Miller82] [Josephson84] [Basili85]. These systems use an associative knowledge base where causal associations between disorders and manifestations are the central component, and inferences are made through a sequential hypothesize-and-test process. An important but as yet unresolved issue in abductive systems for diagnostic problem-solving is what characteristics make a set of disorders a plausible, "best", or "simplest" explanatory hypothesis for observed manifestations. This issue has long been an important one in philosophy [Peirce55] [Thagard78] [Josephson82] as well as in AI [Rubin75] [Pople73] [Pauker76] [Reggia83] [Josephson84], and is not only of relevance to diagnostic problem-solving but also to many other areas in AI (natural language processing, machine learning, etc. [Charniak85] [Reggia85a]). In particular, to the authors' knowledge, all previous suggestions of hypothesis plausibility have generally been proposed primarily on intuitive rather than formal grounds.

Over the last few years we have been studying a formal model of diagnostic problem-solving referred to as *parsimonious covering theory* [Reggia83,85b] [Peng86a]. Recently, we have successfully integrated into this causal reasoning model the ability to calculate the relative likelihood of any evolving or complete diagnostic hypothesis [Peng86b]. As a result an *objective* measure (relative likelihood) can now be used to examine several previous *subjective* criteria of hypothesis plausibility. The rest of this paper examines this issue, and is organized as follows. First, the parsimonious covering model of problem-solving, which is based on an underlying causal relationship and the use of probability theory in this context, are briefly summarized in Sections 2 and 3. Section 4 then examines several different criteria for hypothesis plausibility used in AI systems with respect to whether they lead to the *most probable* diagnostic hypothesis. Situations where the use of each criterion is/is not appropriate are identified. Section 5 concludes by summarizing the implications of these results for AI system development.

2. Parsimonious Covering Theory

Causal associations between disorders and manifestations are the central element of diagnostic knowledge bases in many real-world systems, and parsimonious covering theory is based on a formalization of causal associative knowledge [Peng86a] [Reggia85b]. The simplest type of diagnostic problems in this model, and the one we use in this paper, is defined to be a 4-tuple $P = <D,M,C,M^+>$ where

$D = \{d_1, \ldots, d_n\}$ is a finite non-empty set of disorders;

$M = \{m_1, \ldots, m_k\}$ is a finite non-empty set of manifestations (symptoms);

$C \subseteq D \times M$ is a relation with domain(C) = D and range(C) = M; and

$M^+ \subseteq M$ is a distinguished subset of M.

The relation C captures the intuitive notion of causal associations in a symbolic form, where $<d_i,m_j> \in C$ iff "disorder d_i may cause manifestation m_j". Note that $<d_i,m_j> \in C$ does not imply that m_j always occurs when d_i is present, but only that m_j may occur. D, M, and C together correspond to the knowledge base in an abductive expert system. M^+, a special subset of M, represents the features (manifestations) which are present

* Abductive inference is generally defined to be "reasoning to the best explanation" for a given set of facts, and is distinguished from deductive and inductive inference (see [Peirce55] [Thagard78] [Pople73] [Josephson82] [Charniak85] [Reggia85a]).

in a specific problem. Fig. 1 graphically illustrates the symbolic causal knowledge of a tiny abstract diagnostic problem of this type.

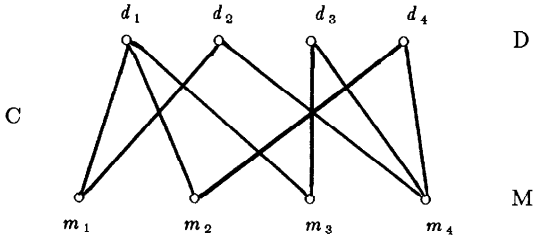

Fig. 1. An example of a very simple abstraction of a diagnostic problem.

Two functions, "causes" and "effects", can be defined in the above framework: for all $m_j \in M$, causes(m_j) = $\{d_i \mid <d_i, m_j> \in C\}$, representing all possible causes of manifestation m_j; for all $d_i \in D$, effects(d_i) = $\{m_j \mid <d_i, m_j> \in C\}$, representing all manifestations which may be caused by d_i. A set of disorders $D_I \subseteq D$ is then said to be a *cover* of a set of manifestations $M_J \subseteq M$ if $M_J \subseteq$ effects(D_I), where by definition effects(D_I) = $\bigcup_{d_i \in D_I}$ effects(d_i). Also, we define causes(M_J) = $\bigcup_{m_j \in M_J}$ causes(m_j).

In parsimonious covering theory, a diagnostic hypothesis must be a cover of M^+ in order to account for the presence of *all* manifestations in M^+. On the other hand, not all covers of M^+ are equally plausible as hypotheses for a given problem. The principle of parsimony, or "Occam's Razor", is adopted as a criterion of plausibility: a "simple" cover is preferable to a "complex" one. Therefore, a plausible hypothesis, called an *explanation* of M^+, is defined as a *parsimonious cover* of M^+, i.e., a set of disorders that both covers M^+ and satisfies some notion of being parsimonious or "simple". Since there is, in general, more than one possible explanation for M^+, and one is often interested in all plausible hypotheses, the set of all explanations of M^+ is defined to be the *solution* of a given problem.

A central question in this theory is thus: what is the nature of "parsimony" or "simplicity"? Put otherwise, what makes one cover of M^+ more plausible than another? A number of different parsimony criteria have been identified both by us and by others doing related work: (1) Single-Disorder Restriction: a cover D_I of M^+ is an explanation if it contains only a single disorder [Shubin82]. (2) Minimality: a cover D_I of M^+ is an explanation if it has the minimal cardinality among all covers of M^+, i.e., it contains the smallest possible number of disorders needed to cover M^+ [Pople73] [Reggia81, 83]. (3) Irredundancy: a cover D_I of M^+ is an explanation if it has no proper subsets which also cover M^+, i.e., removing any disorder from D_I results in a non-cover of M^+ [Nau84] [Reggia84,85b] [Peng86a] [Reiter85] [deKleer86]. (4) Relevancy: a cover D_I of M^+ is an explanation if it only contains disorders in causes(M^+), i.e., every $d_i \in D_I$ must be causally associated with some $m_j \in M^+$ [Peng86a]. Other criteria of parsimony are possible. Assuming at least one manifestation is present, single-disorder covers are minimal. Further, the set of all minimal covers is always contained in the set of all irredundant covers, which in turn is always contained in the set of all relevant covers [Peng86a].

example 1: In Fig. 1, let $M^+ = \{m_1, m_3\}$. Then $D_1 = \{d_1\}$ is a minimal cover of M^+ because it alone covers $\{m_1, m_3\}$. The cover $D_2 = \{d_2, d_3\}$ is irredundant but not minimal because neither d_2 nor d_3 alone can cover $\{m_1, m_3\}$. The cover $D_3 = \{d_1, d_2, d_3\}$ is relevant but redundant because it is a subset of causes($\{m_1, m_3\}$) and one of its proper subsets, namely $\{d_2, d_3\}$, is a cover of M^+. Finally, $D_4 = \{d_1, d_2, d_3, d_4\}$ is an irrelevant cover of M^+ because $d_4 \notin$ causes($\{m_1, m_3\}$).

The single-disorder restriction, while appropriate in some restricted domains [Shubin82] [Reggia85b], is obviously not sufficient for general diagnostic problems where multiple, simultaneous disorders can occur (and thus we will not consider it any further). Minimality captures features and assumptions of many previous abductive expert systems. However, our experience has convinced us that there are clearly cases where minimal covers are not necessarily the best ones. For example, suppose that either a very rare disorder d_1 alone, or a combination of two very common disorders d_2 and d_3, could cover all present manifestations. If minimality is chosen as the parsimony criterion, d_1 would be chosen as a viable hypothesis while the combination of d_2 and d_3 would be discarded. A human diagnostician, however, may consider the combination of d_2 and d_3 as a possible alternative. Minimality also suffers from various computational difficulties [Peng86a]. On the other hand, intuition also suggests that relevancy is too loose as a parsimony/plausibility criterion (in Fig. 1 there are only 2 irredundant covers, but 5 relevant ones among all 10 covers of $M' = \{m_1, m_3\}$). Therefore, solely on an intuitive basis, in our recent work irredundancy has been chosen as the parsimony criterion, and the notion of explanation equated to the notion of irredundant cover. Irredundancy handles situations like that in the above example and avoids some computational difficulties of minimality [Peng86a].

Similar notions are also used in related work by others, although with different emphasis. For example, in de Kleer's work, the notion of "minimal conflict" of an abnormal finding corresponds to causes(m_j), while a "minimal candidate" corresponds to an irredundant cover of M^+ in parsimonious covering theory [deKleer86]. Similarly, in Reiter's work, the notion of "minimal conflict set" corresponds to causes(m_j), "hitting set" to relevant cover, and "minimal hitting set" to irredundant cover [Reiter85]. One reason that we choose the term "irredundancy" rather than "minimality" is to avoid any confusion with the term "minimal cardinality".

3. Hypothesis Likelihood

An alternative approach to determining the plausibility of a diagnostic hypothesis is to *objectively* calculate its probability using formal probability theory. The

difficulty with this approach in the past has been that general diagnostic problems are *multimembership classification problems* [Ben-Bassat80]: multiple disorders can be present simultaneously. A hypothesis $D_I = \{d_1, d_2, \ldots, d_n\}$ represents the belief that disorders d_1 and d_2 and . . . and d_n are present, and that all $d_i \notin D_I$ are absent. Such problems are recognized to be very difficult to handle [Ben-Bassat80] [Charniak83]. Among other things, the set of $2^{|D|}$ diagnostic hypotheses that must be ranked in some fashion is incredibly large in most real-world applications (e.g., in medicine, even very constrained diagnostic problems may have $50 \leq |D| \leq 100$; see [Reggia83]).

Recently we have been successful in integrating formal probability theory into the framework of parsimonious covering theory in a way that overcomes these past difficulties [Peng86b]. This is achieved as follows. In the knowledge base, a prior probability p_i is associated with each $d_i \in D$ where $0 < p_i < 1$. A *causal strength* $0 < c_{ij} \leq 1$ is associated with each causal association $<d_i, m_j> \in C$ representing how frequently d_i causes m_j. For any $<d_i, m_j> \notin C$, c_{ij} is assumed to be zero. A very important point here is that $c_{ij} \neq P(m_j \mid d_i)$. The probability $c_{ij} = P(d_i \text{ causes } m_j \mid d_i)$ represents how frequently d_i *causes* m_j when d_i is present; the probability $P(m_j \mid d_i)$, which is what has been used in previous statistical diagnostic systems, represents how frequently m_j *occurs* when d_i is present. Since typically more than one disorder is capable of causing a given manifestation m_j, $P(m_j \mid d_i) \geq c_{ij}$. For example, if d_i cannot cause m_j at all, $c_{ij} = 0$, but $P(m_j \mid d_i) \geq 0$ because some other disorder present simultaneously with d_i may cause m_j.

By introducing the notion of causal strengths, and by assuming that disorders are independent of each other, that causal strengths are invariant (whenever d_i is present, it causes m_j with the probability c_{ij} regardless of other disorders that are present), and that no manifestation can occur without being caused by some disorder, a careful analysis derives a formula for $P(D_I \mid M^+)$, the posterior probability of any D_I given the presence of any M^+, from formal probability theory. Here D_I, representing a hypothesis, denotes the event that all disorders in D_I are present and all other disorders absent, while M^+, representing the given findings, denotes that all manifestations in M^+ are present and all others absent [Peng86b]. Specifically, we have proven that manifestations are independent under a given D_I, and that $P(m_j \mid D_I) = 1 - \prod_{d_i \in D_I} (1 - c_{ij})$ for $m_j \in M$, $D_I \subseteq D$. Then by Bayes' theorem, it is easy to show that

$$P(D_I \mid M^+) = \frac{\prod_{d_i \in D} (1 - p_i)}{P(M^+)} \cdot L(D_I, M^+) \qquad [1]$$

where $\prod_{d_i \in D} (1 - p_i) \;/\; P(M^+)$ is a constant for all D_I given any M^+. $L(D_I, M^+)$, called the relative likelihood of D_I given M^+, consists of three components:

$$L(D_I, M^+) = L_1(D_I, M^+) \cdot L_2(D_I, M^+) \cdot L_3(D_I, M^+), \qquad [2a]$$

where the first product

$$L_1(D_I, M^+) = \prod_{m_j \in M^+} P(m_j \mid D_I) = \prod_{m_j \in M^+} \left(1 - \prod_{d_i \in D_I} (1 - c_{ij})\right) \qquad [2b]$$

informally can be thought of as a weight reflecting how likely D_I is to cause the presence of manifestations in the given M^+; the second product

$$L_2(D_I, M^+) = \prod_{m_l \in M\text{-}M^+} P(\overline{m_l} \mid D_I) = \prod_{d_i \in D_I} \prod_{m_l \in \text{effects}(d_i)\text{-}M^+} (1 - c_{il}) \qquad [2c]$$

can be viewed as a weight based on manifestations expected with D_I but which are actually absent; and the third product

$$L_3(D_I, M^+) = \prod_{d_i \in D_I} \frac{p_i}{(1 - p_i)} \qquad [2d]$$

represents a weight based on prior probabilities of D_I [Peng86b]. Note that each of these products involves only probabilistic information related to $d_i \in D_I$ and $m_j \in M^+$ instead of the entire knowledge base. For this reason $L(D_I, M^+)$ is computationally very tractable.

Eqs 1 and 2a - d make it possible to compare the relative likelihood of any two diagnostic hypotheses D_I and D_J using

$$\frac{P(D_I \mid M^+)}{P(D_J \mid M^+)} = \frac{L(D_I, M^+)}{L(D_J, M^+)}. \qquad [3]$$

Before we use this objective measure to examine various subjective notions of plausibility, a brief example may be helpful.

example 2: Let the following probabilities be assigned to the problem given in Fig. 1:

$p_1 = .01$	$p_2 = .1$	$p_3 = .2$	$p_4 = .2$
$c_{11} = .2$	$c_{12} = .8$	$c_{13} = .1$	$c_{14} = 0$
$c_{21} = .9$	$c_{22} = 0$	$c_{23} = 0$	$c_{24} = .3$
$c_{31} = 0$	$c_{32} = 0$	$c_{33} = .9$	$c_{34} = .2$
$c_{41} = 0$	$c_{42} = .5$	$c_{43} = 0$	$c_{44} = .8$

Let $M^+ = \{m_1, m_3\}$. Then the relative likelihood of three covers of M^+, $\{d_1\}$, $\{d_2, d_3\}$, and $\{d_1, d_2, d_3\}$, are calculated as follows.

$L_1(\{d_1\}, \{m_1, m_3\}) = c_{11} \cdot c_{13} = .2 \cdot .1 = .02$

$L_2(\{d_1\}, \{m_1, m_3\}) = (1 - c_{12}) \cdot (1 - c_{14}) = (1 - .8) \cdot 1 = .20$

$L_3(\{d_1\}, \{m_1, m_3\}) = \frac{p_1}{1 - p_1} = .01$. Similarly,

$L_1(\{d_2, d_3\}, \{m_1, m_3\}) = (1 - (1 - c_{21}) \cdot (1 - c_{31})) \cdot (1 - (1 - c_{23}) \cdot (1 - c_{33})) = .9 \cdot .9 = .81$

$L_2(\{d_2, d_3\}, \{m_1, m_3\}) = (1 - c_{24}) \cdot (1 - c_{34}) = .7 \cdot .8 = .56$

$L_3(\{d_2, d_3\}, \{m_1, m_3\}) = \frac{p_2}{1 - p_2} \cdot \frac{p_3}{1 - p_3} = .028$. Similarly,

$L_1(\{d_1, d_2, d_3\}, \{m_1, m_3\}) = (1 - (1 - c_{11}) \cdot (1 - c_{21}) \cdot (1 - c_{31})) \cdot (1 - (1 - c_{13}) \cdot (1 - c_{23}) \cdot (1 - c_{33})) = .84$

$L_2(\{d_1, d_2, d_3\}, \{m_1, m_3\}) = (1 - c_{12}) \cdot (1 - c_{24}) \cdot (1 - c_{34}) = .11$

$L_3(\{d_1, d_2, d_3\}, \{m_1, m_3\}) = .00028$.

Thus, $L(\{d_1\}, \{m_1, m_3\}) = .00004$, $L(\{d_2, d_3\}, \{m_1, m_3\}) = .013$, and $L(\{d_1, d_2, d_3\}, \{m_1, m_3\}) = .000026$, by Eq. 2a.

4. Hypothesis Plausibility

As noted earlier, parsimonious covering theory (as well as the work of others cited earlier) captures the basic notion used in many abductive problem-solvers that a set of disorders D_I is an "explanation" (plausible hypothesis) for M^+ if (1) D_I covers M^+, and (2) D_I is "parsimonious". We now examine these intuitive/subjective criteria using the measure $L(D_I, M^+)$ given above, focusing on the question of when a set of parsimonious covers includes the most probable cover.

First, suppose a hypothesis $D_I \subseteq D$ is *not* a cover of M^+. Then there exists at least one present manifestation, say $m_j \in M^+$, that is not covered by D_I, i.e., for all $d_i \in D_I$, $<m_j, d_i> \notin C$ so $c_{ij} = 0$. Then, $L_1(D_I, M^+) = 0$ and hence $L(D_I, M^+) = 0$ (by Eqs. 2b and 2a). That is, any $D_I \subseteq D$ which does not cover M^+ will have zero relative likelihood, and $P(D_I \mid M^+) = 0$. It thus follows that any most likely set of disorders D_I must be a cover of M^+, and that in search for plausible hypotheses only those sets that are covers of M^+ need to be considered (an important savings because usually a large number of D_K in 2^D are not covers).

The more difficult issue in hypothesis evaluation, however, has been precisely defining what is meant by the "best" or "most plausible" explanation for a given set of facts [Thagard78] [Josephson82] [Reggia85c] [Peng86a]. In the context of diagnostic problem-solving, it seems reasonable to correlate such subjective and ill-defined concepts with likelihood, i.e., to prefer diagnostic hypotheses that are more likely to be true based on their posterior probabilities. If one accepts $P(D_I \mid M^+)$ as a measure of the plausibility of D_I, it then becomes possible to objectively analyze the conditions under which different criteria of parsimony seem plausible. Three such criteria were defined in section 2, namely, relevancy, irredundancy, and minimality, and we now wish to consider if and when these criteria identify the most probable diagnostic hypothesis.

Let $D_I \in D$ be a cover of M^+ in a diagnostic problem $P = <D,M,C,M^+>$. For any $d_k \in D - D_I$, it follows from Eqs. 2b - d that

$$L_1(D_I \cup \{d_k\}, M^+)$$

$$= L_1(D_I, M^+) \cdot \prod_{m_j \in effects(d_k) \cap M^+} (1 - c_{kj} + \frac{c_{kj}}{P(m_j \mid D_I)}) \qquad [4a]$$

where $P(m_j \mid D_I) = 1 - \prod_{d_i \in D_I} (1 - c_{ij}) \neq 0$ for all $m_j \in M^+$ since D_I covers M^+.

For $L_2(D_I \cup \{d_k\}, M^+)$ and $L_3(D_I \cup \{d_k\}, M^+)$, it is similarly the case that

$$L_2(D_I \cup \{d_k\}, M^+) = L_2(D_I, M^+) \cdot \prod_{m_i \in effects(d_k) - M^+} (1 - c_{kj}). \qquad [4b]$$

$$L_3(D_I \cup \{d_k\}, M^+) = L_3(D_I, M^+) \cdot \frac{p_k}{1 - p_k}. \qquad [4c]$$

Eqs. 4a - c directly support the analysis of the three types of parsimony in question by permitting the direct comparison of $L(D_I, M^+)$ and $L(D_I \cup \{d_k\}, M^+)$.

relevant covers: Let D_I be a relevant cover of M^+, so by definition D_I covers M^+ and $D_I \subseteq causes(M^+)$. Let $d_k \notin causes(M^+)$, i.e., d_k is irrelevant to M^+, so $d_k \notin D_I$. Then, $D_I \cup \{d_k\}$ is an irrelevant cover of M^+. For such a d_k, all of its manifestations are known to be absent, so using Eqs. 4a, b and c, it follows from the preceding that $\frac{L(D_I \cup \{d_k\}, M^+)}{L(D_I, M^+)} = (\prod_{m_i \in effects(d_k)} (1 - c_{ki})) \cdot \frac{p_k}{1 - p_k}$ because $L_1(D_I \cup \{d_k\}, M^+) = L_1(D_I, M^+)$.

In most real world diagnostic problems, p_k is generally very small. For example, in medicine $p_k < 10^{-1}$ even for very common disorders in the general population, such as a cold or the flu, and is much much smaller (e.g., 10^{-6}) for rare disorders. Thus, $p_k/(1 - p_k) << 1$ usually. The product of $(1 - c_{ki})$'s is also less than 1, and is often much less since it is a product of numbers less than one. Thus, in most applications, $L(D_I \cup \{d_k\}, M^+) << L(D_I, M^+)$ making an irrelevant cover much less likely than any relevant cover it contains. This effect is magnified as a cover becomes "more irrelevant", i.e., as additional irrelevant disorders d_i are included. Thus, generally, it is only necessary to generate relevant covers as hypotheses for which $L(D_I, M^+)$ is calculated, and in most real world problems this represents an enormous computational savings (typically most covers are irrelevant). The only exception would occur when p_k is fairly large, and d_k has few, weakly causal associations with its manifestations. In particular, $L(D_I \cup \{d_k\}, M^+)$ would exceed $L(D_I, M^+)$ where d_k was an irrelevant disorder only if $p_k > 1 / (1 + \prod_{m_i \in effects(d_k)} (1 - c_{ki})) > 0.5$, a distinctly atypical situation as noted earlier. An interesting consequence of this result is that if $M^+ = \emptyset$, since $\emptyset$ is the only relevant cover of such a M^+, the probabilistic causal model generally entails "no disorders are present" as the only reasonable explanation, provided that $p_i \leq 0.5$ for all $d_i \in D_I$. This is consistent with parsimonious covering theory and with intuition.

irredundant covers: If D_I is an irredundant cover of M^+, then by definition no proper subset of D_I covers M^+. For $d_k \notin D_I$ but $d_k \in causes(M^+)$, $D_I \cup \{d_k\}$ is a redundant but relevant cover of M^+. From Eqs. 4a - c, $\frac{L_1(D_I \cup \{d_k\}, M^+)}{L_1(D_I, M^+)} \geq 1$ and $\frac{L_2(D_I \cup \{d_k\}, M^+)}{L_2(D_I, M^+)} \leq 1$. If $p_k << 1$, then $\frac{L_3(D_I \cup \{d_k\}, M^+)}{L_3(D_I, M^+)} = \frac{p_k}{1 - p_k} << 1$. In general it is likely that the decrease in L_2 and L_3 caused by adding d_k will compensate for the increase in L_1 because p_k is typically small. As example 2 shows, although adding d_1 into irredundant cover $\{d_2, d_3\}$ increases L_1 from .81 to .84, it reduces L_2 from .56 to .11 and L_3 from .028 to .00028, thus making the redundant but relevant cover $\{d_1, d_2, d_3\}$ much less likely than the irredundant cover $\{d_2, d_3\}$ (.000026 vs. .013). Therefore, if the prior probabilities $p_i << 1$ for all $d_i \in D$ as in many applications, the most probable covers of M^+ are likely to be irredundant covers, consistent with intuitive arguments made in the past [Nau84] [Reggia85] [Peng86a] [Reiter85] [deKleer86].

However, more care must be applied in restricting hypothesis generation to just irredundant covers. A careful analysis of Eqs. 4a - c should convince the reader that a redundant but relevant cover $D_I \cup \{d_k\}$ might

occasionally be more likely than D_I if d_k is fairly common and $c_{kj} >> P(m_j \mid D_I)$ for some $m_j \in M^+$. This is an intuitively reasonable result, and it represents an insight concerning the nature of "parsimony" that was only recognized after developing the probability calculus summarized in Section 3. However, even in the situation where some redundant cover is more probable than an irredundant cover it contains, such a redundant cover might still be less probable than the most probable irredundant cover. For instance, in example 2, a redundant cover $\{d_1,d_3\}$ has relative likelihood $L(\{d_1,d_3\}, \{m_1,m_3\}) = .000064$ which is greater that of $\{d_1\}$, but still less than that of $\{d_2,d_3\}$ which is an irredundant cover.

minimal covers: It is possible to identify situations where minimal cardinality is a reasonable criterion for hypothesis generation. For example, if, for all $d_i \in D$, the prior probabilities are $p_i << 1$ and are about equal, and the c_{ij}'s are fairly large in general, then a careful analysis of Eqs. 4a - c shows that the most probable covers of M^+ are likely to be minimal covers. In this situation, the ratio between $L(D_I, M^+)$ and $L(D_J, M^+)$ for two different covers D_I and D_J of M^+ will be dominated by the ratio $\frac{L_3(D_J, M^+)}{L_3(D_I, M^+)} \approx (\frac{p_i}{1 - p_i})^{|D_J| - |D_I|}$ which would be very much smaller than 1 if $|D_I| < |D_J|$. Unfortunately, in many real-world diagnostic situations the assumptions needed to make minimality a useful parsimony criterion are violated. In medicine, for example, prior probabilities among diseases and causal strengths vary by as much as 10^6, and therefore minimality is generally not a reasonable criterion to adopt to limit hypothesis generation.

5. Discussion

By applying a form of Bayesian classification extended to work in the framework of parsimonious covering theory, we have been able to examine various intuitive/subjective criteria for hypothesis plausibility in an objective fashion. Consistent with intuition and concepts in parsimonious covering theory, probability theory leads to the conclusion that a set of disorders must be a cover to be a plausible hypothesis. Further, conditions can now be stated (Section 4) for when various criteria of "simplicity" are reasonable heuristics for judging plausibility. For example, minimal cardinality is only appropriate to consider when all disorders are very uncommon and of about equal probability, and causal strengths are fairly large. If some disorders are relatively much more common than others, or if causal strengths in some cases are fairly weak, using minimal cardinality as a heuristic to select plausible diagnostic hypotheses is inadequate. In this latter situation, typical of most real-world problems, the criterion of irredundancy may be appropriate.

Irredundancy is generally quite attractive as a plausibility criterion for diagnostic hypotheses, and a formal algorithm (with proof of its correctness) for generating all irredundant covers of a set of given manifestations M^+ has recently been described [Peng86a]. Unfortunately, there are two difficulties with directly generating the set of all irredundant covers for consideration as diagnostic hypotheses. First, this set may itself be quite large in some applications, and may contain many hypotheses of very low probability. Second, and more serious, it may still miss identifying the most probable diagnostic hypothesis in some cases (see Section 4). This latter difficulty is an insight concerning plausibility criteria that has not been previously recognized.

Fortunately, both difficulties are surmountable. A heuristic function based on a modification of $L(D_I, M^+)$ can be used to guide an A^*-like algorithm to first locate a few most likely irredundant covers for M^+. Then, a typically small amount of additional search of the "neighborhood" of each of these irredundant covers can be done to see if any relevant but redundant covers are more likely. An algorithm to do this and a proof that it is guaranteed to always identify the most likely diagnostic hypothesis has been presented in detail elsewhere [Peng86b].

There are a number of generalizations that could be made to the results presented in this paper, and we view these as important directions for further research. Our use of Bayesian classification with a causal model assumed that disorders occur independently of one another. In some diagnostic problems this is unrealistic, so a logical extension of this work would be to generalize it to such problems. Some work has already been done along these lines in setting bounds on the relative likelihood of disorders with Bayesian classification [Cooper84]. In addition, we have adapted only one method of ranking hypotheses (Bayes' Theorem) to work in causal domains involving multiple simultaneous disorders. It may be that with suitable analysis other approaches to ranking hypotheses could also be adopted in a similar fashion (e.g., Dempster-Shafer theory [Dempster68][Shafer76]). Some initial work along these lines with fuzzy measures has already been done [Yager85].

Supported in part by ONR award N00014-85-K0390 and by NSF Award DCR-8451430 with matching funds from Software A&E, AT&T Information Systems, and Allied Corporation Foundation.

REFERENCES

[1] Basili, V., and Ramsey, C., "ARROWSMITH - P: A Prototype Expert System for Software Engineering Management", *Proc. Expert Systems in Government Symposium*", Karna, K., (ed.), Mclean, VA, 1985.

[2] Ben-Bassat, M., et al, "Pattern-Based Interactive Diagnosis of Multiple Disorders: The MEDAS System", *IEEE Transactions on Pattern Analysis and Machine Intelligence,* **2**, March 1980, pp. 148-160.

[3] Charniak, E., "The Bayesian Basis of Common Sense Medical Diagnosis", *Proc. of National Conference on Artificial Intelligence,* AAAI, 1983, pp. 70-73.

[4] Charniak, E., and McDermott, D., *Introduction to Artificial Intelligence,* Addision-Wesley, Reading, MA., 1985, chapters 8, 10.

[5] Cooper, G., *NESTOR: A Computer-Based Medical Diagnostic Aid That Integrates Causal and Probabilistic Knowledge,* STAN-CS-84-1031 (Ph. D. Dissertation), Dept. of Computer Science, Stanford University, Nov. 1984.

[6] de Kleer, J., and Williams, B., "Reasoning about Multiple Faults", submitted, 1986.

[7] Dempster, A., "A Generalization of Bayesian Inference", *Journal of Roy. Statis. Sci. Ser.* **B30**, 1968, pp. 205-247.

[8] Josephson, J., *Explanation and Induction,* Ph. D. Thesis, Dept. of Phil., Ohio State Univ., 1982.

[9] Josephson, J., Chadrasekaran, B., and Smith, J., "Assembling the Best Explanation", *IEEE Workshop on Principles of Knowledge-Based Systems,* Denver, CO. Dec., 1984.

[10] Miller, R., Pople, H., and Myers, J., "INTERNIST-1, An Experimental Computer-Based Diagnostic Consultant for General Internal Medicine", *New England J. of Medicine,* **307**, 1982, 468-476.

[11] Nau, D., and Reggia, J., "Relationship Between Deductive and Abductive Inference in Knowledge-Based Diagnostic Problem Solving", *Proc. First Intl. Workshop on Expert Database Systems,* Kerschberge, L. (ed.), Kiawah Island, Sc., Oct, 1984, pp. 500-509.

[12] Pauker, S., Gorry, G., Kassirer, J., and Schwartz, M., "Towards the Simulation of Clinical Cognition", *Am. J. Med.,* **60**, *1976, pp. 981-996.*

[13] Peirce, C., *Abduction and Induction,* Dover, 1955.

[14] Peng, Y., *A Formalization of Parsimonious Covering and Probabilistic Reasoning in Abductive Diagnostic Inference,* Technical Report TR-1615 (Ph. D. Dissertation), Dept. of Computer Science, University of Maryland, Jan. 1986a.

[15] Peng, Y., and Reggia, J., "A Probabilistic Causal Model for Diagnostic Problem-Solving", submitted for publication, 1986b.

[16] Pople, H., "On the Mechanization of Abductive Logic", *Proc. of International Joint Conference on Artificial Intelligence,* IJCAI, 1973, pp. 147-152.

[17] Pople, H., "Heuristic Methods for Imposing Structure on Ill-structured Problems: The Structuring of Medical Diagnostics", *Artificial Intelligence in Medicine,* Szolovits, P. (ed.), 1982, pp. 119-190.

[18] Reggia, J., *Knowledge-Based Decision Support Systems: Development Through KMS,* Technical Report TR-1121 (Ph. D. Dissertation), Dept. of Computer Science, University of Maryland, Oct., 1981.

[19] Reggia, J., Nau, D., and Wang, P., "Diagnostic Expert Systems Based on a Set Covering Model", *Int. J. Man-Machine Studies,* Nov. 1983, pp. 437-460.

[20] Reggia, J., and Nau, D., "An Abductive Non-Monotonic Logic", *Proc. Workshop on Non-Monotonic Reasoning,* AAAI, Oct. 1984, pp. 385-395.

[21] Reggia, J., "Abductive Inference", *Expert Systems in Government Symposium,* Oct. 1985a, pp. 484-489.

[22] Reggia, J., Nau, D., Wang, P., and Peng, Y., "A Formal Model of Diagnostic Inference", *Information Sciences,* **37**, 1985b, pp. 227-285.

[23] Reiter, R., "A Theory of Diagnosis from First Principles", TR-187/86, Dept. of Computer Science, University of Toronto, Dec. 1985.

[24] Rubin, A., "The Role of Hypotheses in Medical Diagnosis", *Proc. of International Joint Conference on Artificial Intelligence,* IJCAI, 1975, pp. 856-862.

[25] Shafer, G., *A Mathematical Theory of Evidence, Princeton University Press,* Princeton, NJ, 1976.

[26] Shubin, H., and Ulrich, J., "IDT: An Intelligent Diagnostic Tool", *Proc. National Conference on Artificial Intelligence,* AAAI, 1982, pp. 290-295.

[27] Thagard, P., "The Best Explanation - Criteria for Theory Choice", *Journal of Philosophy,* **75**, 1978, pp. 76-92.

[28] Yager, R., "Explanatory Models in Expert Systems", *Int. Journal of Man - Machine Studies,* **23**, 1985, pp. 539-549.