October 20, 1977

# PROSPECTOR -- A Computer-Based Consultation System for Mineral Exploration

by

P. E. Hart and R. O. Duda

Artificial Intelligence Center
SRI International
Menlo Park, California 94025

Technical Note No. 155

Presented at the Taita Hills Conference on Standards for Computer Applications in Resource Studies (Taita Hills, Kenya, November 8-15, 1977); to be published in <u>International Association for Mathematical Geology</u> (Vol. 10, Nos. 5-6).

## ACKNOWLEDGMENTS

This work was supported in part by the Office of Resource Analysis of the U. S. Geological Survey under Contract 14-08-0001-15985, and in part by the RANN Division of the National Science Foundation under Grant AER77-04499. Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors, and do not necessarily reflect the views of either the U. S. Geological Survey or the National Science Foundation. The work was carried out by an interdisciplinary team of geologists and computer scientists that included Nils Nilsson, Rene Reboh, Jonathan Slocum, Georgia Sutherland, Alan Campbell, Neil Campbell, Marco Einaudi, and Charles Park, Jr. Additional help and suggestions were provided by Gerald Askevold, Joseph Botbol, Donald Singer, and Edwin Tooker. We are grateful to all of these contributors.

# I INTRODUCTION

Our purpose in this paper is to describe a large computer program called PROSPECTOR that is being developed to provide active consultation on problems of mineral exploration and resource evaluation. Three primary considerations motivated us to develop PROSPECTOR: the need for continuing exploration for mineral resources, the difficulty of staying technically abreast of an expanding technical discipline, and the desirability of bringing to bear the knowledge of several specialists on a given resource problem. Recent advances in computer-based consultation systems provided the technical basis for believing that a system like PROSPECTOR can contribute to solving these problems.

We envision two different modes of use for such a system. In the first mode, an exploration geologist starts by telling the program the characteristics of a particular prospect of interest -- the geologic setting, structural controls, and kinds of rocks, minerals, and alteration products present or suspected. The program compares these observations with models of various kinds of ore deposits, noting the similarities, differences, and missing information. The program then engages the geologist in a dialog to obtain additional relevant information and to make an assessment of the mineral potential of the prospect. Our goal here is to provide the geologist with a service comparable to giving him telephone access to authorities on many different kinds of ore deposits. In many cases, the main benefit of the consultation may be to alert the geologist to unsuspected possibilities, and to establish the additional observations that would be most valuable for further exploration.

In the second mode, the program would "talk" not to a person but to a large data base describing mineralized properties. Here the goal would be to screen the data base, either to select for particular commodities or to make regional assessments of mineral resources. In general, this is a more difficult mode of operation, primarily because the facts recorded in data bases often require interpretation; simple mechanical accessing may fail to provide an answer, even though a trained geologist would recognize that relevant information was present. Thus, although we recognize the value of this mode of use, we have begun by addressing the problem of providing interactive consultation to an exploration geologist.

The ability of PROSPECTOR to provide expert consultation rests on a base of knowledge about economic geology. This "knowledge base" has several components, the most important of which are the models that contain geological (and eventually geochemical and geophysical) information relevant to exploration for various classes of ore deposits. The program currently contains three different exploration models, one for Kuroko-type massive sulfide deposits, one for Mississippi-Valley-type carbonate lead/zinc deposits, and one for a major class of near-continental-margin porphyry copper deposits. The models are stored in the computer in a special way, to be described later, that enables PROSPECTOR to use them to reason about geological data. In addition, the overall system has been designed in such a way that its competence can be continually improved by the incorporation of additional models.

Mineral exploration is perhaps as much art as science, and the state of this art does not admit the construction of models as rigorous and complete as, say, those of Newtonian mechanics. This state of affairs has two important effects on the design of PROSPECTOR. First, the system must accommodate plausible or probabilistic styles of reasoning. Second, the models often reflect the subjective judgements of expert economic geologists more than objectively derivable facts. Of course, the use of subjective judgements and probabilities to make technical evaluations is not unique to PROSPECTOR. Subjective probabilities have been used in resource evaluation [Harris et al., 1970; see Harris, 1977 for a comprehensive treatment], while panels of experts have been frequently used in Delphi studies [Linstone and

Turoff, 1975; Ellis et al., 1975] to forecast technological events. What is unique is PROSPECTOR's combination of plausible and logical reasoning using a "knowledge base" supplied by experts to provide a computer-based consultation service.

As the foregoing description suggests, three different groups of people are involved in the design and use of PROSPECTOR: computer scientists, who design the computer programs that provide the framework of the system; expert economic geologists, who provide the knowledge base for the system; and the end-users of the system who are seeking consultation about particular prospects of interest to them. We should emphasize here that the end-user is expected to be a trained geologist, not a layman. PROSPECTOR cannot make direct observations and must therefore depend upon the skill of the user in this regard.

To the best of our knowledge, PROSPECTOR represents the first attempt to build a computer system able to consult actively on problems of mineral exploration. The general notion of a computer-based consultation system, however, has been explored before. Procedures for performing the required plausible and logical reasoning have been developed through computer science research on artificial intelligence [Nilsson, 1971; Raphael, 1976]. These procedures have been applied in several fields, with the most advanced development being in the field of medicine. In particular, computer-based consultation systems have been developed for infectious diseases [Shortliffe, 1976], for glaucoma [Weiss et al., 1977], for kidney disease [Pauker et al., 1976], and for a substantial part of internal medicine [Pople, 1977]. While these systems are rather recent and are still experimental, some have already attained levels of competence that in their specialties rival the performance of expert physicians. These successes provided an additional motivation for us to achieve similar performance in the area of mineral exploration.

The remainder of this paper is organized as follows: Section II presents an example of an interactive computer session with PROSPECTOR as it currently exists, calling attention to some of the features of the design. Section III describes some technical details regarding the form in which geological knowledge is stored and the mechanisms for drawing plausible inferences from specific observations. Section IV concludes with some comparisons with other approaches to the general problem, and some observations on current and prospective work.

## II AN EXAMPLE

This section contains an example of an interactive session with the current PROSPECTOR system. The data presented reflect an actual prospect in Canada; however, our purpose in presenting this example is not to discuss the geology of the particular prospect, but rather to illustrate as simply as possible some of the features of PROSPECTOR.

The transcript of the consultation session has been lightly edited in the interest of brevity. The user's input is underlined, and some explanatory comments (enclosed in square brackets) have also been added.

[After logging on to the computer and requesting access, the user receives the following.]

```
        PROSPECTOR ...... Mark II ...... Version 2
                  13-Oct-77 10:06:26
```

Do you want instructions? <u>YES</u>

PROSPECTOR, an experimental computer based consulting system, is designed for use both in searching for ore deposits and in evaluating the mineral potential of large geographic areas. Being in the early research stages, PROSPECTOR is based on only three models, C. F. Park's model for Kuroko-type massive sulfide deposits, N. Campbell's model for a class of Mississippi-Valley-type lead/zinc deposits, and M. Einaudi's model for a class of near-continental-margin porphyry-copper deposits.

In using PROSPECTOR, you will be asked to answer questions about the existence of various kinds of evidence. Answers to most questions will be numbers from -5 to 5. These numbers indicate the certainty of your response, with 5 indicating that the requested evidence is definitely present, -5 that it is definitely absent, and 0 that you have no opinion one way or the other.

In addition to supplying answers, you can respond at any time with the following commands:

```
          ? -- Prints a rephrased version of the question, often with
               additional comments or suggestions.
        WHY -- Prints the reason that PROSPECTOR asked the question.
  SUMMARIZE -- Prints a summary of the conclusions at this point in the
```

consultation.

VOL -- Lets you volunteer information to PROSPECTOR.

CHANGE -- Lets you change your answer to a previous question.

COMMENT -- Lets you type comments which will appear in your transcript file. (Two consecutive carriage returns will end the comment.)

WAIT -- Interrupts the session to allow modification of the program. (Should be used by qualified users only.)

QUIT -- Terminate execution of the program.

HELP -- Prints this explanation.

Do you wish to see intermediate results? <u>NO</u>

Do you want to volunteer any information? <u>YES</u>

When you are finished volunteering, please type DONE
Do you want to see the instructions? <u>YES</u>

PROSPECTOR can accept certain simple English statements that describe the prospect. You can name the types and ages of the rocks present, and the minerals and alteration products that are present or suspected. You can indicate uncertainty either through wording such as "There might be sphalerite," or by a more terse statement such as "Sphalerite, 3." When you are finished volunteering, please type DONE.

1:<u>THERE ARE CARBONATES</u>
(CARBONATES) (4)

[The computer acknowledges the first volunteered evidence; the certainty is taken to be 4.]

2:<u>THERE IS DOLOMITE</u>
(DOLOMITE) (4)

3:<u>THERE IS GALENA</u>
(GALENA) (4)

4:<u>THERE MIGHT BE SPHALERITE</u>
(SPHALERITE) (2)

5:<u>DONE</u>

[The user is finished volunteering evidence, and PROSPECTOR takes the initiative.]

--------------------

Models are (PCDA MSD MVTD)   [These are abbreviations of the three current models.]

...
...

[A few initial questions pursued hypotheses (i.e., models) that were ruled out. We continue the session with the hypothesis that proved to be the best one available.]

...
...

You told me about:
- CARBONATES (4)   [The computer reminds the user of volunteered evidence.]

6--To what degree do you believe that:
(* Prospective ore body lies in carbonate sediments) ? 3

You told me about:
- GALENA (4)

You suspected:
- SPHALERITE (2)

Your statements imply:
- Sulfides (4)

7--To what degree do you believe that:
(* Prospective ore body contains galena and sphalerite) ? 2

You told me about:
- DOLOMITE (4)
- CARBONATES (4)

8--To what degree do you believe that:
(* Prospective ore body involves dolomite) ? ?

[The user responds to the question with "?" and a clarification of the question is furnished.]

* Assuming that there is an ore body in carbonate sediments, to what degree do you believe that some significant dolomitization of the host carbonates has occurred?
Note: the exact type of dolomitization is not important. 3

[The user responds now with a certainty measure of 3]

9--To what degree do you believe that:
(* Fluid inclusion data suggest that mineralization took place above 160 degrees C) ? 0

10--To what degree do you believe that:
(* Prospective ore body contains non-telethermal alteration products) ? 4

11--To what degree do you believe that:
(* Other evidence indicates that mineralization took place above 160 degrees C) ? 0

12--To what degree do you believe that:
(* Evidence of metamorphism) ? WHY

[The user is interested in the reasoning behind this question.]

* The evidence of high temperature mineralization is discouraging for the prospective ore body being a MVTD. However, if the high temperatures were due to subsequent metamorphism, then this discouraging evidence should be discounted or ignored. To what degree do you believe that:
(* Evidence of metamorphism) ? -4
Your statements imply:
- Sulfides (4)

13--To what degree do you believe that:
(* Prospect contains sulfide minerals other than those of Fe, Pb and Zn) ? -3

14--To what degree do you believe that:
(* Any prospect assay indicates more than one ounce per ton of silver) ? 0

15--To what degree do you believe that:
(* Prospect contains fragmental host rock) ? 5

16--To what degree do you believe that:
(* Ore mineralization is stratabound) ? 0

17--To what degree do you believe that:
(* An unconformity or disconformity, or several of them, overlie the prospective ore body) ? 0

18--To what degree do you believe that:
(* Favorable stratigraphic sequence for a Mississippi Valley type deposit) ? ?

[The question requires clarification.]

* The stratigraphic sequence almost always found with a MVTD is as follows: sandstone overlain by a carbonate sequence containing the ore bearing limestone and dolomite, in turn overlain by shales or, more rarely, by a dense impermeable limestone or dolomite. Ignoring faulting, erosion, or subsequent deposition of material, to what degree do you believe that this describes the original stratigraphic sequence? 2

My certainty in (* Suggestive evidence for MVTD) is now: 4.997
Do you want to pursue this further? YES

> [The MVTD model is defined in terms of essential evidence and "suggestive evidence." Since the belief in suggestive evidence is nearly at its maximum value of 5, and since PROSPECTOR realizes that subsequent questions cannot affect this portion of the MVTD model by a substantial amount, the user is given the option of discontinuing this portion of the consultation.]

You told me about:
- CARBONATES (4)

19--To what degree do you believe that:
(* Carbonate sediments containing the prospective ore body are shallow-water shelf-type sediments deposited under stable tectonic conditions) ? 3

20--To what degree do you believe that:
(* Presence of hydrocarbons) ? -5

21--To what degree do you believe that:
(* Other evidence indicates that the prospect has characteristics which favor hydrocarbon accumulation) ? 2

22--To what degree do you believe that:
(* Prospect has positive geologic features) ? ?

* Positive geologic features include domes, anticlinal folds, fore reef structures and aft reef structures. To what degree do you believe that such features are present? 3

My certainty in (* Suggestive evidence for MVTD) is now: 5.0
Do you want to pursue this further? <u>NO</u>

[PROSPECTOR is ready to summarize its interpretation.]

On a scale from -5 to 5 ,
my certainty in (MVTD) is now: .8995

[The certainty of .8995 would be interpreted as perhaps mildly encouraging.]

In summary, the following have been observed:

The major favorable conclusions were:
- (* Suggestive evidence for MVTD) (5.0)
- (* Prospective ore body lies in carbonate sediments) (3.0)
- (* Right hydrothermal alteration) (3.0)
- (* Prospective ore body contains galena and sphalerite) (2.0)

The major unfavorable conclusions were:
- (* Right temperature of mineralization) (-3.996)

[The conclusions concern major concepts within the MVTD model. The observations mentioned below are the more important one supporting the overall conclusion.]

The encouraging observations were:
- (* Prospect contains fragmental host rock) (5.0)
- (* Prospective ore body involves dolomite) (3.0)
- (* Carbonate sediments containing the prospective ore body are shallow-water shelf-type sediments deposited under stable tectonic conditions) (3.0)
- (* Prospect has positive geologic features) (3.0)

The discouraging observations were:
- (* Prospective ore body contains non-telethermal alteration products) (4.0)
- (* Evidence of metamorphism) (-4.0)

[This type of deposit is associated with low-temperature mineralization. The primary reason for a relatively low certainty of occurrence is the evidence of high-temperature mineralization, with no evidence of metamorphism as an "excuse."]

I can not do any more with this hypothesis,
Do you want to pursue another one? <u>QUIT</u>

# III REPRESENTING AND USING GEOLOGICAL KNOWLEDGE

## A. Inference Rules

For PROSPECTOR to make use of such models as the one mentioned above, each model must be stored in the computer in a form different from ordinary prose or textbook descriptions. Instead, a model must be highly structured so that computer programs can draw inferences by examining the parts of the model and the relations among them.

We have chosen to structure model information as a collection of rules of plausible inference, termed more simply "inference rules." An inference rule has the form:

```
      IF    E-1  AND  E-2  AND ...  AND  E-n,
    THEN    (to degree LS, LN)      H.
```

The rule is interpreted as meaning "The n pieces of observed evidence E-1 through E-n suggest (to some degree) the hypothesis H." A probability of truth is assigned to all observations and to all hypotheses, and the inference rules specify how the probability of an hypothesis being true is changed by the observation of evidence. The way that the two numbers LS and LN establish the "strength" of the rule will be described shortly.

If the rule mentions only a single piece of evidence E, we can represent it graphically as shown in Figure 1. Using the semantic-network terminology introduced by Hendrix (1975), we refer to the rectangles as "spaces," and to the arrows connecting the spaces as "arcs."

As a simple example, consider one rule in our porphyry copper model relating to the potassic zone of a porphyry deposit. This rule states


FIGURE 1 AN INFERENCE RULE

IF Abundant quartz sulfide veinlets with no apparant alteration halos,

THEN (LS, LN) Alteration favorable for the potassic zone.

According to the model, observation of this evidence is quite encouraging -- though not conclusive -- that there is alteration that is characteristic of the potassic zone of this class of porphyry copper deposits. One the other hand, even known absence of this evidence is only somewhat discouraging for this conclusion. In general, we need to be able to say both how encouraging it is to find the evidence present, and how discouraging it is to find it absent, and this is why two numbers -- LS and LN -- must be provided by the expert for each rule.

It frequently happens that the hypothesis of one rule mentions the evidence of another. For example, our porphyry copper model also includes the following two rules:

Rule 1: IF Volcanic rocks in region are contemporaneous with the intrusive system (coeval volcanic rocks),

THEN (LS-1, LN-1) Favorable level of erosion for porphyry copper deposit.

Rule 2: IF Favorable level of erosion for a porphyry copper deposit, THEN (LS-2, LN-2) Favorable regional environment for porphyry copper deposit.

Because the rules mention each other, they chain together as shown in Figure 2. In general, the rules in PROSPECTOR connect together in various ways -- through chains, through several pieces of evidence bearing on the same hypothesis, and through the same piece of evidence bearing on several different hypotheses. Thus, the collection of rules forms an inference network, such as the one shown in Figure 3. Each space at the top of the network represents a hypothesis about the existence of a particular type of ore deposit. Notice that a typical intermediate space such as E-5 plays two roles: it provides evidence for the spaces above it (E-2 and E-3) and it acts as a hypothesis for the spaces below it (E-8 and E-9).

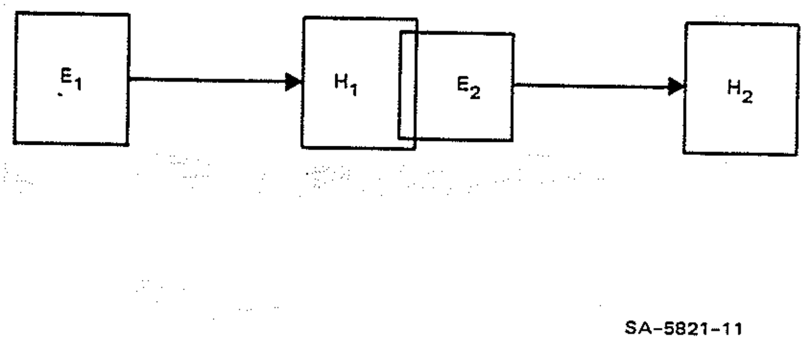


FIGURE 2 CHAINING OF INFERENCE RULES

PROSPECTOR attempts to evaluate the promise of a prospect by trying to estimate the likelihood of each top-level hypothesis. In a given case, most hypotheses will be easy to rule out, so attention will be concentrated on only a small number of active possibilities. By the same token, however, PROSPECTOR will overlook even a strong prospect if it has no model, and hence no top-level hypothesis, for the type of mineral deposit presumed present.

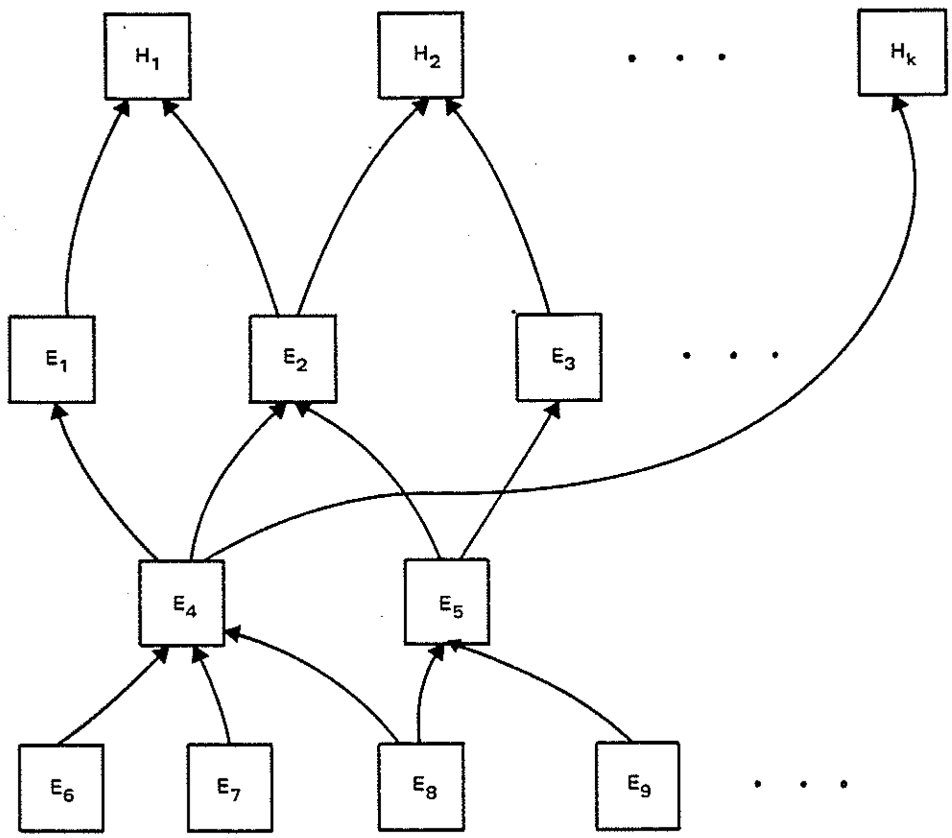


FIGURE 3 A SIMPLE INFERENCE NET

## B. Probabilistic Reasoning

An inference network of the form shown in Figure 3 is used to represent an important part of the knowlege PROSPECTOR has about various types of ore deposits. When engaged in consultation about a particular prospect, the system also needs to record geological evidence as it is provided by the user. This evidence is stored as probabilities that are associated with each space in the network. To use our earlier example, suppose that the user tells the system that "pillow structures are certainly present." PROSPECTOR records this by setting the probability of the space "pillow structures" to one. At any point in the consultation, some spaces will not have been given a probability based on the user's observations. These spaces have "default probabilities" that are initially provided by the expert geologist at the time the model is constructed. Such probabilities, known technically as prior probabilties, are available for every space in the network.

The principal form of reasoning in PROSPECTOR is the propagation of probabilities through the inference network. As an example, suppose in Figure 3 that the user provides some geologic evidence regarding space E-4 by changing the probability of that space to a new value. This change should have an effect on the probabilities of spaces E-1 and E-2, which in turn should change the probabilities of the top-level hypotheses H-1 and H-2.

Propagation of probabilities is accomplished through the application of a form of reasoning known as Bayesian decision theory [Raiffa, 1968]. This theory prescribes a method for propagating a probability from the evidence E of a rule to the hypothesis H. Propagation throughout the inference network is then a matter of iterating this procedure. We shall briefly describe the basis of the method for the simple case in which there is only a single piece of evidence E relevant to an hypothesis H. The general case, with an extended discussion, is given in Duda et al., 1976.

Bayesian inference is based on an elementary theorem of probability theory called Bayes rule. For our purposes, the so-called odds-likelihood form of the rule is most convenient. This form relates three quantities involving E and H: the prior odds* $O(H)$ on the hypothesis, the posterior odds $O(H|E)$ on the hypothesis given that the evidence E is observed to be present, and the sufficiency measure LS mentioned previously. By Bayes rule, we can write

$$O(H|E) = LS \times O(H),$$

where LS is defined by

$$LS = \frac{P(E|\ H)}{P(E|-H)} \ ,$$

where $P(E|H)$ is the probability of obtaining the evidence E given that the hypothesis H is true, and $P(E|-H)$ is the probability of obtaining the evidence E given that the hypothesis H is not true.

The quantity LS has a standard interpretation in statistics, and is called the likelihood ratio. Thus, Bayes rule tells how the odds on the hypothesis of a rule are updated by observing the presence of the evidence for the rule: the prior odds are simply multiplied by LS. An analogous formula tells how the odds on the hypothesis are updated if the evidence is observed to be absent: in that case, the prior odds are multiplied by the necessity measure LN.

Direct application of Bayes rule leads, therefore, to simple formulas for updating the probability of a hypothesis given that the user observes either that the evidence is definitely present or that it is definitely absent. In practice, as was illustrated in the interactive example session, the user is often unable to make such definite statements. Typically, the user is prepared only to indicate a degree of confidence that the evidence sought is present. In this case, a formula for updating the probability of the hypothesis can be derived that effectively interpolates between the two extreme cases of perfect certainty.

--------

* The odds on any evidence (or hypothesis) are just the ratio of the probability in favor of the evidence to the probability against the evidence. Probabilities and odds are therefore freely interchangeable through this simple relation.

Bayes rule thus gives a theoretically sound method for propagating odds (or the equivalent probabilities) through an inference network. In practice, however, certain difficulties arise that require a modification of the purely theoretical solution above. The principal difficulty stems from the fact that all the probabilities involved are subjectively provided by people, rather than being mathematically derived from some abstract, objective calculation. These subjective probabilities are almost always internally inconsistent, and application of the theoretical formulation above can lead to nonsensical results, such as the computation of probabilities greater than one. These difficulties, their resolution, and a number of generalizations of the method to cases of practical importance, are all treated in Duda et al., 1976. Most of that treatment is primarily of technical interest. One significant finding of our work, however, is of general interest and is worth mentioning here.

We have found, almost without exception, that people have great difficulty in assigning numerical values to subjective probabilities. For example, suppose we ask the user a question of the form "Is E present on the prospect?" and suppose that the user seriously doubts, but cannot rule out, its presence. Ideally, we would like the user to indicate this state of affairs by giving a probability to the system, say 0.1. Unfortunately, this might be higher than the prior probability assigned to this observation by the expert, and the system could interpret the user's response as indicating the possible presence of the evidence. In general, neither the expert nor the user can be relied upon to assign accurate probability values to situations, particularly when the situations are rare events. For example, is the prior probability that an ore body exists on a prospect one in a thousand, one in a million, or something else entirely?

These difficulties are intuitively important, and a more extended discussion could point out their consequences in detail. To overcome the problem, we have set up arbitrary scales that avoid the need to specify probability numbers. Thus, users can indicate their degree of belief that a piece of evidence E is present by using a scale of -5 to 5, as shown in the example case earlier. With this scale, a positive response will always lead to a probability higher than the prior probability, and a negative response will always lead to a probability lower than the prior probability. Analogously, experts can indicate the prior odds on hypotheses by using terms such as "very rare" or "abundant." PROSPECTOR converts responses on these arbitrary scales to probabilities, and all internal calculations use probabilities and odds. By using these arbitrary scales, PROSPECTOR can effectively calibrate all external communication on one common probability scale.

## C. Detailed Representations of Rules

The foregoing sections describe how PROSPECTOR stores models of ore deposits as collections of inference rules, and outlines the Bayesian computation of probabilities that allow the effect of a piece of evidence to be propagated through an inference network. If these were the only mechanisms employed by the system, it would be seriously deficient in several ways. Many of these deficiencies would relate to the fact that the system would have no "understanding" of the content of the rules, much less of the whole collection. Of course, "understanding" is a subtle concept to come to grips with (for people or for computers), but a small start can be made by noticing that each rule can be broken into parts and the parts related to each other.

To motivate this discussion, consider evidence requested in two very simple hypothetical rules:

Rule 5: IF pyrite in veinlets is present,
THEN H-5,

and

Rule 6: IF sulfides are present,
THEN H-6.

A user able to observe pyrite as requested in Rule 5 has surely also observed the presence of sulfides, while a user unable to observe any sulfides will surely be unable to observe pyrite. PROSPECTOR needs two mechanisms if it is to deal with this sort of elementary but pervasive reasoning. Obviously, some sort of taxonomy of minerals will be needed in order to infer that pyrite is a sulfide. Less obviously, the rules must be stored in such a way as to reveal the meaning of their parts. In this example, the internal representation of "pyrite in veinlets" must permit the system to notice that "pyrite" is part of the statement. (Notice, incidentally, that it would be unsatisfactory for the system merely to scan for the key-word "pyrite," for then incorrect inferences would be made from such statements as "absence of pyrite" or "morphology of pyrite.")

We address the general set of problems alluded to here by representing the details of rules by a network structure. This kind of representation is called a "semantic network", and has been developed through work in artificial intelligence for just such purposes. In their full generality, semantic networks are a complicated subject in their own right, and even a modest exposition of the topic would be beyond the scope of this paper. However, it is easy to see from Figure 4 how a semantic network can be used to solve our example problem, and thus to appreciate how semantic networks are employed in PROSPECTOR.

Each heavily outlined space to the left in the figure corresponds to a space of an inference network, as discussed up to this point. Each of these spaces, though, now has an internal structure. The evidence space for Rule 6 shows that the evidence sought is the presence of anything composed of sulfide minerals. The evidence space for Rule 5 shows that what is sought is the presence of anything composed of pyrite in the form of veinlets. To the right in the figure is a portion of a taxonomy of minerals, which allows the system to make the elementary deduction that pyrite is a member of the sulfide group. The structure displayed, then, allows PROSPECTOR to make the inferences needed in this example. For a full-scale system, the semantic network would be much

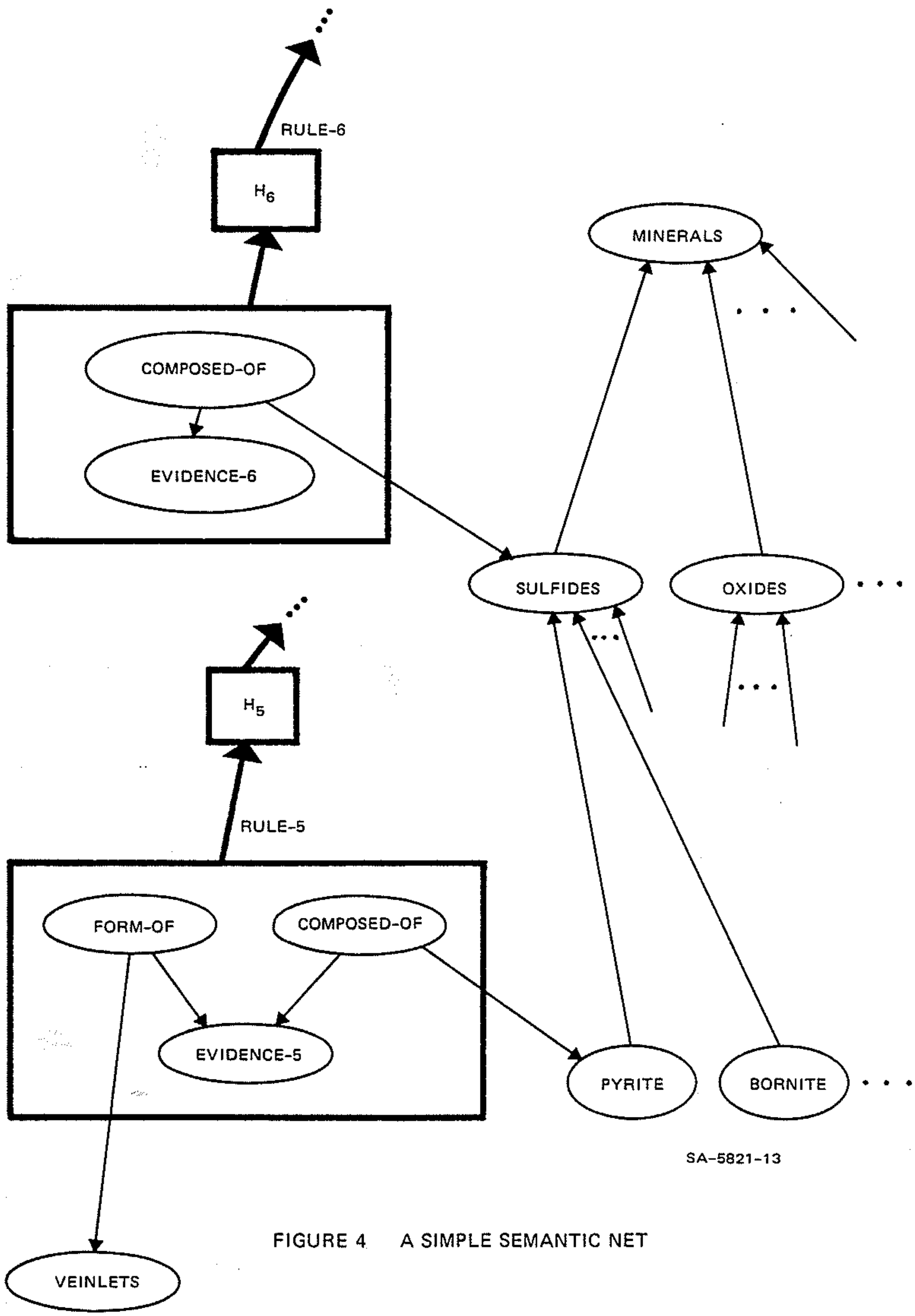


FIGURE 4 A SIMPLE SEMANTIC NET

larger, with many hundreds of spaces and arcs, and a much elaborated-taxonomy that includes not only minerals and rocks, but also other concepts such as "geologic ages" and "geologic forms."

## D. Control of Focus of Attention

We have thus far described how geological models of ore deposits are represented in PROSPECTOR, and how plausible reasoning can be accomplished using Bayesian decision theory as a foundation. For PROSPECTOR to bring this reasoning process to bear on a particular problem, it must have a way to decide at any point what the promising hypotheses are, and what evidence to seek to resolve these hypotheses. This topic is an example of a very general issue in computer science that is often referred to as "flow of control." Many elaborate control strategies are possible in our case, but we have decided to use a rather straightforward one.

Our control strategy rests on partitioning the operation of PROSPECTOR into two stages. In the initial stage, we assume that the user of the system has volunteered a set of initial observations. The first task of the system is to match these observations against the set of stored rules. When a match with a particular evidence space is found, the system propagates the effect of that evidence through the network. When this has been done for all volunteered observations, the system inspects the probabilities of the top-level spaces (i.e., the spaces corresponding to the various models of ore deposits) and selects the most probable model for further work.

At this point, the system enters the second stage of operation. All rules that bear on the selected hypothesis are examined, and the rules are ordered on the basis of their significance. (We use a measure of significance that computes the expected change in the probability of the hypothesis if the user were asked for the evidence required by the rule.) The most significant space is selected as the current focus of attention and the process is repeated. Occasionally a different top-level hypothesis may become more probable than the one currently being worked on. When this happens, the new hypothesis is selected for further investigation and the process of ordering incoming rules is repeated.

## E. Explanations and Evaluations

There are several ways in which PROSPECTOR communicates its findings to the geologist using the system. The most direct form of result is simply the probability associated with each top-level hypothesis corresponding to the presence of a particular type of ore deposit. The user can request this information at any time during the consultation session, but often the most useful information concerns not the specific probabilities but rather the reasoning process that produced them. Fortunately, the structure of the PROSPECTOR system makes this reasoning accessible. In particular, the system can examine the influence of any given piece of field data on the probability of any hypothesis. This ability makes it possible to provide two very useful types of information to the user. First, the system can call attention to the most significant evidence, both favorable and unfavorable, that the user has provided. Second, by performing "hypothetical" calculations, the system can tell the user what additional field data should be obtained. The "summarize" feature illustrated in Section II depends on this ability. Advice of this sort, which can be very valuable in making efficient use of field time, is analogous to the suggestions that might be made by a chief geologist reviewing a project.

The ability to examine chains of reasoning also in principle allows PROSPECTOR to show the user the specific rules of inference that were needed to reach a given conclusion. Since economic geology contains an important judgemental component the user may not necessarily agree with the expert-provided rules, and may accordingly wish to modify the conclusion reached. This capability has not yet been implemented in the current version of the system; when it is, it will increase the extent to which PROSPECTOR aids, but does not control, the exploration process.

# IV DISCUSSION

We have given a general view of a computer-based system designed to aid people in problems of mineral exploration and resource evaluation. In this concluding section, we contrast our approach with some more traditional alternatives, describe the current status of the system, and outline some plans for future developments.

## A. Other Inference Methods

The methods used by PROSPECTOR to draw inferences are by no means the only way of associating hypotheses with related evidence. To the contrary, there are many general approaches to this problem, ranging from the most elementary common-sense methods to highly mathematical ones.

Perhaps the simplest aid for associating observed evidence with tentative conclusions is an ordinary checklist. However, this paper-and-pencil method is able to deal with only the most elementary sorts of situations, and is not ordinarily suggested for serious work.

A more useful technique involves decision-tree structures such as the one shown in Figure 5. Each node in the tree corresponds to a piece of evidence sought. For example, E-1 might be "Is the prospect in a volcanic province?" Depending on the answer, either E-2 or E-3 would be asked next, and so on until one of the bottom nodes was reached. Each of these bottom nodes would correspond to a hypothesis -- say, the presence of a particular type of ore deposit.

Decision trees have some attractive features, not the least of which is their simplicity, but for problems of any complexity they present difficulties that may not be obvious at first glance. For example, they are not readily adapted to situations in which the user is

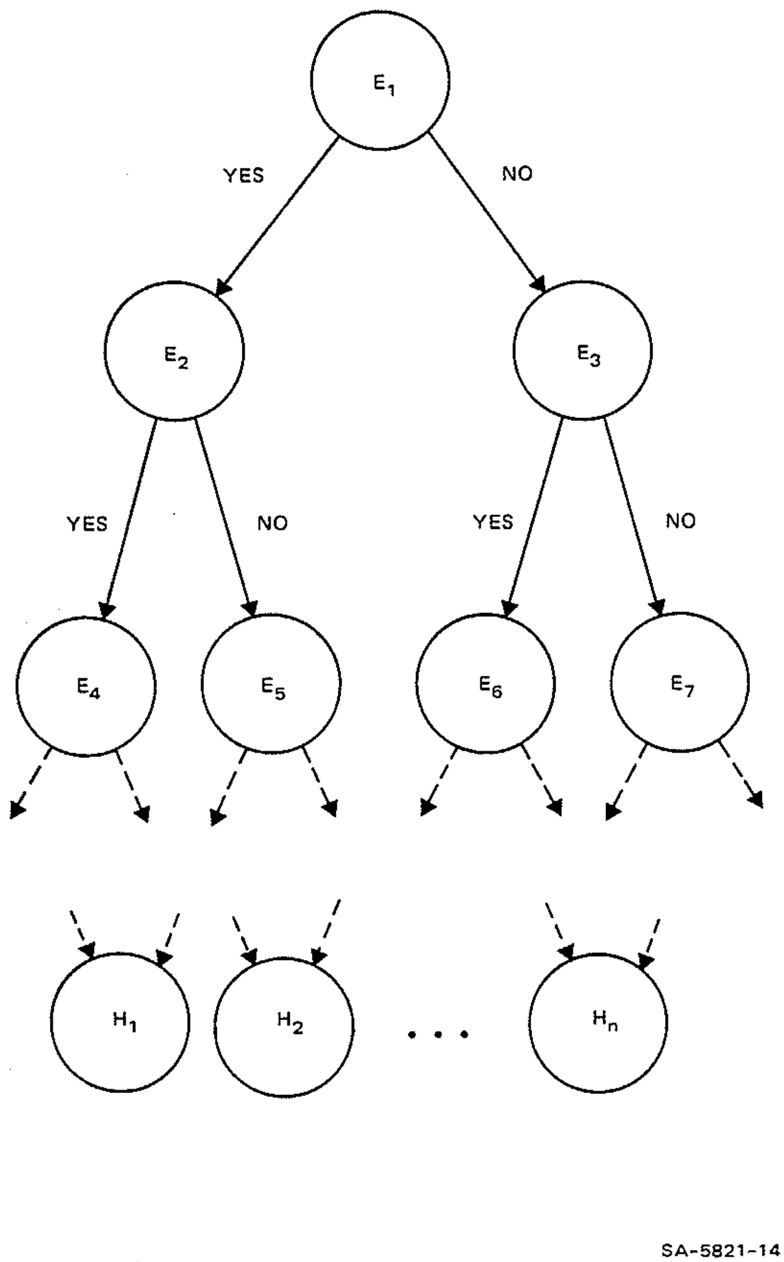


FIGURE 5 A DECISION TREE

uncertain about the answer to a particular question. An "incorrect" answer may lead to questions from the system that are recognized as obviously inappropriate. If this happens, it is difficult for the user to backtrack through the tree to decide which answer to change. In the same vein, decision trees usually do not permit the user "I don't know" responses. Decision trees also are unable to accept volunteered evidence from the user. If the user wants to tell a decision-tree-based system about a piece of evidence, he simply has to wait until he is asked. In contrast, we have seen that PROSPECTOR can easily deal with a wide range of user responses, including "don't know's" and degrees of certainty, and can accept evidence volunteered by the user.

One of the most serious problems with decision trees is the difficulty with which they are modified, since any change or restructuring of a node can affect the logical flow of questions throughout the tree. This deficiency is particularly critical when dealing with a massive and changing body of professional knowledge. By contrast, PROSPECTOR represents knowledge as rules of inference specifically because these modular portions of knowledge can be easily modified or supplemented. It is worthy of mention that although a user of PROSPECTOR is asked a sequence of questions, and therefore may suspect that a decision tree-methodology is being employed, in fact the design of PROSPECTOR rests on entirely different principles.

Probability theory and statistics furnish a classical body of techniques for drawing inferences from uncertain evidence. While applicable to a wide range of problems, a traditional limitation on their utility is the explicit or implicit need to specify joint probability distributions over all evidence and hypotheses that will be considered. These distributions either must be assumed or must be estimated from samples. For problems of moderate size, and with sufficient data, reasonable estimates may be made. Mineral exploration and resource evaluation are so complex, however, that direct estimation of the resulting high-order probability functions is impractical.

Faced with this difficulty, purely statistical approaches usually make assumptions that either reduce the problem to one of estimating a number of parameters, or that allow the high-order probabilities to be partitioned into a number of low-order ones. These low-order probabilities may then be estimated or, as in PROSPECTOR, they may be subjectively obtained from experts. One major difference between PROSPECTOR and more traditional statistical approaches is the "vocabulary" provided for linking together these low-order probabilities. Classical approaches allow only probability functions (conditional probabilities are most often used) to specify these links. PROSPECTOR, in contrast, admits the much richer language of arbitrary rules of inference coupled with special logical structures such as taxonomies.

Another important difference between PROSPECTOR and classical statistical methods is the ease with which explanations of reasoning can be produced. PROSPECTOR can examine its own chain of reasoning, and if need be can display the inference rules used, in order to expose the basis of its conclusions to critical evaluation. In contrast, purely statistical methods are unable to do much more than "replay" the formal computations that have been performed.

## B. Current Status

PROSPECTOR is in a rather early experimental stage. It is implemented in the LISP programming language, a language specially designed for work of this sort, on a Digital Equipment Corporation PDP-10 computer. A consultation session on this computer, using our existing models, generally would cost no more than a few dollars. Hardware developments in microelectronics continue to reduce the cost of computation by a factor of perhaps 10 every seven years, so computation costs can be expected to remain low even though PROSPECTOR will grow in size.

We currently have three ore deposit models installed in the system. The first model was of Kuroko-type massive sulfides, and was provided by

Prof. Charles F. Park, Jr. The second model installed was of Mississippi- Valley-type carbonate lead/zinc deposits, and was provided by Dr. Neil Campbell and Dr. Alan Campbell. The most recent model is of a subtype of near-continental-margin porphyry copper deposits, and was provided by Prof. Marco T. Einaudi.

Since we are still in an experimental stage, it is worth examining our experience in eliciting these models from senior members of the economic geology community. Each model is considerably more refined than its predecessor, in terms of both its geological content and the sophistication with which it is used. As a measure of this, the porphyry model is roughly twice as large and complex as both earlier models taken together. This increase is related in part to the large amount of work that has been done on porphyry deposits, but is also a consequence of the experience we have gained in structuring models of ore deposits. About 50 hours of interviewing time was required to define the porphyry model in terms of inference rules.

## C. Future Plans

Our immediate plans call for increasing the quantity and quality of ore deposit models. We plan in the near future to add models of stratiform chromite deposits and of nickel sulfide deposits. Ultimately, a practical system may have on the order of 25 to 50 different models.

A second extension of PROSPECTOR will be in the direction of accepting and using additional types of field data. We plan in the near future to incorporate within PROSPECTOR a computer program for statistically comparing geochemical characteristics of a prospect with those of known producing districts. This program, based on the work of Dr. Joseph Moses Botbol of the U.S. Geological Survey [Botbol et al., 1976], will enable us to deal with quantitative geochemical data, and also illustrates the nondoctrinaire design philosophy we have adopted.

Most of our discussion has centered on the use of PROSPECTOR as an exploration tool. As we mentioned in the Introduction, an important additional use of the system would be to aid in tasks of regional resource assessment, and here several possibilities are present. One is to concentrate on the analysis of data bases representing properties with mineral potential. Such databases, which have become increasingly prominent in the past several years, contain geologic and other data on properties with presumed mineral potential. PROSPECTOR could screen such a database, property by property, and assess the likelihood of a deposit for each.

A second possibility would be to follow the more common practice of partitioning a region into a grid of cells and applying PROSPECTOR to the task of assessing the potential of each cell. It would be necessary to select a cell size such that at most one deposit could occur in it; the cell probabilities would be combined to form an estimate of the number of deposits in the region. Of course, PROSPECTOR could be given data on the general geologic characteristics of the region so that manually entered data would need specify only those cell characteristics that differed from the general geologic picture.

Both possibilities lie in the future but illustrate the potential utility of a computer-based system that can aggregate geological knowledge from a team of experts and can actively apply that knowledge to specific exploration problems.

## REFERENCES


1. Botbol, J. M., R. Sinding-Larsen, R. B. McCammon, and G. B. Gott, "Characteristic Analysis of Geochemical Exploration Data," U. S. Geological Survey Open File Report OF 77-349, U. S. Geological Survey, Reston, Virginia, 1976.

2. Duda, R. O., P. E. Hart, and N. J. Nilsson, "Subjective Bayesian Methods for Rule-Based Inference Systems," National Computer Conference 1976 (AFIPS Conference Proceedings Vol. 45), pp 1075-1082, 1976.

3. Duda, R. O., P. E. Hart, N. J. Nilsson, and G. L. Sutherland, "Semantic Network Representations in Rule-Based Inference Systems," Technical Note 136, Artificial Intelligence Center, SRI International, Menlo Park, California, March 1977; to appear in Pattern-Directed Inference Systems, D. A. Waterman and F. Hayes-Roth (eds.), Academic Press, 1977, in press.

4. Ellis, J, D. P. Harris, and N. Van Wie, "A Subjective Probability Appraisal of Uranium Resources in the State of New Mexico," Grand Junction, Colorado, U.S. ERDA, December 1975, 97pp.

5. Harris, D. P., G. Berry, and A. Freyman, "The Methodology Employed to Estimate Potential Mineral Supply of the Canadian Northwest -- an Analysis Based Upon Geologic Opinion and Systems Simulation," Proc. of the Ninth Int. Symposium on Decision Making, APCOM, June 14-19, 1970, 51pp.

6. Harris, DeVerle P., Mineral Endowment, Resources, and Potential Supply: Theory, Methods for Appraisal, and Case Studies, in press. (Preliminary edition available from the author.)

7. Hendrix, G.G., "Expanding the Utility of Semantic Networks Through Partitioning," Proc. Fourth Int. Joint Conf. on Artificial Intelligence, Tbilisi, Georgia, USSR, September 3-8, 1975.

8. Linstone, H. and M. Turoff (eds.), The Delphi Method--Techniques and Applications, Addison-Wesley Publishing Co., Reading, Massachusetts, 1975.

9. Nilsson, N. J., Problem Solving Methods in Artificial Intelligence, McGraw-Hill Book Co., New York, 1971.

10. Pauker, S. G., G. A. Gorry, J. P. Kassirer and W. B. Schwartz, "Towards the Simulation of Clinical Cognition: Taking the Present

Illness by Computer," American Journal of Medicine, Vol.60, pp 981-996, 1976.

11. Pople, H. "The Formation of Composite Hypotheses in Diagnostic Problem Solving: An Exercise in Synthetic Reasoning," Proc. 5th Int. Joint Conference on Artificial Intelligence, pp 1030-1037, MIT, Cambridge, Massachusetts, 1977.

12. Raiffa, H., Decision Analysis, Addison-Wesley Publishing Co., New York, 1968.

13. Raphael, B., The Thinking Computer, W. H. Freeman and Co., San Francisco, California, 1976.

14. Shortliffe, E. H., Computer-Based Medical Consultations: MYCIN, American Elsevier Publishing Co., New York, 1976.

15. Weiss, S. M., C. A. Kulikowski, and A. Safir, "A Model-Based Consultation System for the Long-Term Management of Glaucoma," Proc. 5th Int. Joint Conference on Artificial Intelligence, pp 826-833, MIT, Cambridge, Massachusetts, 1977.