August 1985

(12)

Report No. STAN-CS-85-1067
Also numbered HPP-84-2

Acquiring, Representing, and Evaluating A Competence Model of Diagnostic Strategy

by

William J. Clancey

Department of Computer Science

Stanford University
Stanford, CA 94305

In M. Chi, R. Glaser, and M. Farr (Eds.), The Nature of Expertise (pp. 343-418). Hillsdale, New Jersey: Lawrence Erlbaum Associates, 1984.

ACQUIRING, REPRESENTING, AND EVALUATING A COMPETENCE MODEL OF DIAGNOSTIC STRATEGY

William J. Clancey

Stanford Knowledge Systems Laboratory
Department of Computer Science
701 Welch Road, Building C
Palo Alto, CA 94304

The studies reported here were supported (in part) by:

The Office of Naval Research
Personnel and Training Research Programs,
Psychological Sciences Division.
Contract No. N00014-85K-0305

The Josiah Macy, Jr. Foundation
Grant No. B852005
New York City

The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the Office of Naval Research or the U.S. Government.

Table of Contents

List of Figures

Abstract

NEOMYCIN is a computer program that models one physician's diagnostic reasoning within a limited area of medicine. NEOMYCIN'S diagnostic procedure is represented in a well-structured way, separately from the domain knowledge it operates upon. We are testing the hypothesis that such a procedure can be used to simulate both expert problem-solving behavior and a good teacher's explanations of reasoning.

The model is *acquired* by protocol analysis, using a framework that separates an expert's causal explanations of evidence from his descriptions of knowledge relations and strategies. The model is *represented* by a procedural network of goals and rules that are stated in terms of the effect the problem solver is trying to have on his evolving model of the world. The model is *evaluated for sufficiency* by testing it in different settings requiring expertise, such as providing advice and teaching. The model is *evaluated for plausibility* by arguing that the constraints implicit in the diagnostic procedure are imposed by the task domain and human computational capability.

This paper discusses NEOMYCIN's diagnostic procedure in detail, viewing it as a memory aid, as a set of operators, as proceduralized constraints, and as a grammar. This study provides new perspectives on the nature of "knowledge compilation" and how an expert-teacher's explanations relate to a working program.

1. Introduction

Over the past decade, a number of Artificial Intelligence programs have been constructed for solving problems in science, mathematics and medicine. These programs, termed "Expert Systems" (Feigenbaum, 1977, Duda and Shortliffe, 1983), are designed to capture what specialists know, the kind of non-numeric, qualitative reasoning that is often passed on through apprenticeship, rather than being written down in books. However, these programs are not generally intended to be *models* of expert problem-solving, neither in their organization of knowledge nor their reasoning process. Consequently, difficulties have been encountered in attempting to use the knowledge formulated in these programs outside of a consultation setting, where getting the right answer is mostly what matters. Their application to explanation and teaching, in particular, (Clancey, 1983a, Swartout, 1981, Brown et al., 1977), has necessitated closer adherence to human problem-solving methods and more explicit representation of knowledge. That is, building expert systems whose problem solving must be comprehensible to people requires a close study of the nature of expertise in people.

NEOMYCIN (Clancey and Letsinger, 1984, Clancey, 1984) is a consultation system whose knowledge

base is intended to be used in a tutoring program. While MYCIN (Shortliffe, 1976) is the starting point, we have significantly altered the representation and reasoning procedure of the original program. Unlike MYCIN, NEOMYCIN's knowledge is richly organized in multiple hierarchies; distinction is made between findings and hypotheses; and the reasoning is data- and hypothesis-directed, not an exhaustive, top-down search of the problem space. Most importantly, for purposes of explanation and teaching, the reasoning procedure is abstract, separate from knowledge of the medical domain. The knowledge base is also broadened to take in many disorders that might be confused with the problem of meningitis diagnosis, the central concern of the MYCIN program. Together, the knowledge base and reasoning procedure constitute a model of how human knowledge is organized and how it is used in diagnosis.

In practical terms, we are interested in determining what we can teach students about diagnosis and how this knowledge might be usefully structured in a computer program. In general terms, we want to know what design would enable an expert system to acquire knowledge interactively from human experts, to explain reasoning to people seeking advice, and to teach students. Figure 1-1 shows how a program like NEOMYCIN relates to these three perspectives, providing an idealized overview of our goals.

In teaching, GUIDON2 will use NEOMYCIN's knowledge to model a student's problem solving. A strong parallel occurs in the process of building NEOMYCIN: "Knowledge acquisition" is a process of modeling a human expert's problem solving, in which the modeler is the learner and the expert is the teacher. Similarly, to provide explanations of advice, a "user model" of the client is required. In all three settings--teaching, knowledge acquisition, and consultation explanation--a model is constructed of the person interacting with the program and a common knowledge base (NEOMYCIN) is used. We give different names to the modeling process--student modeling, knowledge acquisition, and user modeling--but the principles are essentially the same. We must determine: What is this person telling me about what he knows? What does he want to know about my knowledge? The purpose of NEOMYCIN research is to determine what kind of representation of knowledge facilitates interacting with people in these three settings--as teacher, learner, and expert problem solver. Indeed, we take the strong stand that a program is not an "expert" system, and certainly not a model of reasoning, unless it is proficient in these multiple, complex settings (see (Anderson and Bower, 1980) for a similar discussion).

We don't have such a central program today, and most knowledge acquisition is done between people. But we can still capitalize on the analogies to learn how people organize their knowledge, how they model other people's knowledge, and how they explain what they know in dialogues. For

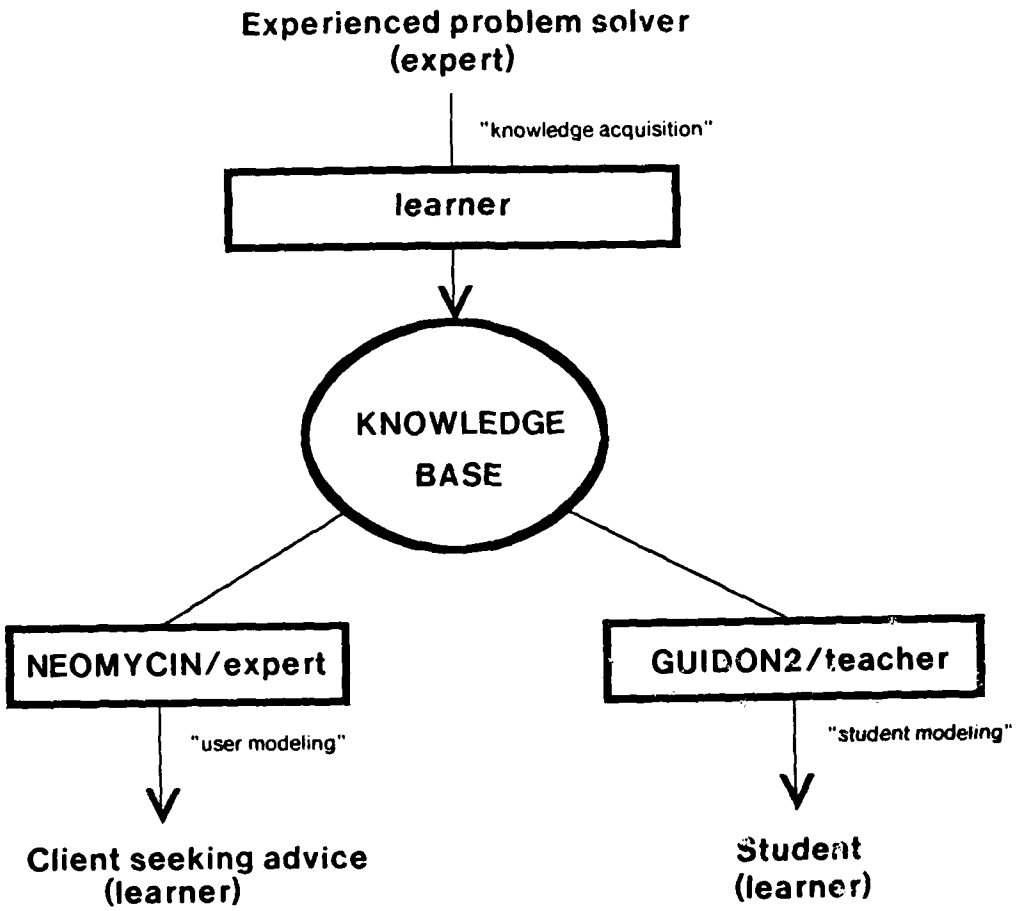


Figure 1-1: Three perspectives for acquiring, representing, and evaluating expertise

example, we can compare a physician's explanations in knowledge acquisition dialogues to what he tells his students in the classroom. What we learn from this study can be incorporated in a user modeling program. All along we refine our model of diagnostic reasoning.

There are many overlapping perspectives to such a study. For example, in modeling medical diagnosis, we must sort out modeling of disease processes, general search procedures, explanation techniques, pedagogical strategies for interrupting students, and so on. In this paper, we examine NEOMYCIN as it is currently constructed from the perspective of what we might call *the psychology of medicine*. We are interested in issues of model acquisition, representation, content, and evaluation. In particular, we will consider the following questions:

1. Why does NEOMYCIN work? How could a model derived from a problem-solvers' explanations about his behavior actually solve problems? That is, what must be true about an *explanation* of reasoning for it to be part of a procedural model?

2. What aspects of the model are *empirical*, based on observations of an expert's behavior and his explanations? What aspects are *rational*, based on mathematical and logical assumptions about the nature of knowledge and the task domain?

3. What capabilities of human reasoning are assumed by the *procedural language* for representing diagnostic strategy? How are considerations of *cognitive economy* incorporated?

4. What constraints imposed by the problem space are implicit in the *content* of the diagnostic procedure? What *correctness* and *efficiency* considerations derive from these task constraints?

5. What must be true about the nature of expertise and task domains for a model of reasoning to be expressed as an *abstract procedure*, wholly separate from the domain knowledge it operates upon?

6. Given that expert knowledge is highly "compiled" into domain-specific form and novices do not always know the right procedures, whom does NEOMYCIN model? If NEOMYCIN's abstract procedure of diagnosis is a *grammar*, constituting a model of *competence*, what are the difficulties of extracting such a grammar from expert behavior?

7. What part do multiple settings for using expertise play in evaluating the *sufficiency* of the model? How can knowledge of the underlying cognitive and task constraints be used to evaluate the *plausibility* of the model?

In pursuing these questions, we adopt different perspectives for formalizing and studying the

model. We view it as:

- *an opportunistic strategy for remembering "compiled knowledge" of disorders*--emphasizing that diagnosis is an indexing problem. The diagnostic procedure operates upon a network of stereotypic knowledge of disorders, that is, knowledge derived from experience of diagnosing many cases, not a working model of the human body and how it can be faulted;

- *a set of operators for establishing the space of diagnoses*--emphasizing that diagnosis is at heart a search problem whose bounds must be established and explored systematically;

- *a procedure derived from cognitive, sociological, mathematical and case-experience constraints*--emphasizing that the determinants of efficiency and correctness are implicit in the procedure, below the level of diagnostic behavior;

- *a grammar for parsing information-gathering behavior*--emphasizing the domain-independent character of the diagnostic procedure, how it selects from a well-structured "lexicon" of medical knowledge and specifies the "discourse structure" of the diagnostic interview.

Building a large, complex program is necessarily iterative, with early versions serving as sketches of the idealized model. Like artists, we start with an idea, represent it, study what we have done, and try again. The state of AI and computational modeling is such that an exhibit hall of completed paintings would be very small. NEOMYCIN is not a completed program, but a sketch that this paper studies and critiques. It is reasonable to address the above questions now to lend some methodological clarity to the enterprise.

Four major sections follow. In the *acquisition* section we illustrate how we collect and parse diagnostic behavior. (A detailed protocol analysis appears in Appendix II.) In the *description* section, we present an overview of our perspective on the search problem of medical diagnosis. (The entire diagnostic procedure appears in Appendix IV.) The *representation* section describes NEOMYCIN's strategy and domain knowledge architecture in detail, along with a summary of constraints implicit in the procedure. Finally, the *evaluation* section considers tests for determining the sufficiency and plausibility of the model. We conclude by considering what NEOMYCIN reveals about the nature of expertise and its implications for teaching.

2. Acquiring the model: Knowledge engineering and protocol analysis

2.1. Related work and scope of effort

In conventional knowledge engineering (Hayes-Roth, et al., 1983), an expert system is constructed by an interview process. A program is constructed and critiqued in an iterative manner. In this way, the resident "expert" frequently picks up the jargon and tools of artificial intelligence: He learns how to formalize his knowledge in some structured language, using editing programs and explanation systems to construct a "knowledge base" with the desired problem-solving ability.

NEOMYCIN was constructed in a different way. Our teaching goals required that we improve MYCIN's representation. We found that MYCIN's rule formalism made it necessary to proceduralize all knowledge, combining facts with how they were to be used (Clancey, 1982, Clancey, 1983a). With this experience in mind, we decided not to devise yet another formalism by which an accommodating physician might distort what he knew. Instead, we started (in 1980) by presenting problems to the physician to learn about his knowledge and methods from scratch. Our original objective was just to make explicit a taxonomy of diseases and subtype relations among findings; but the clarity of the approach used by our expert (and its difference from MYCIN's) ultimately encouraged us to construct the model that became NEOMYCIN's diagnostic procedure.

This investigation was influenced in many ways by previous work. For example, Pauker and Szolovits (Pauker and Szolovits, 1977) constructed a model of diagnostic reasoning, called PIP, concurrent with the development of MYCIN. Thus, we knew that a psychological approach, instead of a purely engineering approach, could be used for constructing an expert system, without a loss in problem-solving performance. Other studies, such as (Miller, 1975, Rubin, 1975, Pauker et al., 1976, Elstein et al., 1978, Kassirer, 1978) and (Benbassat and Schiffmann, 1976) strongly suggested that diagnostic strategy constitutes a separate, significant body of knowledge that might be interesting to formalize independently of medical facts themselves. Furthermore, previous research in teaching problem-solving strategies with instructional programs using AI techniques (e.g., (Papert, 1980, Brown et al., 1977, Wescourt and Hemphill, 1978)), suggested that it would be useful to go beyond MYCIN's purely domain-specific rules and make explicit the underlying general search procedure.

In related psychological research, Feltovich, Johnson, and Swanson (Feltovich et al., 1980) used fixed-order diagnostic problems to demonstrate the effect of knowledge organization on reasoning. Could we formalize an ideal organization of knowledge for MYCIN's meningitis domain? In AI, Davis

(Davis, 1980) designed a construct he called a "metarule" for controlling reasoning, but he had presented only two examples in MYCIN's domain. Could this representation be generalized for formalizing a complete diagnostic procedure? Concurrent studies at the Learning Research Development Center and CMU (Anderson et al., 1981, Chi, et al., 1981, Feltovich et al., 1980, Larkin, et al., 1980) were concerned with modeling differences between experts and novices in geometry and physics problem solving. Could we "decompile" MYCIN's knowledge into the components an expert had learned from experience and compiled into specific procedures and rules? Finally, in our previous research (Clancey, 1983a, Clancey, 1984), we had found a convenient epistemologic framework for characterizing the content of an explanation. Could this be used for directing and analyzing a knowledge acquisition dialogue?

In summary, the process of acquiring the NEOMYCIN model from expert interviews is disciplined by three greatly different perspectives:

- Psychology: The new program, unlike MYCIN, should embody a model of diagnosis that students can understand and use themselves. Moreover, a program that captures general principles of data- and hypothesis-directed reasoning can be used as the basis for a student model (Section 5.3.3).

- Knowledge Engineering: The new program, unlike MYCIN, should separate control knowledge from the facts it operates upon. The diagnostic procedure should be represented in a well-structured way, just like the medical knowledge, so that it will be accessible for explanation and interpretation in student modeling. (See (Clancey, 1985a) for detailed discussion.)

- Epistemology: The new program, unlike MYCIN, should distinguish among findings, hypotheses, evidence (finding/hypothesis links), justifications (why a finding/hypothesis link is true), structure (how findings and hypotheses are related) and strategy (why a finding request or hypothesis comes to mind). (See (Clancey, 1983a) for detailed discussion, plus Section 4.)

Besides not filling in some pre-determined representation, we have been wary of incorporating ad-hoc features into the model, just because the computer allows them. In particular, we are especially wary of all scoring mechanisms: We want every hypothesis and finding request to be based on explicit principles or totally arbitrary. It is essential that NEOMYCIN avoid numeric calculations that cannot be expressed in terms of facts and procedures known and followed by people. We use MYCIN's evidence-weighing scheme (certainty factors) to signify strength of association (Section 4.2.4), but focus decisions, such as selecting a hypothesis to test and finding to request, primarily follow from relations among findings and hypotheses (such as "sibling," and "necessary cause").

Furthermore, in proceeding in this principled way, we have avoided making the mechanisms more complex than our empirical observations of physicians' reasoning or the cases to be solved warrant. For this reason, we have not included in the model diagnostic considerations that play an important part in several other programs (Pople, 1982, Pauker and Szolovits, 1977, Chandrasekharan et al., 1979). These include: differentiation of the disease on the basis of organ system involvement; a problem-oriented approach (trying to explain the data); consideration of multiple causes; and use of probabilistic information. We have minimized these concerns by focusing on diagnosis of meningitis and diseases that might be confused with it. Of course, some of these considerations may be incorporated as we continue to develop the program.

Our research approach could be characterized as "making a push to the frontier." Some of our results might not stand up because the problems considered are not broad enough. But we will have demonstrated, as a first attempt, that certain epistemologic and knowledge engineering distinctions are useful for constructing a program that can solve problems and explain what it knows.

As another perspective, we want to determine what good teachers know about their own knowledge and problem solving methods that students would profit from being taught. In assembling a runnable computational model, we must fill in some details, such as strength of belief and activation of memory. We do this in a minimal way, devising just enough mechanism to get the behavior we want (on our small set of test cases). So, for example, we use the MYCIN certainty factor mechanism because it is convenient and simple enough. We have much to learn about what teachers *know* about their knowledge and problem solving, and much of what we do falls in the realm of the traditional computer science problem of designing an appropriate programming language to encode these structures and procedures. Thus, our first interest is to replicate what people know about what they do, only secondarily to formalize models of how the mind works (e.g., activation of knowledge), and not at all to derive mathematically optimal models that might replace or augment what people do.

With our objective of constructing a tutoring program with useful capabilities, the purpose of NEOMYCIN research is not to make the best medical diagnostic program, but to demonstrate a representation methodology for separating kinds of knowledge and formalizing strategies in domain-independent form. The problem domain is sufficiently complex to be challenging, and we have formalized a sufficient subset of diagnostic strategies to provide an interim report on our approach. We have uncovered a number of cognitive problems of interest that have been little studied, particularly how focus of attention changes during diagnosis.

2.2. The hypothesize and test theory of diagnosis

In studying diagnostic behavior, we used the epistemologic framework mentioned above and evolved a set of terms for describing the process of diagnosis. Terms that will appear frequently in subsequent sections, such as "task" and "differential," are defined in Appendix I.

In addition, we began with the traditional model of diagnosis, which says that each request for case information, some finding, directly relates to some hypothesis (Figure 2-1). This model suggests several problems for investigation (points corresponding to numbers in the figure):

1. Where do the initial hypotheses come from?

2. How does the problem solver choose a finding to confirm or test a hypothesis?

3. What causes attention shift to a new hypothesis?

4. How does the problem solver know when he is done?

We define a *diagnostic strategy* to be the control structure that regulates these four decisions. This hypothesize and test theory drove our initial investigations, but the NEOMYCIN model eventually became much more complex.

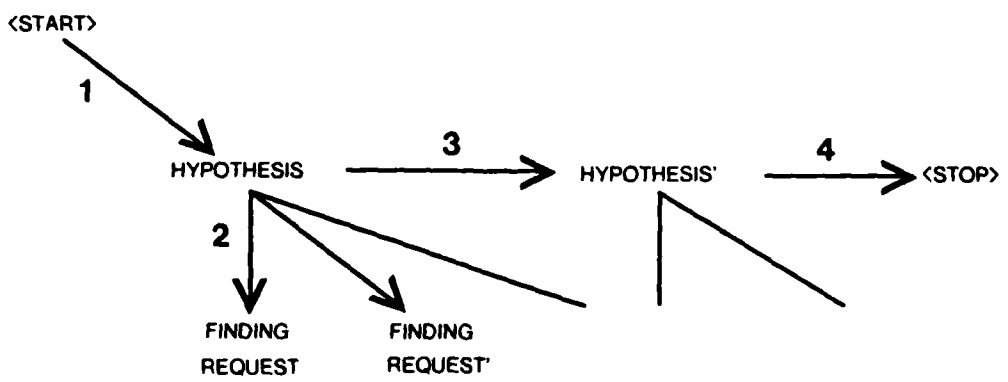


Figure 2-1: Hypothesize and test theory of diagnosis

2.3. Knowledge acquisition technique

With our interest in formalizing the reasoning process of diagnosis, it is particularly important to allow the expert to request problem findings in whatever order he desires. Our main concern is to determine what task and domain knowledge leads to each finding request. Contrary to the protocol-collection procedure most often used today (Newell and Simon, 1972, Ericsson and Simon, 1980, Kassirer, et al., 1982, Kuipers and Kassirer, 1984), with a minimal number of interruptions, we frequently ask the expert specific questions. In retrospect, this is not always done in a consistent way, and is sometimes so late that the expert has clearly moved ahead (see Line 30 in Appendix II). However, the expert appears to be quite tolerable of interruptions, perhaps from his teaching experience, though of course he might not be typical in this respect.

The questioning techniques we use are listed here, in somewhat idealized form.[1]

- Epistemologic distinctions:
    - Be concerned about the specificity of a finding request: Is it a general maneuver or does he have a specific hypothesis in mind?
    - When asking why a finding came to mind, distinguish between strategic and causal explanations.
    - Distinguish between substances and processes; watch out for composed explanations that leave out intermediate processes or refer to substances as if they were processes.
    - Do not delve into explanation of causal mechanisms that goes beyond the expert's level of reasoning.
    - Ask for definitions and try to detect synonyms, which might be mistaken for different entities.
- Interactive considerations:
    - Immediately after a finding is requested, and before supplying the information, ask why the finding came to mind (otherwise new hypotheses might be used to rationalize the request).
    - When the expert indicates that he has formed some hypotheses, ask him to list his

[1] Typical of our attempt to apply expertise in multiple settings, we use such generalizations of our own behavior as expectations of what a student or client watching NEOMYCIN might want to know.

differential (this encourages completeness).

- When a specific hypothesis is being tested, ask about ordering of data requests: Are these "routine" questions for the hypothesis, or has the expert been reminded of some particular correlation or causal process?

- When the expert appears to be changing his task and/or focus without commenting, confirm this and find out why.

- Watch for assumptions made by the expert: What is he inferring from the context of his dialogue with you and not explicitly confirming? Ask why certain questions were not asked.

2.4. Illustration of level of protocol analysis

We introduce our analysis of an expert's problem solving and explanation protocol with an excerpt (Figure 2.4) from the end of the case we analyze in Appendix II. Phrases are broken to separate different kinds of statements; **MD** = the medical expert, **KE** = the knowledge engineer. (Again, we choose the term "knowledge engineer" to make clear that this is not presented as a formal psychological experiment.) Brief annotations illustrate our terminology. Annotations always *precede* the protocol section they pertain to.

The analysis shows how findings, hypotheses, and tasks are typically related. Lines L5 to L7 are most interesting in this aspect. Here we see plainly the interaction of task knowledge (stating a list of tested hypotheses), focus of attention (hematoma), and application of domain knowledge (what causes hematoma). One hypothesis in focus, hematoma, was tested by considering what could have caused it. (Interestingly, the physician is so caught up in his role as clinician, he addresses the KE as if he were the patient.)

It is also worth noting that the expert states in L2 that he is planning to go back to ask for more information. Again, in L9 he characterizes his own behavior in general terms. This is typical of the abstract statements this expert makes about diagnosis. His "explanations" of what he does abstractly characterize his problem-solving procedure: "formulate a differential" and "ask more questions." An important aspect of these explanations is that they are not arbitrary "rationalizations," but are abstract descriptions of a procedure that could generate his finding-requests and hypotheses. They do not necessarily correspond to steps of a procedure that he consciously considers, but are rather the "syntax" of his behavior. The expert's statements constitute a set of tasks and goals that can be fleshed out as an executable procedure. This is

A task has been completed...

L1 MD: I've gotten a pretty good data base,

A new task is planned...

L2 so I am going to go back and just ask a couple more questions.

There is a differential...

L3 I have formulated in my own mind what I think some of the possibilities are.

L4 KE: Can you tell me what you think are some of the possibilities?

The differential is stated...

L5 I think that there is a very definite possibility that this patient does not have an infectious disease. She could have brain tumor, or a collection of blood (hematoma) in her brain from previous head trauma

In reviewing, the expert notices that the task

"PURSUE-HYPOTHESIS (focus = mass lesion)"

was not completed; all of the causes have not been considered.
So the problem-solving process shifts task and focus:

task: TEST-HYPOTHESIS (hematoma)
evidence rule: head-trauma -> hematoma
task: FINDOUT (head-trauma)

L6 (that is a question I should have asked, by the way...)

L7 Have you had any recent head trauma?

L8 KE: Head trauma, no.

L9 MD: You'll find that this happens to physicians. As they formulate their differential diagnosis and then they go back and ask more questions.

L11 KE: What comes after...?

L10 MD: Then I would say a chronic meningitis.

Figure 2-2: Example protocol analysis

obviously important if the model we construct from the expert's explanations is to solve problems successfully and to be useful in teaching. We know that our expert was an unusually good teacher, so we cannot expect that every expert's explanations would have this property.

Finally, this excerpt illustrates how during the process of reviewing the differential (a task) the expert realizes that a hypothesis should be tested or refined (broken into subtypes or causes). We do not view this as an error on his part. Rather, as the expert says in L9, reviewing is a deliberate maneuver for being complete; it helps bring other diagnostic tasks to mind. NEOMYCIN does not behave in this way because it is a simplified model that does not precisely model how knowledge of diseases is stored or recalled. This level of modeling may very well be useful for understanding the basis of diagnostic strategies, as well as for considering the space of alternative strategies people are capable of and the causes of errors.[2]

3. Overview of the diagnostic model

3.1. Flow of information

Figure 3-1 provides an overview of the flow of information during diagnosis. The loop begins with a "chief complaint," one or more findings that supposedly indicate that the device is malfunctioning. These findings are supplied by an *informant*, who has made or collected the observations that will be given to the problem solver. By forward reasoning, hypotheses are considered. They are focused upon by a general search procedure, leading to attempts to test hypotheses by requesting further findings.

Keep in mind that this diagram shows the flow of information, not the invocation structure of the tasks. TEST-HYPOTHESIS regains control after each invocation to FINDOUT and FORWARD-REASON. Similarly, the subtask within ESTABLISH-HYPOTHESIS-SPACE that invoked TEST-HYPOTHESIS will regain control after a hypothesis is tested. Tasks can also be prematurely aborted

---

[2] As will become clear later, we might link NEOMYCIN's metarules to the domain memory model used by Kolodner in the CYRUS program (Kolodner, 1983). In this paper, we present prosaic summaries of the underlying memory constraints (Appendix IV and Section 4.3), many of which bear striking resemblance to Kolodner's results, such as the importance we give to disease process features for differentiating among diseases.

and the "stack popped" in the manner described in Section 4.1.[3]

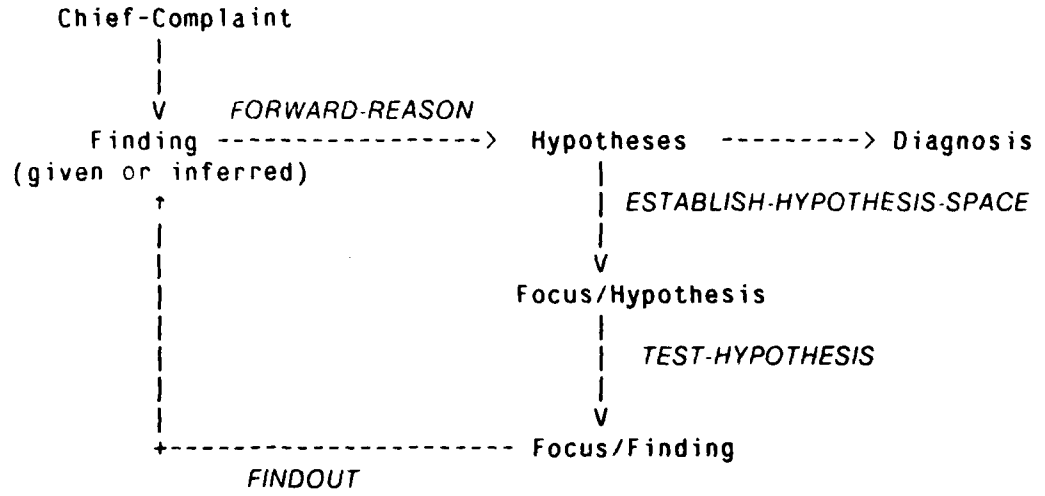


Figure 3-1: Flow of information during diagnosis
(Tasks appear in capitalized italics.)

3.2. Tasks for structuring working memory

Figure 3-2 shows the general calling structure of tasks in the diagnostic procedure. An important perspective behind the design of this procedure is that the diagnosis can be described abstractly as a process in which *the problem solver poses tasks for himself in order to have some structuring effect on working memory*. Metarules for doing a task bring appropriate sources of knowledge to mind. Thus, it is very important that the procedure is structured so that the tasks make sense as things that people try to do.

Diagnosis involves repetitively deciding what data to collect next, generally by focusing on some hypothesis in the differential. If we examine the kind of explanations a physician gives for why he is requesting a finding, we find that most refer to a hypothesis he is trying to confirm; this is the conventional view of diagnosis. But we find that a number of requests are *not directed at specific hypotheses* or *relate to a group of hypotheses*. The problem solver describes *a more general effect that knowledge about the finding will have on his thinking*. For example, information about pregnancy

---

[3] An obvious alternative design is to place tasks, particularly PROCESS-FINDING and PURSUE-HYPOTHESIS, on an agenda, so findings to explain and hypotheses to test can be more opportunistically ordered (e.g., see (Hayes-Roth and Hayes-Roth, 1979)). It is possible that the procedural decomposition of reasoning in NEOMYCIN, which suitably models an expert's deliberate approach on relatively easy cases, will prove to be too awkward for describing a student's reasoning, which might jump back and forth between hypotheses and mix data- and hypothesis-directed reasoning in some complex way.

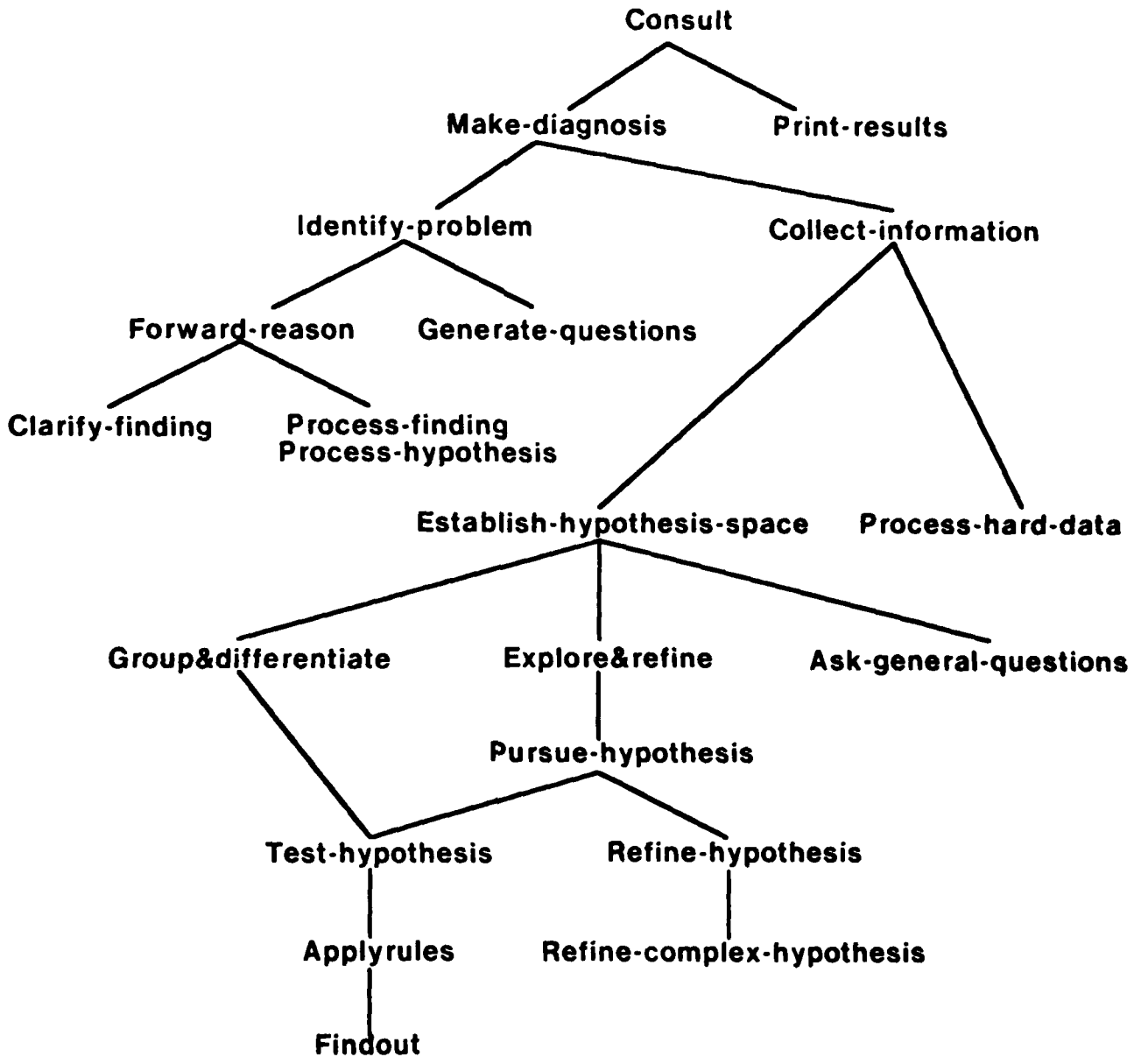


Figure 3-2: NEOMYCIN's diagnostic strategy.
(All terminal tasks shown here except PRINT-RESULTS invoke FINDOUT directly or through APPLYRULES.)

would "broaden the spectrum of disorders" that he is considering. He considers fever and trauma, very general findings, in order to "consider the things at the top." Thus, besides being focused on particular hypotheses, finding requests are intended to affect the differential in some way, for example, to restrict it categorically or to rule out unusual causes. We call the overall task of collecting circumstantial evidence (history and physical) "establishing the hypothesis space" because it is oriented towards circumscribing the space of diseases that must be considered.

Structurally, we relate this heuristic search to multiple hierarchical organizations of disorders. Figure 3-3 illustrates our model in general terms. The problem solver receives initial information that "places him in the middle" of some hierarchical organization of known diseases. Here, we show an etiological hierarchy (defined later). In the protocol we analyze in Section II, "chronic-meningitis" was first considered, not "infection", something at the top of the hierarchy, or "tb-meningitis" something at the bottom. The process of diagnosis then involves massaging this set of initial guesses by first "looking up" for general evidence that establishes the class, and then "looking down" to be as specific as possible. To establish a diagnosis, the physician must not only attempt to collect direct evidence for it, he must establish paths upwards through his multiple hierarchies in which the diagnosis is contained.

Put another way, the physician tries to form a set of possibilities that includes the "right answer" and then narrows down the possibilities to a small, treatable number. This is why a premium is placed on questions that would "broaden the spectrum of possibilities that must be considered" or, alternatively, lend confidence that the typical, a priori most likely diseases under consideration are appropriate.

To repeat the main point, *we explain finding requests in terms of the effect they are intended to have on the differential.* And moreover, at each point, as findings are requested that could have a certain effect, we say that the *task* of the problem solver is to bring about this effect on his thinking, to change what he is considering or give him confidence in some respect. Each effect provides structure to the problem in some way: characterizing, refining, or confirming the causes that must be considered. Figure 3-4 shows graphically how each of the operators affect the space of hypotheses.[4] This analysis is of course strongly inspired by Simon's study of the role of the problem space and how it pertains to ill-structured problems (Newell and Simon, 1972, Simon and Lea, 1979). Pople, in work

---

[4]The objective is to put the "right answer" into the box labeled "differential." Possible answers, hypotheses, are put focused on, confirmed, grouped, differentiated, and refined. The box is broadened to include other hypotheses by asking general questions. Determining a finding may involve requesting it or determining another finding. Findings must be explained (accounted for causally) with respect to the differential.

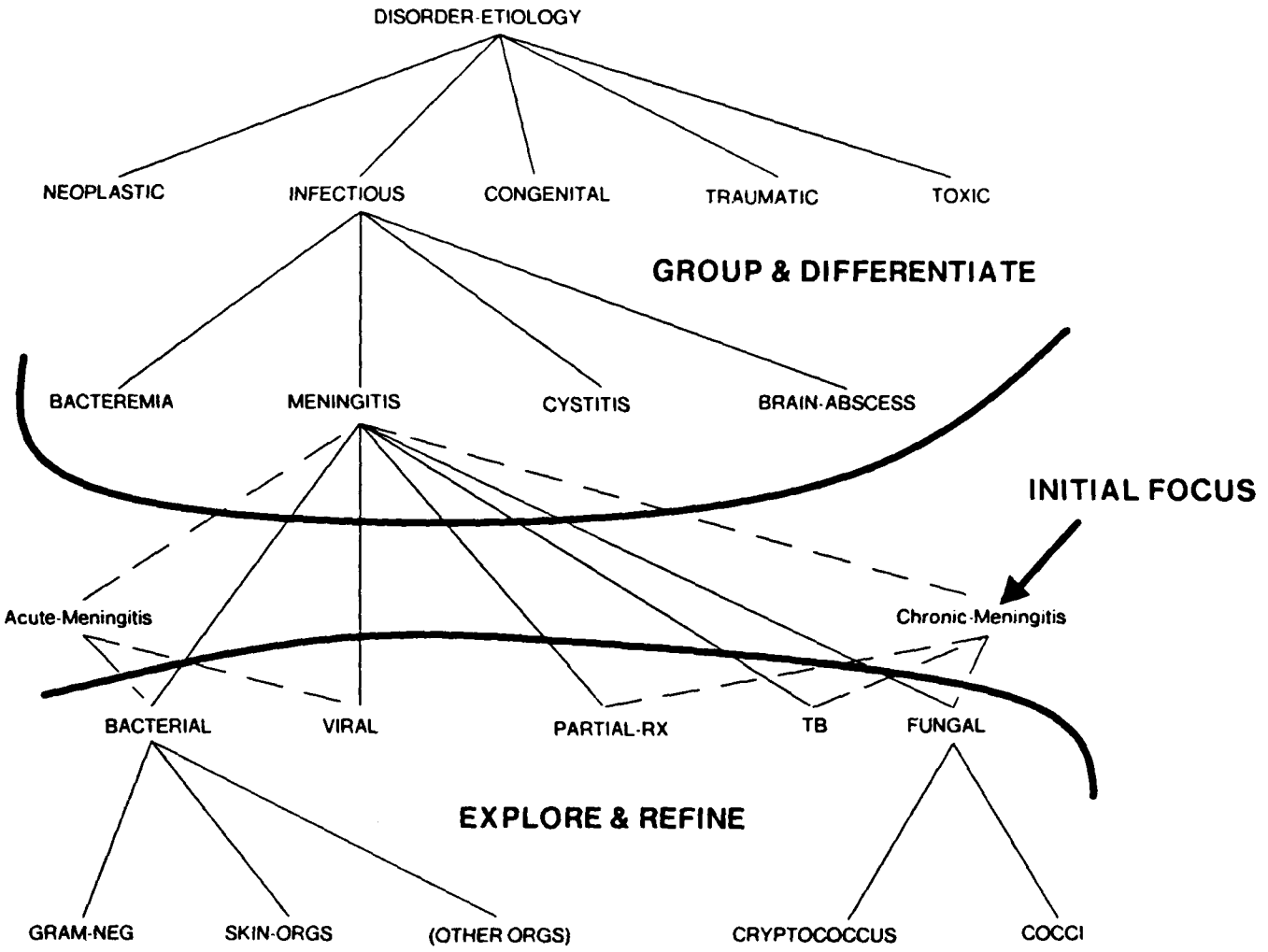


Figure 3-3: Overview of diagnostic search in an etiologic hierarchy: Initial information brings problem-solver to an intermediate hypothesis; it must be confirmed by considering classes containing it, and then it must be refined by considering more specific disorders.

concurrent to ours, has developed this point very well and appears to adopt the same "task-oriented" terminology for the proposed CADUCEUS follow-on to INTERNIST (Pople, 1982). Patil (Patil, 1981) has defined operators for constructing alternative causal models to explain findings on multiple levels of detail. Returning to Elstein's study of medical problem solving (Elstein et al., 1978), we find similar experiments and analyses of how a physician reasons about alternative formulations of the problem he is trying to solve. Finally, the idea of an *information gathering strategy* for classifying objects or phenomena was pioneered by Bruner (Bruner, et al., 1956) in experiments that allowed the problem solver to order his data requests, so the different strategic motivations could be studied.

3.3. Problem formulation and other approaches to diagnosis

It is worth noting that this model of diagnosis differs from a Bayesian model in its emphasis on a structured search. The problem solver is not just working with lists of diseases. There are general maneuvers for contrasting, exploring, and seeking evidence in terms of *relations* among diseases. Nor is this model what medical students are taught in textbooks. Students are commonly given an outline of all data that they might collect, organized by "social history," "previous illness," and so on, suggesting that medical diagnosis is a process of collecting data in a fixed order. The result is that students sometimes collect information by rote, without thinking about hypotheses at all!

The aspect of problem solving that involves forming a set of initially unrelated hypotheses and then finding ways to group, contrast, and refine them is often called "initial problem formulation." The capabilities of NEOMYCIN (and systems like PIP (Szolovits and Pauker, 1978) and CADUCEUS (Pople, 1982)) should be contrasted with the exhaustive top-down analysis used by programs like MYCIN and CENTAUR (Aikins, 1980). In a sense, the process of "looking up" into categories serves as a "big switch" as conceived in the General Problem Solver (Newell and Simon, 1972). It is the operation of viewing the overall problem in dramatically different ways: Did the patient fall and hit his head? Does he have an emotional problem? Is there a congenital weakness in the vascular system? Is there a tumor? Has the patient been infected by a virus? Did the patient consume something toxic? Diagnosing each of these dramatically different process requires bringing specialized knowledge into play. So we might imagine constructing specialized subsystems of knowledge to deal with infectious disease diagnosis, psychological analysis, and toxic drug disorders, and integrating them by the GROUP-AND-DIFFERENTIATE procedure of comparing and contrasting likely categories of disease.

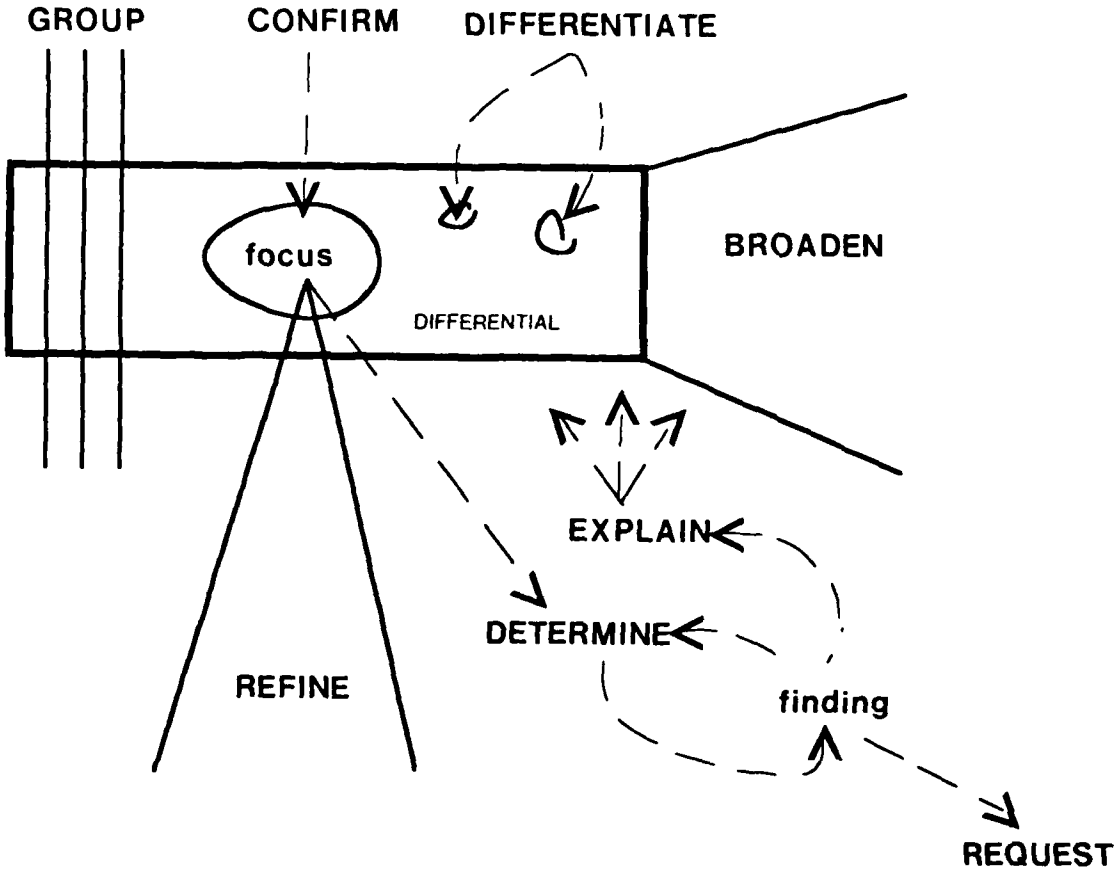


Figure 3-4: Graphic interpretation of each task as an operator for affecting working memory. (See text for elaboration.)

3.4. A causal model of what happened to the patient

So far we have described diagnosis in terms of heuristics for carrying on an efficient search of a combinatorially large space. However, it must be remembered that a diagnosis is not just a label, but constitutes *a model of the patient*. This model is a causal story of what happened to bring the patient to his current state of illness. The general questions of diagnosis regarding travel, job history, medications, etc. (the categories emphasized to a student) seek to circumscribe the external agents, environments, or internal changes (due to age, pregnancy, other diseases) that may have affected the patient's body. Thus, "establishing the hypothesis space" is more precisely characterized as "establishing the space of causes."

The following protocol excerpt provides a typical causal story, showing how a finding request is intended to establish the space of causes that must be considered.

KE: **What about pregnancies? Why is that important?**

MD: **When I asked about compromised host, that includes a wide spectrum of problems. The pregnant woman is probably the most common compromised host, in that during the pregnancy period women are more susceptible to dissemination of certain types of infections, and cocci is a classic of that. Whereas most of us would localize cocci in the lungs, pregnant women disseminate cocci to the meninges more commonly. The same thing happens with TB.**

KE: **Would it be fair to say that the question about pregnancy is not necessarily specific to the possibility of a cocci infection, but is of more general interest?**

MD: **Yes, I think it is of more general interest. It is pertinent to cocci, but would also be considered perhaps in other areas, because it would change your thinking a bit, the pregnant woman having a little different spectrum of infection than a regular, normal person.**

Here the expert supplies a causal explanation for how pregnancy effects the body, mentioning the very important concept of "dissemination"--spread of an infectious agent in the body. In trying to establish a causal story of an infectious disease, the physician looks for general evidence of exposure, dissemination, and impaired immuno-response--all of which are necessary for an infection to take place, regardless of the specific agent. Importantly, diseases can be ruled in or out on the basis of general evidence for these *phases* in the causal process, so the physician needn't directly try to rule in or out all of the specific diseases. Thus, the process of establishing the space of causes reduces to considering broad categories of evidence (e.g., "compromised host" implies impaired-immuno-response), rather than focusing narrowly on every specific causal mechanism and agent that might be involved. Moreover, this might be generalized even further by characterizing some causal

stories as "unusual" and others as "typical." Thus, establishing the space of possibilities reduces to determining whether the patient is "typical," or whether "unusual processes" might be occurring. In this style of diagnosis, characteristic of our domain, diagnosis is categorical, with essentially no concern for low-level causal arguments.

In his analysis of the patient, the physician's "process-oriented approach" is manifested in several ways. The most obvious are the general questions (ASK-GENERAL-QUESTIONS) for determining whether the patient has had related problems in the past. This is a key maneuver for circumscribing the problem space. For example, by asking if the patient has been hospitalized, one learns about all serious illnesses the patient has had. This is an excellent starting point for determining what causal processes might be implicated in the current disease. Learning that there have been no previous hospitalizations, illnesses, medications prescribed, etc., the problem solver can be reasonably sure that he has an accurate data base for making decisions: He knows what has affected this patient and can infer that everything else is "typical" or "what one might expect." Thus, the use of general questioning is perhaps the most heuristically powerful technique in medical diagnosis. The anatomically-oriented "review of systems" is similar, particularly as a spatial reminder of possible diseases, but it is not used by NEOMYCIN.

Constructing a model of the patient is often described informally as forming a "picture of the patient." The physician establishes the sequence in which findings were manifested and factors this with information about prior problems and therapies, using their time relations to match possible causal connections. For example, a fever might be a precursor to an illness that later manifests itself by abdominal pains. Thus, the physician is not just matching a set of symptoms to a disease, he is matching the order in which the symptoms appeared and how they changed over time to his knowledge of disease processes--a much richer organization than a mere list of symptoms. The physician remembers the sequence, knowing what symptoms to expect or to ask about, from his knowledge of the underlying causal process that relates the symptoms to one another.

Another way to understand the importance of process knowledge is to consider logically the importance of differentiating between hypotheses. In a pure sense, this does not mean to confirm them independently, but to gain information that will favor one and disfavor another. This is the sense in which diagnosis is a process of modeling the patient. When the interpretation is ambiguous, it is necessary to gain more information. Discrimination in this way presupposes that there is some *dimension* for comparison. That is, we must have some common way for viewing the competing diseases. In NEOMYCIN, we call this the *disease process frame*. Its *slots* are the features of any disease--where it occurs, when it began, its first symptom, how the symptoms change over time,

whether it is a local or "systemic", etc. This frame applies to more than disease processes, of course. For example, it can be used in the "oil spill problem" (Hayes-Roth, et al., 1983) to diagnosis the causes of oil spills by their frequency, amount, change over time, periodicity, and location in the network of drainage ditches.

The following excerpt from a class discussion with our expert illustrates how this kind of process orientation is critical to causal reasoning.

TEACHER: Think of the common anemias that a young person might get, and think of anemia in general. There are two ways to look at it. You start out with an adequate number of red cells and you reach the point of being anemic, there are two ways you can do it. You're losing blood excessively, or you're not making enough to replace your normal losses. Those divide anemia into two major categories. Production deficits or loss of blood. So you can talk about reasons that a young person might lose blood.

Basically to lose enough blood to become anemic either you are losing it in your stool, GI bleeding, ... what's a good question about GI bleeds, or the most common reason for blood loss in the United States is what? What physiologic function causes people to lose blood?

STUDENT:
Menstruation. She said that it was normal.

TEACHER: Normal. Normal menstrual periods, okay. So now the question is if you don't get a good history for excessive blood loss then you question, are people producing blood adequately? You can have some serious derangement in productions such as sickle cell anemia, or they may not have the basic substrates.

Even here, causal reasoning is categorical, with general consideration of production deficiency, loss of product, or substrate (input) limitation.

3.5. Structure of knowledge

The hypothesis space is structured in many different ways, with different purposes. For example, an etiological taxonomy, based on the *ultimate origins* of disorders, can be contrasted with an "organ system taxonomy," also used in medicine, which is a strict hierarchy by location of the disorder. Siblings of the etiologic taxonomy are alternative causes for a given disease process, which is why the etiological taxonomy is favored over the organ system taxonomy for focusing search during diagnosis.

The task of establishing the hypothesis space blends the good human ability to detect familiar

patterns (by data-directed associations) with a critical analysis that considers alternatives and unusual possibilities, with different indexing schemes used for these purposes. Studies indicate that the medical expert differs from a novice precisely by his ability to call to mind useful categories of disease (Feltovich et al., 1980). For example, in diagnosis of congenital heart disease, the expert learns the list of causes associated with abnormal noises on the left side of the heart. Feltovich calls this the *logical competitor set.* Significantly, this grouping is often orthogonal to the traditional hierarchies given in textbooks. Similarly, a subset of hypotheses can be remembered by labelling them, as in meningitis we refer to "the unusual causes of bacterial meningitis." Thus, over time the expert evolves a complex organization of hypotheses that is more finely indexed than a simple hierarchy (Feltovich et al., 1980). He efficiently circumscribes the possible causes by relating a familiar interpretation with unlikely, but important causes that might be confused with it.

3.6. Activation of knowledge

Modeling human reasoning requires some model of the *activation* of knowledge. The idea is basic in medical diagnosis: Any given fact about the patient might have many real world implications, but only those relevant to diagnostic hypotheses should come to mind. As a simple example, consider a physician told that the patient has pets. The expert, diagnosing a possible infectious disease, might ask, "Does the patient have turtles?" *Some sort of intersection match* has occurred that activated Salmonella as a diagnosis (because it is a bacterial infectious disease). If the leading hypothesis had been cancer, it is less likely that the Salmonella association with turtles would have come to mind when pets were mentioned. If so, we would say that a shift in focus of attention occurred. A model of data- and hypothesis-directed reasoning, such as NEOMYCIN, must specify how data is used and how focus of attention changes.

Most programs use a form of "spreading activation" (Anderson and Bower, 1980, Rumelhart and Norman, 1983, Szolovits and Pauker, 1978) by which knowledge structures are brought into consideration based on their proximity. NEOMYCIN's model incorporates these dimensions:

- *Context:* In simple terms, this concerns when relations between findings and hypotheses are realized. The value of known findings is realized when a new hypothesis is triggered (see PROCESS-HYPOTHESIS). Support for previously considered hypotheses (ancestors and immediate descendents of the differential) is realized when a new finding is received (see PROCESS-FINDING). These are called *focused forward-inferences.*

- *Strength of association:* "Antecedent rules" are applied immediately (discussed in Section 4.2.4).

- *Level of effort:* Intermediate subgoals are only pursued when applying "trigger rules," interpreting "hard findings," or deliberately attempting to confirm a hypothesis.

3.7. Summary of NEOMYCIN's reasons for gathering information

One measure of complexity of NEOMYCIN's model of diagnosis is the number of reasons for requesting a finding. In MYCIN the only reason for asking a question was to apply a rule that concluded about some "goal." This is analogous to the hypothesis and test, "single-operator" view presented in Figure 2-1. NEOMYCIN's tasks in essence give more structure and meaning to the data-gathering process. Besides testing a hypothesis, the program has the following direct motivations for gathering information (with related task in parentheses).

- *follow-up questions that specify previous information* (Given that the patient has a fever, the program will ask what the temperature is.) (CLARIFY-FINDING).

- *process-oriented follow-up questions* (When did a headache begin, how severe is it, where is it located?) (CLARIFY-FINDING)

- *process-oriented discrimination questions* (To discriminate between meningitis and brain-abscess, determine if the disorder is spread throughout the central nervous system or is localized.) (GROUP-AND-DIFFERENTIATE)

- *triggered questions* (Given that the patient has a stiff neck, we might immediately ask whether he has a headache or other neurological symptoms, because of the possibility that this might be meningitis.) (FORWARD-REASON)

- *general questions to determine the availability or presence of findings and tests* (To determine whether the CSF is cloudy, a lumbar puncture must be taken.) (FINDOUT)

- *general questions to establish that the relevant history is complete* (Has the patient been hospitalized recently? Is he taking any medications?) (ASK-GENERAL-QUESTIONS)

The expert-teacher's directives to students are the primary source for formulating the tasks of NEOMYCIN's diagnostic procedure (Appendix III).

4. Representing the model: Strategy and domain knowledge

NEOMYCIN's abstract and explicit diagnostic procedure distinguishes it from other AI programs. The procedure is *abstract* because it is separated from the domain knowledge--a feature common to frame-oriented systems. The procedure is *explicit* because it is represented in a well-structured way,

not arbitrary code--a feature common to rule-based systems.[5] Here we discuss these two knowledge representations.

4.1. Representing strategy: Tasks, metarules, and end conditions

As already described, the strategy part of the model is represented as subprocedures we call tasks. Each task has an *ordered* list of rules, sometimes called a "rule set," associated with it.[6] We call them *metarules* because they reason about which domain rules (more generally, "domain relations") should be applied to the problem. The metarules determine which causal, subtype, definition, or disease process relations will be exploited for purposes of broadening the differential, contrasting hypotheses, focusing on a hypothesis, refining a hypothesis, confirming a hypothesis, or determining whether a finding is present.

For example, the FORWARD-REASON metarule that says, "If there is a red-flag finding, then do forward reasoning with it," is using the relation "red-flag finding" to index the knowledge base. More specifically, this metarule causes red-flag (or significant, abnormal) findings to be considered first. We say that the relation "red-flag finding" *partitions* set of findings. This is the typical way in which metarules use relations that organize domain knowledge to select findings, hypothesis, and relations to apply to the problem at hand. To the degree that a concept like "red-flag finding" can be given a consistent meaning in several problem domains, the diagnostic procedure is domain independent. It is plausible that we might construct such a theory of knowledge organization because relations like "red-flag finding" are completely defined by how they are used by the diagnostic procedure.

A task has associated with it a description of how its metarules are to be applied. (To "apply a rule" means to determine whether the "if part" of the rule is satisfied [i.e., the rule "succeeds"], and if so, to carry out the action specified in the "then part" of the rule.) There are four possibilities:

1. *simple, try-all*: all of the metarules are applied once in sequence (a simple procedure of multiple steps).

2. *simple, don't-try-all*: the metarules are applied in sequence until one succeeds, then the task is complete (control returns to the calling task) (a "do one" selection).

---

[5] That is, the procedure is expressed in a language for which we can write an interpreter that can reason about how tasks are invoked, as well as their input and output. The notation is *declarative*. (Rumelhart and Norman, 1983) provides a good, up-to-date discussion of the declarative/procedural distinction.

[6] Currently, there are 45 tasks and 80 metarules; thus the procedure is highly structured, with relatively few steps or methods for achieving any one task.

3. *iterative, try-all*: the metarules are applied in order, repetitively, until no rule succeeds (a simple loop; NEOMYCIN currently has no tasks of this type, probably because "try-all" suggests constantly changing methods or following a breadth-first approach).

4. *iterative, don't-try-all*: the metarules are applied in order, with control returning to the head of the list each time a rule succeeds, until no rule succeeds (a "pure production system").

The "if part" of a metarule generally examines the working memory and domain knowledge. The "then part" invokes another task, applies a domain rule, or requests a finding of the informant.

A task generally has an argument, known as the *focus* of the task, that part of the working memory it is operating upon (a finding, hypothesis, or domain rule). A task can have only one focus, but it might be a list, such as the entire differential.

A history is kept of which tasks have been done, recording the focus, if appropriate. Metarules reference this history, for example to determine if a particular hypothesis has been pursued. Other bookkeeping, such as resetting global registers that characterize the state of the differential, is handled by rules applied before or after the task metarules.

A task may have an *end condition*, which is evaluated whenever a metarule succeeds. If it is satisfied, the task is aborted. Importantly, end conditions can be inherited from tasks higher on the stack, and each task along the way will be aborted. End conditions describe either *preconditions*, which must be true for it to make sense to be doing the task (see end condition of EXPLORE-AND-REFINE) or *what the task is trying to achieve* (when it can be halted--see GENERATE-QUESTIONS). NEOMYCIN'S end conditions all refer to the differential: the presence of strong evidence for a "competing" hypothesis; the presence of a hypothesis in a new, unexplored category; an "adequate" differential to begin a diagnosis. Some tasks are always allowed to go to completion (indicated by an end condition of DONTABORT). We can think of the end condition mechanism as a means for "backing out of a procedure" when it becomes inappropriate or its goal is no longer of highest priority.

In summary, the knowledge for applying tasks--knowledge for controlling metarules, focusing, bookkeeping, and interrupting--constitutes a knowledge base in its own right.

Figure 4-1 summarizes how the diagnostic procedure interacts with domain knowledge. Figure 4-2 shows a task definition and a metarule expressed in internal form, using the MRS language, a form

of predicate calculus (Genesereth et al., 1981). (In MRS notation, **$X** will match whatever term is in the database and once bound will maintain that value in the rest of the expression). Note that intermediate relations, such as "active hypothesis," are also defined by rules written in MRS. Further details about the advantages of the MRS notation and NEOMYCIN's procedural language for representing strategy appear in (Clancey, 1985a).

In general, new strategies are expressed by writing new metarules and tasks and defining appropriate new structural relations for indexing domain knowledge. In summary, the control language constructs include: tasks, controlled metarules, problem-solving history, end conditions, primitive actions (ask, conclude, apply a rule), and a relational language for organizing domain knowledge (referenced by the conditional part of metarules). Domain knowledge and its organization is considered in the next section.

4.2. Representing domain knowledge: States, relations, and strengths

The domain knowledge consists of states, unary and binary relations defined on states and other relations, and information about the strength of relations.

4.2.1. States

There are two kinds of states: findings and hypotheses. *Findings* are observations describing the problem. There are two kinds of findings: soft (circumstantial or historical) and hard (laboratory or direct measurements). *Soft findings* tend to be categorical, weak, and easily determined. *Hard findings* are specific, strong, and often costly, dangerous, or time-consuming to determine. *Hypotheses* are partial descriptions of the disorder process causing the findings, that is, hypotheses explain the findings and constitute the problem-solver's diagnosis.[7]

4.2.2. Causal and subtype relations

Findings and hypotheses can be related by cause and subtype. Various larger structures are built out of these parts:

- *Etiological taxonomy* -- a subtype hierarchy of hypotheses. These are the ultimate causes of disorders. For example, in medicine, these hypotheses include poisoning, an injury from falling down, infection by a virus, and psychological problems (refer to Figure 3-3). Associated with each hypothesis are findings or other hypotheses that it causes or

---

[7] Technically, distinctions among states, such as "hypothesis," "soft finding" and "red-flag finding" are unary relations, which we express in metarules as (HYPOTHESIS $STATE), (SOFT-FINDING $STATE) and (RED-FLAG-FINDING $STATE). The states themselves are relations (e.g., (HEADACHE $PATIENT)), though as shorthand we write them as atomic propositions (e.g., HEADACHE). Thus, we write (HYPOTHESIS HEADACHE), rather than (HYPOTHESIS (HEADACHE $PATIENT)).

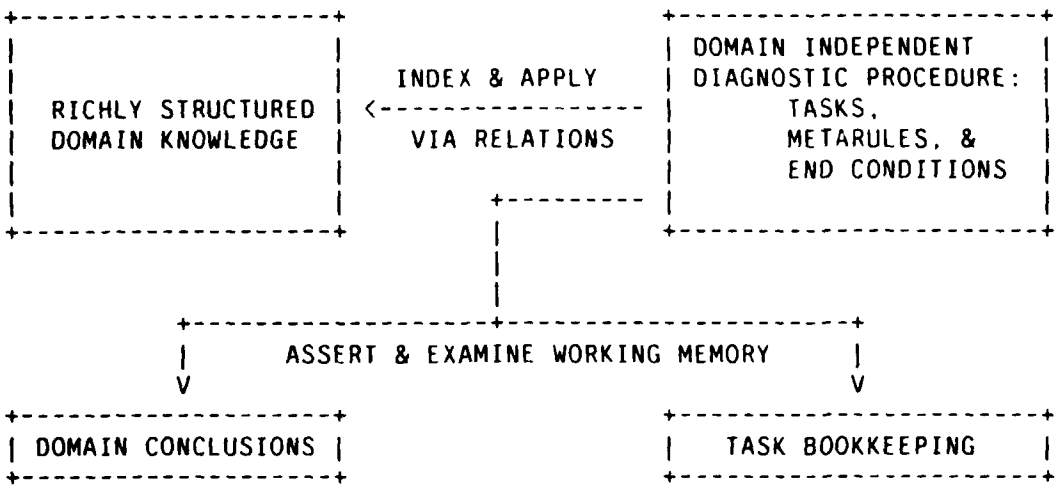


Figure 4-1: Interaction of working memory with domain and strategic knowledge: A domain independent language of relations partitions domain knowledge, enabling a domain independent procedure to index and selectively apply facts

<Task Control Knowledge>

```
(TASKTYPE PROCESS-FINDING SIMPLE)
(TASK-TRY-ALL-RULES PROCESS-FINDING)
(ENDCONDITION PROCESS-FINDING DONTABORT)
(TASKFOCUS PROCESS-FINDING $FOCUS-FINDING)
(LOCALVARS PROCESS-FINDING (RULELST SUPERFINDINGS FOCUSQS))
(ACHIEVED-BY PROCESS-FINDING (METARULE069 ...))
(DO-AFTER PROCESS-FINDING (RULE381))
```

<Typical Metarule>

```
(IF (AND (SOFT-FINDING $FOCUS-FINDING)
         (ACTIVE-HYP $HYPOTHESIS)
         (EVIDENCE-FOR $FOCUS-FINDING $HYPOTHESIS $RULE $CF)
         (UNAPPLIED $RULE))
    (TASK APPLYRULE $RULE))
```

<Auxiliary Rule>

```
(IF (OR (DIFFERENTIAL $HYPOTHESIS)
        (AND (DIFFERENTIAL $H1)
             (CHILD $HYPOTHESIS $H1))
        (AND (DIFFERENTIAL $H2)
             (TAXONOMIC-ANCESTOR $HYPOTHESIS $H2)))
    (ACTIVE-HYP $HYPOTHESIS))
```

Figure 4-2: Internal form of the task PROCESS-FINDING and one of its metarules ("apply rules using the finding to conclude about a hypothesis in focus")

are caused by it. Hypotheses lower in the tree inherit properties of all hypotheses on the path to the root ("ANY-DISORDER"). Thus, bacterial-meningitis has manifestations common to all infectious processes, such as fever and inflammation. The leaf-node hypotheses are the most specific causes, usually those that can be treated to alleviate the disorder.

The etiological taxonomy is actually a "tangled hierarchy" based on process relations. Proceeding below INFECTIOUS-PROCESS, the relations of each level are: "location," "chronicity," "class of causal agent," and "causal agent." For example, children of MENINGITIS are ACUTE-MENINGITIS and CHRONIC-MENINGITIS. Thus, each level of the taxonomy further characterizes *the kind of process* in some way. Under this interpretation, the top level of the etiological hierarchy pertains to events in the life process of the device: design, birth, ingestion, growth, injury, etc. We have found this characterization of the etiological taxonomy to be useful in our initial attempts to apply it to computer software diagnosis.

There may be multiple etiologies requiring treatment. For example, a traumatic injury, such as falling and hitting one's head, can cause certain forms of bacterial meningitis. Here the treatable cause is really two etiologies: the bacteria must be treated and, if the patient is elderly, some means must be found to prevent the patient from falling again. (In medicine, this relation is sometimes called a "complication" (Szolovits and Pauker, 1978).)

- *Causal network* -- hypotheses that characterize general states, neither findings (directly observed) nor etiologic hypotheses (pertaining to specific processes), which are related by cause. To give them a name, we call these general characterizations of abnormal conditions in the device *state/categories*. An example in medicine is "unusual space-occupying substance in the brain," a non-observable condition, which can have many etiologies. We have found it useful to distinguish between *substances* (or structural features) and *processes*. This does not lead to a complete causal model, but it does provide a useful discipline for our level of representation.[8]

- *Hypothesis subtype hierarchies* -- hypotheses (either etiologic or state/category) related by subtype. For example, INTRACRANIAL-MASS has subtypes INTRACRANIAL-TUMOR, INTRACRANIAL-HEMATOMA, and INTRACRANIAL-MASS-OF-PUS. Substances are subtypes of substances; processes are subtypes of processes.

---

[8]One potential difficulty is that this representation is more principled than common medical knowledge. For example, in some cases we found that our expert made no distinction among a substance causing a lesion, the lesion itself, and its functional effects. Thus, a tumor is referred to as a type of lesion, a bit like saying that a pair of scissors is a kind of cut. Traversing a more articulated network may require different strategies than those used by the physician. Indeed, to turn the argument around, composition of relations through "compilation," or blurring of cause/subtype distinctions, as we observed in our expert, may be useful for efficient search. See (Clancey, 1985b) for further discussion.

- *Finding subsumption hierarchies* -- a presupposition hierarchy of findings. For example, HEADACHE subsumes HEADACHE-SEVERITY, HEADACHE-DURATION, etc., because consideration of headache severity presupposes that the patient has a headache. In NEOMYCIN, a subsumption hierarchy is just a concise way of expressing inference relations among findings. Subsumption can be further characterized by relations such as "component of" and "specialization of"--distinctions we have not yet found to be useful for performance, but that might be useful for teaching.

4.2.3. Source, world-fact, definitional and process relations

Other domain relations are:

- *Source* -- a finding can be the source of a set of findings that are collected together. For example, the complete blood analysis is the source of the white cell count.

- *World-fact* -- findings can be related by factual relations based on what is usually true about the world. For example, males do not become pregnant; we can't determine directly if a 1 year old has a headache; adults do not frequently suffer from ear infections. Because there tends to be a different underlying relation for each case we have encountered, this knowledge is currently proceduralized in NEOMYCIN in the form of "don't ask" rules. For example, "if the patient is under 2 years old, don't ask if he has a headache."

- *Definitional* -- a finding can be defined in terms of other findings. For example, a neonate is a person under five months of age.

- *Process feature* -- a finding or hypothesis can characterize in more detail the process partially described by another finding or hypothesis. For example, the patient's temperature characterizes the finding that he has a fever. A pain can be characterized by location and change in severity over time. Every hypothesis in the etiological taxonomy can be characterized by a set of similar process features. Thus, each process feature constitutes a relation upon which a generalization hierarchy can be based. For example, an organ-involvement hierarchy of hypotheses is based on a hierarchy of locations. (While our work has clarified these distinctions, in our limited domain and with our current knowledge base, we use such multiple hierarchies only in the most limited way.)

Figure 4-3 summarizes how findings and hypotheses can be related.

4.2.4. Strength of a relation

Associated with causal relations is a "certainty factor" (CF), as used in MYCIN. For convenience in associating a CF with a causal relation between states, and to signify that the association is a heuristic that omits details, the relation is called a *rule* and given a name. For example, "double vision

```
FINDING  subsumes
         is source of
         is further characterized by (process features are)
         defines
         is usually related to (don't ask when)

       FINDING  is evidence for (causes or caused by)

              HYPOTHESIS  has process subtypes
                          is etiologic parent of
                          is caused-by

                     HYPOTHESIS
```

Figure 4-3: Summary of basic domain relations in NEOMYCIN.

is caused by increased intracranial pressure" is a rule with CF 0.8. We call the "if-part" of the rule the *premise* and the "then-part" the *conclusion*.[9] A rule premise is stated as a conjunction and each part involving a finding or hypothesis is called a *conjunct*.

Certainty is dynamically propagated through the network of states by a fairly complicated scheme. Basically, the maximum positive certainty is propagated upwards and the minimum negative certainty downwards through the multiple hierarchies. Assuming a closed world, a parent will be negative if all of its children are negative. Assuming mutual exclusivity, a sole believed child will inherit all the belief of its believed parent. The "cumulative" CF used in reasoning combines the CF directly inferred from rules with the propagated certainty.

A rule whose strength is very strong might be labeled as being an antecedent or trigger rule. These are defined in terms of activation criteria:

- A causal relation that is *definite*, having a certainty of 1.0, is generally labeled as an *antecedent rule*, so named because the rule will be considered, as part of the program's forward reasoning, when the premise of the rule is known to be true. For example, the double-vision rule is so labeled, so the program will conclude that the patient is experiencing increased intracranial pressure just as soon it learns that the patient has double vision.

- If an antecedent rule is also labeled as a *trigger rule*, then the program will attempt to satisfy the premise of the rule (by gathering additional findings if necessary), as soon as

---

[9] Technically, we should call the "if-part" the *antecedent* and the "then-part" the *consequent*, but we reserve these terms for characterizing the indexing schemes for applying rules.

some specified part of the premise (one or more conjuncts) is satisfied.

4.3. Implicit constraints of the diagnostic procedure

Metarules for tasks, as well as subtasks in the action of a metarule, are often ordered, and the criteria for this ordering is not explicit in the model. These ordering criteria are *constraints* which the problem-solver is trying to satisfy or which are imposed by his reasoning ability. From our study of the metarules, we have identified several sources of constraints in diagnosis:

- *Cognitive Economy*--to incur the least costs in terms of mental effort, acting within the constraints of human memory and reasoning capability, specifically:
    - the size or organization constraints of memory for holding the current problem description and partial solution ("working memory"),
    - the organization of domain knowledge ("long-term memory"),
    - the manner in which knowledge is retrieved ("activation criteria").
- *Computational or mathematical constraints*--properties of combinatorial, categorical, and probabilistic search.
- *Assumptions about the world*--disorder patterns, determined by the frequency of problems previously encountered, in turn determined by device weaknesses and external influences on devices. These assumptions or expectations can be used to constrain search.
- *Sociological economy*--to make the correct diagnosis, with the least expenditure of money and time, with due regard for the value placed on life and equipment, and efficiently communicating information needs and decisions.

In using a categorical search, asking general questions first, requesting hard data sparingly after consideration of soft data, maintaining focus until leads have been exhausted, etc., the problem solver is satisfying these constraints. We make an attempt in Appendix IV to indicate how the constraints are evidenced by individual metarules and their ordering. The main constraints of concern are correctness, efficiency (speed), and minimizing mental effort. Correctness is best evidenced by the systematic search of ESTABLISH-HYPOTHESIS-SPACE; efficiency, by the categorical reasoning of GROUP-AND-DIFFERENTIATE and the use of general questions by FINDOUT; and minimizing mental effort, by the nature of focus changes in PROCESS-FINDING and EXPLORE-AND-REFINE. The constraints can also be grouped in terms of the problem solver's goals

(reflecting cognitive and sociological constraints) and constraints imposed by the task domain (mathematical and statistical).

Each task corresponds to some condition the problem solver is trying to make true: the metarules and task control knowledge constitute a procedure for making the condition true. We say that tasks *proceduralize* constraints (VanLehn and Brown, 1979), that is, they seek to *satisfy constraints by conditional actions*. For example, one of the correctness constraints relevant to EXPLORE-AND-REFINE is that all hypotheses placed on the differential must be pursued eventually. One of the ordered metarules for this task says, "If there is a sibling of the current focus that has not been pursued, then invoke PURSUE-HYPOTHESIS with the sibling as focus." Thus, subtasks with a given focus are invoked to satisfy constraints.

The structural properties of NEOMYCIN's domain knowledge reveal an interesting set of cognitive and task domain constraints. However, these properties are a strong reflection of the cases the model has been developed upon, so they are just a set of unrefuted or convenient (known to be false in general) assumptions.

- Every problem that will be encountered can be uniquely characterized in terms of some single disorder that has been diagnosed before (an assumption known to be false in general). These "etiologies" can be organized hierarchically in multiple ways, particularly according to process relations.

- Evidence for disorders is generally weak, requiring categorical reasoning and inheritance of belief.

    - There are no "deep" causal models that explain the normal functioning of the device's behavior (an assumption known to be false in general). Therefore, reasoning does not benefit from complete structural (anatomical) information about the device.

    - There are few "pathognomonic" findings, that is findings that clearly identify the disorder.

- Nevertheless, groups of findings strongly "trigger" hypotheses because of the high frequency with which the disorder exhibits that pattern of findings, the disorder's relatively high a priori probability over other hypotheses that explain the findings, and/or it is a serious and treatable disorder.

- Patterns in finding/hypothesis relations make it possible to characterize findings as "non-specific" vs. "red-flag," "a good general question," "a good follow-up question."

The tasks and metarules are deliberately formalized at a level of detail that will be useful for providing explanations to a student in a tutoring system. However, it is becoming apparent that constraint information is essential for deciding what parts of the model should be emphasized during teaching and what parts might differ with individual abilities and preferences. For example, we might explain student errors by systematically relaxing the constraints of the procedure. We are currently extending the model to include annotations that indicate: what is arbitrary and not part of the model (e.g., order of GENERATE-QUESTIONS metarules); what may reasonably vary among individuals (order of PROCESS-FINDING metarules); what no person could logically expect to do differently (doing FORWARD-REASON before information is received); what individuals might do differently, but which would violate the principles of the idealized model (e.g., doing EXPLORE-AND-REFINE before GROUP-AND-DIFFERENTIATE).

Note that NEOMYCIN's procedure doesn't reflect some of the most important constraints useful for the "present illness interview," namely the constraints of human interaction that require the problem-solver to paraphrase finding requests in multiple ways and to cross-check information ("interface constraints"). We assume that the informant speaks the model's language and is always reliable (see FINDOUT). Interactional methods for talking to patients is certainly a key part of what students learn in the classroom diagnosis games. In the six classroom transcripts we have analyzed, one-third of the teacher's interruptions (10 of 30) are directed at giving practical advice of this sort.

In summary, at this stage in NEOMYCIN's development we are developing a procedural language that enables the program to articulate its reasoning. By studying the procedures we write down in this language, we may become able to represent them at a more principled level, in terms of the constraints they seek to satisfy. (See (Clancey, 1985a) for a significant expansion of this point. Also see section 5.3.2 for a discussion of an expert's awareness of constraints on his behavior.)

5. Evaluating the model: Sufficient performance and plausible constraints

Having considered how NEOMYCIN's model is acquired and represented, we now turn to its evaluation: A general discussion of what the program really is, what it says about the nature of expertise, and what its limitations are. Evaluation is very difficult. At this time, we can only hope to explicate the issues and discuss how we're handling them, rather than describe formal, completed experiments.

In considering evaluation, we take NEOMYCIN as it exists today as an incomplete artifact, and we ask,

"What is it?" What kind of program is it? What is its basis in fact? What does it tell us about human reasoning? About knowledge engineering? About computational modeling? This is an opportunity to take stock of the enterprise, criticize the program, and try to determine what has been accomplished.

Four perspectives are useful for evaluating the program, to be considered in this order:

1. *Performance:* Does the program run? Does its behavior (question asking and diagnosis) suitably match, on some domain of problems, *the expert behavior we seek to model?*

2. *Articulation:* Is the level of explicitness of the representation appropriate? Do the program's explanations of its behavior correspond to the statements made by an expert teacher explaining the tasks and rationale of diagnosis to students?

3. *Accuracy:* Does the program model human reasoning? Are the constraints of the tasks what experts seek to satisfy in their problem solving? Are the implicit assumptions about correctness, efficiency, and cognitive economy justified?

4. *Completeness:* Is the program a comprehensive model of diagnostic reasoning? Are the domain knowledge structures and search techniques complete for some domain of problems?

The first two perspectives are concerned with the *sufficiency* of the model for different settings requiring expertise (refer to Figure 1-1 in Section 1). The second two perspectives examine whether this is a *plausible* model of human competence and whether it fully captures the full range of human diagnostic behavior. We evaluate NEOMYCIN's acquisition and representation from these perspectives in the sections that follow.

5.1. Performance of the model: Problem solving

Perhaps a non-trivial point, a pre-requisite for claiming that NEOMYCIN is a model at all is that it runs: It "computes" behavior that we can match against the behavior of people. This is a property of the representation of the diagnostic procedure; it is structured into recursive subprocedures, with control information for stopping and printing results. Its activities are to gather information and construct a *solution*. Contrast this with the constraints (given in Section 4.3) which the tasks implicitly satisfy. Such statements might capture what problem solvers try to accomplish and the background in which they work, but they do not specify the *process* by which consideration of specific domain knowledge and actions taken in the world interact. NEOMYCIN's metarules combine considerations of domain knowledge (via indexing relations) and working memory to conditionally invoke the right subtasks (with the right focus) to satisfy the task constraints.

NEOMYCIN solves problems at least as well as MYCIN. In particular, its conclusions are reasonably close to MYCIN's for the ten cases used in a double-blind evaluation of MYCIN (Yu et al., 1979). However, we demand much more of NEOMYCIN. Unlike MYCIN, it should:

- Reason in a focused, hypothesis-directed way. For example, if the infection is chronic, it should not explore acute subtypes of meningitis. In contrast, MYCIN's question-asking is undirected and exhaustive for all types of meningitis.

- Consider meningitis from initial information and decide what tests to request, such as a lumbar puncture. MYCIN is told that the patient has meningitis and that certain laboratory tests are available. NEOMYCIN must begin with more general, non-specific findings, such as "headache" and "malaise," consider meningitis, and decide when a lumbar puncture would be too dangerous to do.

- Consider competitors of meningitis and know when they are more likely. MYCIN has no knowledge of migraine, tension-headache, brain abscess, etc. NEOMYCIN carries on a "differential diagnosis," knowing when to consider these competitors and how to contrast them.

- Reason more generally about findings, for example, determine what lab test to request, based on subtype and definitional information.

There are other differences in performance (e.g., as specified in the task FINDOUT and FORWARD-REASON), but these are the main ones. Our main technique for testing (and developing) the program is to run cases with different correct diagnoses, but having very similar initial findings. This tests the program's ability to elicit relevant additional information and to adopt different lines of reasoning appropriately. Trivially, the program should not always pursue meningitis. The same evaluation technique is essential for measuring completeness of the model as well. Evaluation of the order of questioning pertains most closely to matters of accuracy and is considered in that section.

A not-insignificant question is, "Why does NEOMYCIN work correctly at all?" There are two aspects to this. First, how can abstract explanations given by a physician (e.g., "look for associated symptoms"), coded as tasks and metarules, produce the right answer? Second, what is the nature of reasoning that allows us to completely separate the domain knowledge from the reasoning procedure? The issue of explanation is treated here; the more general characterization of reasoning is treated in the final section of the paper.

It is plausible that the expert's explanations should constitute at least the outline of an effective procedure. Recall from Section 3 that all behavior is explained in terms of the *effect* it will have on the

expert's thinking. He says, "I'm trying to form and test my hypothesis set in some way." Indirectly, we take this to be his general *task* at that point--what he is trying to do--and write rules that will invoke that task and carry it out. A procedure written to have *the same effects on working memory* will generate the same questions as the expert, with the same final diagnosis, and can be characterized abstractly by the same explanations supplied by the expert.

The question has a deeper side, however. Do NEOMYCIN's metarules really come from the expert? What do we supply from our knowledge of the constraints of diagnosis? All of the major tasks bear some relation to the expert's explanations, visible most clearly in the classroom discussions when he tells students what they should and should not be doing. (Recall the examples in Section 3.7.) Most of the rules for FORWARD-REASON, FINDOUT, and ESTABLISH-HYPOTHESIS-SPACE are *inferred* from conclusions the expert states and the questions he asks. But the nature of the inferences are different. For example, FORWARD-REASON and FINDOUT consist of lists of metarules using straightforward domain relations such as SUBSUMES. That is, we inductively abstract patterns from expert behavior, based on our evolving knowledge of the relations among findings and hypotheses. The simple co-appearance of findings in a problem solution is often sufficient to suggest metarules. (For example, the subsumption relation among findings suggests why "travel" would be mentioned at the same time as "lived in Mexico.")

However, ESTABLISH-HYPOTHESIS-SPACE is a *procedure involving search of a taxonomy*. We have to infer both the domain relations and subprocedures from patterns in the expert's questions. Explanations point the way at critical times, and the classroom discussions seem to confirm most of our analysis, as strategies we learn inductively are often stated explicitly in class (particularly the idea of looking up, then down the etiological taxonomy). But, most of our confidence in the completeness of the procedure is based on *mathematical considerations of set manipulations*, concepts the expert never mentioned. The idea of getting the right answer into the differential, even at just the highest categorical level, and then winnowing down makes good mathematical sense. In this way, the metarules are designed to work: The constraints of set theory are adhered to at every turn.

In summary, NEOMYCIN's model is not supplied directly by the expert. It is *constructed* by relating his behavior to mathematically logical maneuvers within the data- and hypothesis-driven reasoning scheme. However, our views are strongly guided by the expert's emphasis on what he is trying to do--what new evidence can accomplish in terms of getting the right answer.

The relation of *empirical* and *rational* approaches for constructing a model has been a subject of much debate (e.g., see (Anderson and Bower, 1980)). Our methodology is summarized in Figure 5-1.

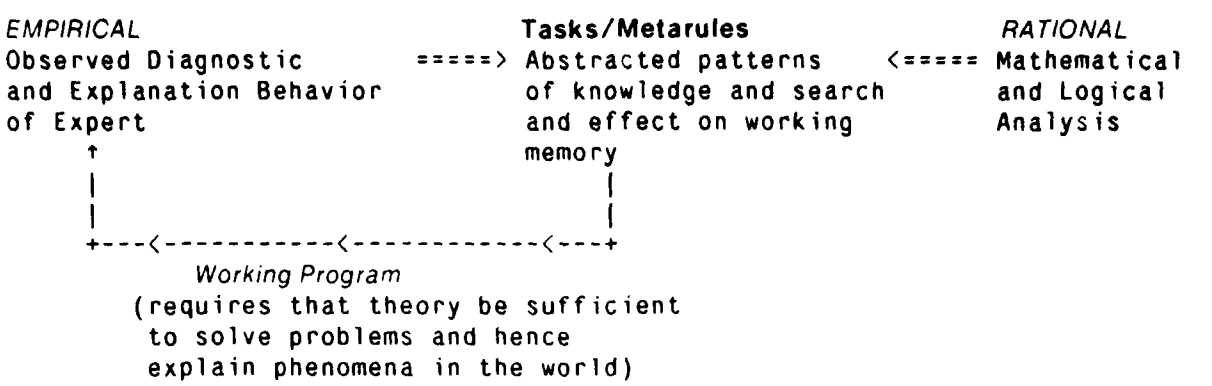


Figure 5-1: Combined empirical and rational methodology
[After (Anderson and Bower, 1980)]

Given the logical basis for much of the model, we might wonder whether we could construct a proof that the program will always output the right diagnosis. One approach is to break the proof into parts:

1. Prove that the hypothesis that explains the findings or some more general hypothesis will be put in the differential.

2. Prove that it and its ancestors will be examined, and

3. Prove that it will be refined to its subtypes and causes.

There are many subtle interactions to consider. For example, considering a hypothesis requires inferring evidence for it by some rule. A rule not applied immediately might be considered later. If a rule is not a trigger rule, it still might be invoked by the GENERATE-QUESTIONS task, but this task won't be invoked if the differential is already "adequate." Thus, a hypothesis might not be considered if belief in some alternative explanation is strong enough. Also, the problem ultimately reduces to proving that the knowledge base's finding/hypothesis relations are complete and correct, a difficult assumption to start with and difficult to prove independently.

However, this analysis can be used to complement the usual test of running cases. Stepping through it, we discovered that NEOMYCIN did not examine ancestors of state/category hypotheses--a GROUP-AND-DIFFERENTIATE metarule was missing. We conclude that this approach is a worthwhile cross-check for developing the model.

5.2. Performance of the model: Articulating reasoning

Evaluating the explanation capability of NEOMYCIN is perhaps best done in a tutorial setting. Does the program use appropriate terminology? Does the program explain its question-asking with appropriate generalizations? A prototype explanation system demonstrates during problem-solving that the program's level of representation is apparently close to the terminology used by the expert (Hasling, 1984). Major explanation issues as we begin to use NEOMYCIN for teaching include: The proper mix of abstract and concrete statements, terminology (e.g., task names like ESTABLISH-HYPOTHESIS-SPACE have to be restated), and use of a model to selectively present and summarize reasoning.

One very interesting test of the ability of the program to articulate its reasoning involves use of a "student modeling" program. We have transcripts of discussions of six cases in a classroom, in which one student interviews (and diagnoses) another student who is pretending to have a particular illness. Can we combine a program that uses NEOMYCIN's model with some (hopefully) simple pedagogical rules, to predict not only when the teacher will interrupt the student/physician but (because of model violation) predict as well what he will say? To do this, we would need more case discussions in NEOMYCIN's domain or would need to expand the program's domain of expertise.

5.3. Accuracy of the model

By reducing the metarules to constraint assumptions, and separating out accuracy of the *implementation* of the constraints, arguments about accuracy reduce to showing that the principles upon which the model is based are valid. NEOMYCIN's design, in which the reasoning procedure is stated in a special, well-structured language, completely separately from the domain knowledge, helps makes these principles clear. We start by writing down how knowledge, working memory, and task behavior interact, then we study what we have written down. With the components of the model factored out this way, each can be examined for plausibility: Could human knowledge be structured hierarchically with multiple indices? Could working memory include a list of hypotheses? Does NEOMYCIN allow its differential to get "too long"? Is the recursive, single-argument invocation structure of tasks plausible? Similarly, we might evaluate the end condition mechanism, means for restoring context, etc. In fact, there are three considerations, though with some common constraints: the *task/metarule control language*, the *content of the metarules*, and the *representation of domain knowledge*.

5.3.1. Competitive argumentation

Our primary technique for constructing the model is a form of "competitive argumentation" described by Van Lehn (VanLehn, 1984, VanLehn, 1983). We enumerate alternative designs and choose among them in a principled way. For example, in the extended protocol (Appendix II, line 5), observe that the expert mentions evidence for increased intracranial pressure and goes on to use this information immediately. When NEOMYCIN was first given this case, it gathered additional information because "diplopia" did not make increased intracranial pressure certain. Why didn't the expert do this? We list some alternative "designs":

1. The expert *had* made a definite conclusion; NEOMYCIN's evidence rule is incorrect.

2. The expert knew of nothing that could disconfirm his current belief in increased intracranial pressure, and he believed that the current evidence was fully reliable, not susceptible to retraction. So there was no need to gather additional evidence: the current belief was high enough to be useful in any way.

3. The expert used the information tentatively, planning to try to disconfirm the hypothesis or the single finding upon which it was based, should this conclusion play a pivotal part in the final analysis (e.g., should it suggest that an dangerous, invasive test is necessary). That is, he is capable of retracting conclusions and reconsidering his decisions.

Having listed these, we can now argue about whether other alternatives should be included, as well as which is most likely. Furthermore, given that most researchers would probably opt for the third ("allow retractions") alternative, and NEOMYCIN now uses the second ("assume reliability"), we can proceed to construct cases in which the program's behavior would fail to be an accurate model of how people reason, thus testing the hypothesis that NEOMYCIN is inaccurate in a particular way.[10]

5.3.2. Difficulties of extracting principles from compiled knowledge

One effect of experience is that simple domain facts are proceduralized into specific rules for using them and rules for controlling reasoning are composed and generalized. This effect is called "knowledge compilation" (Neves and Anderson, 1981). In attempting to formulate a competence model, we want to carefully decompose these rules and state how knowledge is used, separately from the facts themselves. That is, we want to "decompile" expert knowledge, to the extent possible, to get at the primitive knowledge organization and control that lies behind it. Evaluation of accuracy of

---

[10] Indeed, taking this example, the inability to change conclusions that have been used to form other conclusions is very basic. We should examine the entire model critically from this perspective. For example, we are probably missing FORWARD-REASON metarules that detect that a prior conclusion must be changed or task interruptions (end conditions) that trigger reconsideration of the patient model

the model takes place at this lower level.

However, separation of domain facts and abstract control may be difficult if compilation occurs in a principled way. A result of compilation might be systematically mistaken for a new principle, a primitive step of the diagnostic strategy. For example, consider a case in which a finding counts against a hypothesis. Suppose further that the hypothesis has not been considered yet, but is a child of some hypothesis that is about to be refined. Now, would the negative evidence be *consciously* noticed by problem solver at refinement time, when the children are logged as hypotheses to pursue (placing them in the differential), or would it not occur until the problem solver focuses on that hypothesis and tries to confirm it? (Similarly, if you are using an agenda, do you note the evidence while putting the task of pursuing the hypothesis on the agenda [and decide not to schedule it], or when you go to do the task?) There appear to be no simple answers. It all depends on how long ago the finding was revealed, what the problem solver was thinking about at the time, how strongly he is swayed by other hypotheses, etc.

A similar example suggests that we are dealing with a general problem about attention and focusing. Does the problem solver notice that a task such as testing a hypothesis is trivially done in some context when looking for a new focus (e.g., in EXPLORE-AND-REFINE when examining hypotheses to pursue). Or is this noticed after the operation is scheduled and begun? Put another way, should the metarule predicate do look-up only and require the invoked task to observe and record completion?

In an expert, compilation of knowledge probably combines scheduling and task behavior. In a novice, the separation might be more complete, so his behavior is methodical, but rigid, clumsy, and inefficient by not being adapted to routine problems. This suggests that NEOMYCIN is a model of *competence*--what the expert is capable of doing (at the task level), rather than the actual operations (*performance*) he does for any given case. He is traveling on familiar roads and takes shortcuts that are compositions of primitive steps.

In building NEOMYCIN, it has been difficult to isolate unambiguous, principled paths by which the expert indexes knowledge. In some cases, more than one inference path is possible. Indeed, when information is useful for more than one inference path, it tends to become one of the "important general questions I always ask" rather than "something I need to confirm a specific hypothesis" (see Figure 5-2). In general, it can be unclear whether the expert is *indexing via findings*, asking things he knows will usefully modify his differential, versus *indexing via hypotheses* that he currently cares about. As expert reasoning tends to be more data-directed (Chi, et al., 1981), subgoals are set up by

"trigger rules" (see PROCESS-FINDING in Appendix IV), rather than arising from a hypothesis-directed line of questioning (TEST-HYPOTHESIS). Rubin's model (Rubin, 1975) and ours differ in this respect. In fact, trigger rules occupy an interesting mid-way point in our model: They are a form of "compiled" knowledge that beginners need to be taught immediately if they are not to be extremely inefficient. Follow-up questions (CLARIFY-FINDING) are another manifestation of compiled knowledge that must be distinguished from deliberate attempts to confirm a hypothesis.

A model of competence is an idealized, "interpreted" statement of expert reasoning--the conscious steps an expert follows when reasoning in "careful" mode, rather than routinely solving problems. We claim that the expert's knowledge, full of shortcuts as it is, can be expanded into principled steps (or alternative principled procedures).[11] A principled procedure is an "interpretive simulation" in which the outward behavior of data requests and conclusions is matched, but many intermediate steps (e.g., decide to EXPLORE-AND-REFINE, choose a focus, REFINE-HYPOTHESIS, TEST-HYPOTHESIS, choose a finding) would only be consciously followed by a beginner (knowing the right procedure) or an expert faced with a difficult problem.

Furthermore, we must distinguish composition of procedure and medical knowledge with compilation of the medical knowledge base itself. As a set of schemas characterizing diseases, domain knowledge is knowledge of patterns in the world. The problem solver asks, "Of all the problems I have encountered in the world or am likely to encounter, what are the common causes, the serious findings, the general questions important to ask early on, important causes, and useful follow-up questions?" These patterns all relate to importance in terms of *usefulness* (of a finding, based on the number of evidence links or its ability to discriminate) and *likelihood* (of a hypothesis). Thus, by case experience or general knowledge of the problem population, associations are specialized and abstracted, moving to the level of *heuristic knowledge* as opposed to simple facts about cause and subtype. By some form of structural analysis, it may become possible to derive a theory of when a finding would be a good general, trigger, or follow-up question in a given domain. (See (Clancey,

---

[11] For example, we disallow a rule of the form, "Headache and fever triggers meningitis," because fever is evidence for an infection and meningitis is a kind of infection. The link between fever and meningitis should be made via propagation of belief from the parent, infectious-process. Otherwise, the evidence of a fever is considered redundantly. However, we allow a specialized rule stating "headache and high fever," or its more correct generalization, "headache and evidence for a fulminating infection," because the information about severity is not factored into the belief that the patient has an infection. In general, when we study a rule of the form "A implies B," we must always ask whether there is some hypothesis X in the knowledge base, where X implies B, meaning that the new rule should state that A implies X. In the example given here, we might also decide to have fever trigger infectious-process, and write an ordinary evidence rule of high CF that headache implies meningitis. If the patient has a fever, infectious-process will be triggered; meningitis will then be "active" and noticed should it become known that the patient has a headache (see PROCESS-FINDING in Appendix IV and the metarule stated in Figure 4-2).

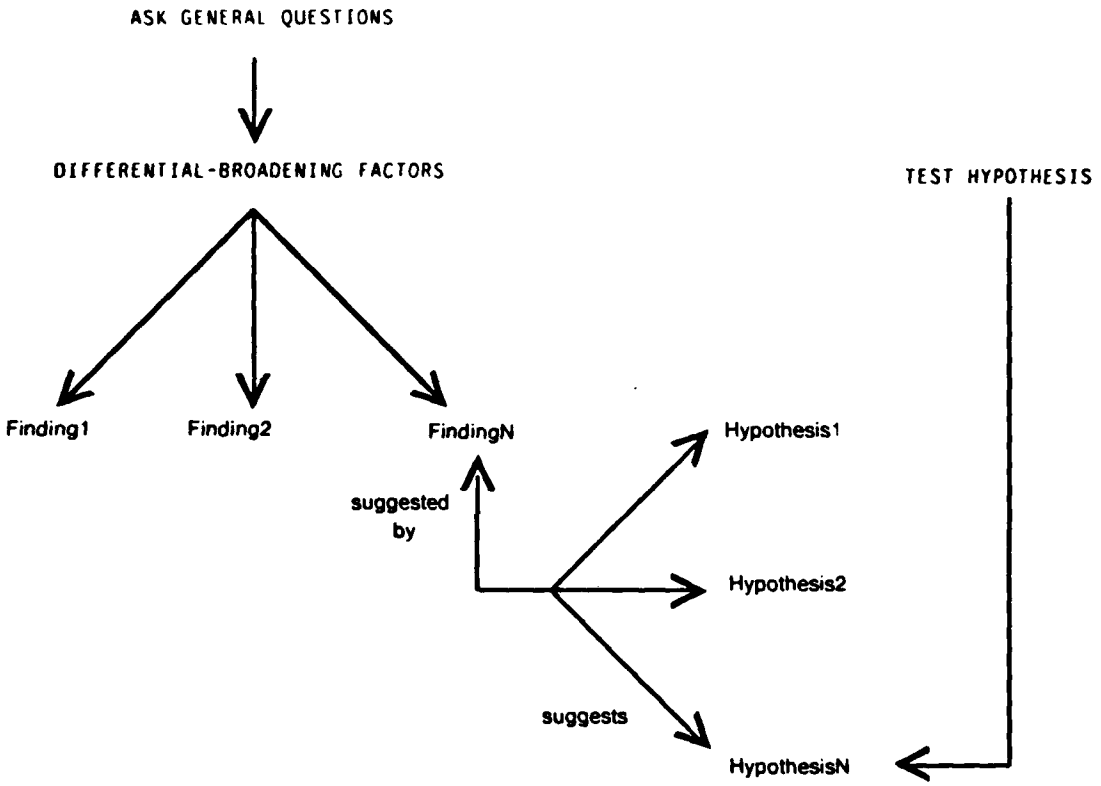


Figure 5-2: Finding request interpreted as a "compiled" general question or a deliberate attempt to confirm a hypothesis

1985a) for further discussion.)

In summary, in identifying primitive steps and knowledge relations in the diagnostic model, we need to clear about:

- **Kinds of knowledge.** Figure 5-3 summarizes the basic elements of NEOMYCIN's diagnostic model. The model consists of domain knowledge relations (kinds of patterns), reasoning tasks for using this knowledge (a classification procedure concerning focus and activation of associations), and constraints that could be used to derive the procedure (the rationale for the procedure).

- **Kinds of "knowing."** We claim that a good teacher knows the domain relations and the general tasks for manipulating the differential. He can talk about this knowledge; it is not just reflected in his behavior. In classroom explanations, the teacher also mentions many social constraints, as well as some logical constraints (regarding search of trees) and some case experience constraints (such as correlations among findings). This is the substance of what we want to teach students.

  However, some of parts of NEOMYCIN's procedure, particularly FORWARD-REASON, describe what experts do and are essential to construct a complete, runnable model. We believe that these tasks, corresponding to the "cognitive constraints," are generally not consciously considered by experts and needn't be taught. These tasks are not known in the same sense that "serious causes of sore throat" are known; they are automatic, they are how the mind does diagnostic classification. Perhaps FORWARD-REASON and its metarules are more a description of how the hardware works, rather than of a particular software program or strategy. Does ESTABLISH-HYPOTHESIS-SPACE fall in between, so that grouping and refining categories is automatic, but profits from conscious direction (to be aware of and cope with knowledge gaps)? Thus, given that NEOMYCIN is a model of what experts *do*, we must distinguish between the processor and the program, and then overlay a secondary description of what experts *know about what they do*.

  We might conclude that a good teacher knows much more about problem solving than the average practitioner. But it is interesting to conjecture that the mark of an expert is precisely this *metaknowledge* of how he reasons: He knows that there are procedures, that these procedures derive from constraints that problem solving must respect, and that there is a mode of reflective reasoning for checking his behavior for completeness and consistency, both for solving difficult problems and justifying his conclusions (teaching).

- **Origin and development of knowledge.** As discussed in this section, associations can be learned directly by rote (e.g., trigger rules), composed from primitive associations (e.g., headache and fever suggesting meningitis), generalized from experience (e.g., patterns of serious causes of a disease), or instantiated from more general principles

(e.g., testing a given hypothesis might be learned as a specific set of things to do, following the principles for testing any hypothesis in general). Complicating the analysis, *what is compiled from experience* by one problem solver might be taught by rote to another. Finally, in relating behavior to motivational principles or a plan, we must remember that even a sequence of behavior could be generated by more than one plan. It is even possible that automatic behavior is non-deterministic, in the sense that the problem solver's actions are explained by multiple plans (compiled paths of association) and no single intention consciously produced his actions.[12]

The decomposition of knowledge types in NEOMYCIN has allowed us to make substantial progress towards characterizing what physician teachers know and communicate with their students. However, we have barely begun to properly account for the origin and development of this knowledge.

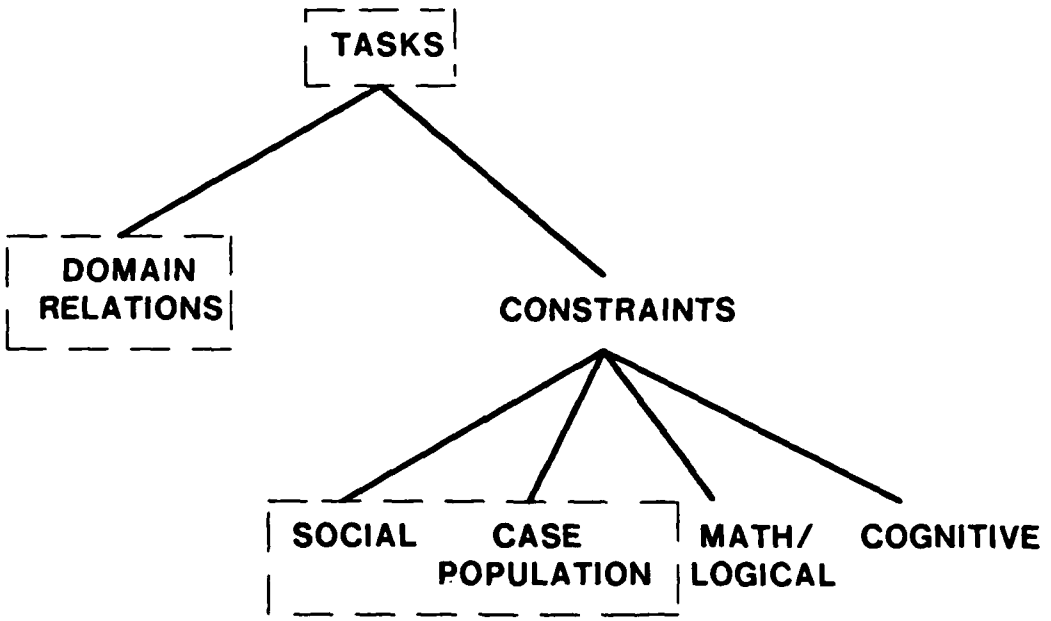


Figure 5-3: Types of knowledge relating to diagnostic strategy. Boxes indicate what a physician teacher can articulate.

[12] John Seely Brown, personal communication

5.3.3. Using a competence model to explain variant behavior

By assumption, the "careful mode" of reasoning is principled. A good way to extract these principles is to give experts difficult problems. In this way we characterize the nature of expertise and how experts and novices might differ. In particular, as already suggested, a principled analysis of mechanisms has real relevance for explaining errors that people make in diagnosis.

A good example of a principled error appears in the classroom excerpt of Figure 5-4. Several students are interviewing the student W1, who is pretending to be a patient. The students' questions about sore throats are not random. The students appear to be looping in the task of CLARIFY-FINDING, following the principle of characterizing a finding in terms of the process (see Figure 5.3.3, parse 1). The error or misconception is that not every process question you might ask will be useful. If the students know the strategy of characterizing a finding, they are applying it at the right time with the right focus, but their knowledge base is not right: What are the useful follow-up questions to ask about a sore throat? In fact, there might not be any in general; instead a causal analysis should be undertaken (form a hypothesis and test it).

Given that the "useful follow-up questions" are determined by case experience, this analysis suggests that some parts of "compiled knowledge" may normally be taught directly, rather than learned from experience. That is, *experiential knowledge--knowledge about how to efficiently solve problems given a certain population of cases--may be learned by apprenticeship, rather than individual practice.* Trigger rules and useful general questions, two other forms of "compiled knowledge" in NEOMYCIN, are probably also taught directly to students.

An alternative analysis of the sore throat protocol is that the students might not know what causes a sore throat, so their differential is inadequate. They might be following the strategy of ELABORATE-DATUM, a subtask of GENERATE-QUESTIONS, attempting to elaborate known symptoms until some new clue triggers a hypothesis. This illustrates how we might explain student behavior in a principled way in terms of the expert's diagnostic procedure operating on different domain knowledge. Having stated the procedure separately from the medical knowledge, we have a basis for inferring what students are doing, the state of their working memory (e.g., an inadequate differential), and hence their knowledge of domain relations. Thus, even if we don't need to teach the diagnostic procedure, it is useful for motivating teaching of domain facts and detecting deficiencies.

We can of course generate an infinity of interpretations if we relax the assumption that the student's procedures are correct. For example, perhaps stuck with an inadequate differential, the students don't know enough to do GENERATE-QUESTIONS, but are instead attempting to "repair"

W2: Have you had a lot of sore throats?

W1: No.

M1: So your throat is getting worse? Is that what you are saying?

W1: Well, it's really bothering me and it just keeps dragging on. And before when I've had a sore throat, I had it for a few... a couple days.

M1: I see.

W1: It would be gone, but it just keeps dragging on and I'm just feeling terrible.

M2: Does anything make the sore throat better? Have you tried gargling?

W1: Um, well I haven't really done too much about it. I just thought it would go away, but it hasn't and as they said I'm just... I'm feeling really tired and not feeling very good.

M1: Your sore throat is always as painful when you get up in the morning or is getting worse during certain time of the day?

W1: Well I guess I haven't noticed too much difference.

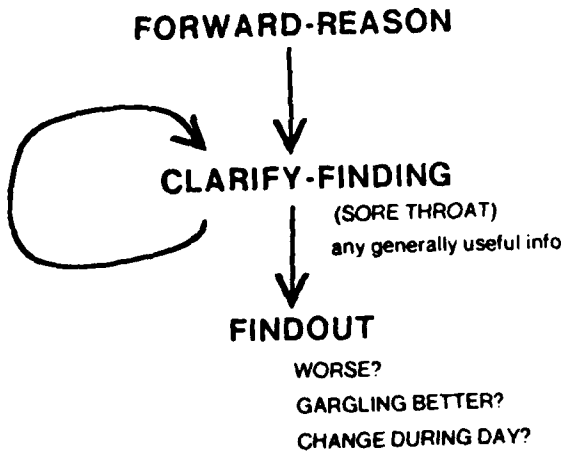
M1: I see.

TEACHER:
Let me ask you a question. When you ask these questions about whether gargling makes it better or worse, or whether it's better certain times of the day, are you thinking about how that's going to help you move down different differential diagnoses?

M1: Uh huh.

Figure 5-4: Classroom discussion illustrating a diagnostic error

ALTERNATIVE PARSE #1:
Same strategy, different knowledge



ALTERNATIVE PARSE #2:
Same strategy, different working memory

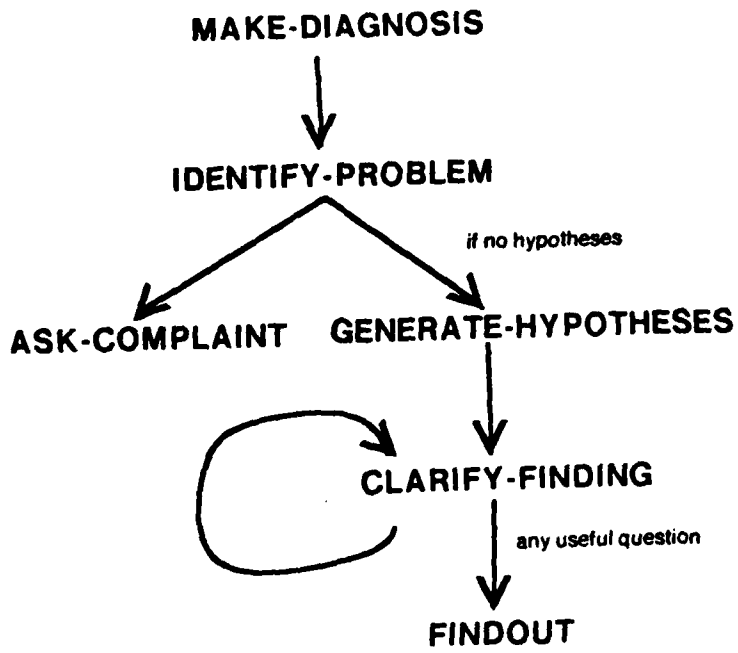


Figure 5-5: Alternative parses of student behavior shown in Figure 5-4

their procedure. They can't continue, so they are looping on the last successful operation. In addition, they might not know the useful follow-up questions to ask, but they know the principle that allows them to generate candidates. This kind of analysis could be pursued by competitive argumentation.

As another example of an incorrect procedure, consider the issue of when TEST-HYPOTHESIS can be interrupted. Suppose that a finding becomes known that is relevant to some hypothesis, previously considered, but that is not the current focus. Under what conditions does the problem solver notice the association and when will he actually shift attention to pursue the other hypothesis? Under one scheme, used by NEOMYCIN, "processing a finding" means deliberately widening attention to notice relevance to any activated hypothesis. Under another scheme, the problem solver might only observe relevance of findings to his current focus. The narrowly-focused problem solver might never realize the significance of data to other hypotheses he cares about.

The very notion of a "task" as something that the problem solver does deliberately, a thinking problem he imposes upon himself, allows us to distinguish among problem solvers according to the tasks they bring upon themselves in various situations, such as when a new finding is revealed. When distinctions in the model have implications for correctness of the diagnosis, it will be important that the model be annotated at this level of detail, so the teaching program can know and point out the important tasks the students are failing to do.

5.4. Completeness of the model

While "accuracy" is concerned with the correctness of the assumptions and constraints of the diagnostic procedure, "completeness" is concerned with coverage of the model: Does a wider population of problems require more problem-solving techniques? Given the association between metarules and constraints, this question approximates asking whether we have identified all of the relevant constraints that the task demands and taken into account all of the relevant capabilities of human reasoning.[13] As already stated, NEOMYCIN's problem domain does not require all forms of diagnostic reasoning that have been studied elsewhere. Without attempting to examine the underlying issues, we simply list many of the limitations we know about:

- Reasoning about structure and function of the body (Genesereth, 1984, Davis and Lenat, 1982).

---

[13] Naturally, testing the program for accuracy may suggest ways in which the program is incomplete (e.g., the possibility of retracting conclusions).

- Analogical reasoning using "device models" (Gentner and Stevens, 1983).

- Interview techniques for getting reliable information from laymen (e.g., common sense ways of detecting weight loss; finding out whether the patient has had rheumatic fever; knowing what the "white pill" is).

- Description of causality and disease processes on multiple levels of abstraction (Patil, 1981).

- Distinguishing among different forms of "subsumption."

- Temporal reasoning: onset and progression of disease.

- Using probabilistic information about findings, such as frequency information to bias and rule out hypotheses.

- Determining whether there is adequate evidence for a hypothesis should be contextual, taking into account other hypotheses and unexplained findings (Cohen and Grinberg, 1983).

- The problem solver must strive for a coherency by explaining the "important" findings and explaining findings inconsistent with each other or which violate expectations formed by his hypotheses. The program's "differential" should be a "case specific model" (Patil, et al., 1982) that merges findings and hypotheses.

- A real-world expert must deal with multiple, interacting, concurrent problems. The problem solver must separate causes from complications (Rubin, 1975, Szolovits and Pauker, 1978, Pople, 1982).

- NEOMYCIN's causal network is too simplistic to determine the completeness of its strategies. For example, when the causal connections between data and the taxonomy are long and complex, it is not feasible to follow each path (possible cause), testing and confirming intermediate states along the way (Pople, 1982). However, as mentioned in Section 4.2, such an articulated model may even require different strategies than used by people, for it poses different search problems. We speculate that experts are searching a highly composed model of disorders, not based on clear subtype and causal distinctions, but allowing for highly efficient search.

- Urgency, cost, the ability to treat a disease, and human values in general must be factored into the model explicitly.

Demonstrating the difficulty of this problem, the exclusions are more complex than what the model

includes. Of course, the aim of the work has been to develop a representation useful for teaching, not the most comprehensive model of diagnosis. It is premature to "flesh out" the model in all possible ways. However, gaps in the model require that we argue for its extensibility, particularly within the task/metarule/endcondition framework, which is the main product of this effort. Here the main considerations are both *psychological*, at the level of interrupting and restoring focus of attention and meta-level reasoning about an agenda of tasks, and *representational*, at the level of belief maintenance, the constructed model of the problem, and intersection-search procedures

5.5. Summary of evaluation

We have argued that evaluation of accuracy and completeness of the model should focus on the assumed constraints pertaining to knowledge structure, task requirements, human memory, and reasoning. Evaluation of performance and articulateness requires exercising the program in different, complex settings, including consultation, teaching, and learning. More specifically, we find ways in which the same knowledge must be used in multiple ways. We examine how a particular knowledge organization (e.g., subsumption) is used by different strategies and how a given strategy is applied in different contexts for a single case. Multiple cases enable us to vary the task, preventing us from tailoring strategies to particular cases, and revealing not only where the model falls short, but what properties of the task domain made the model appear adequate in other cases. Applying the model to other domains, such as computer software failure diagnosis, further reveals unprincipled or inadequately specified parts of the model (e.g., what is an etiological taxonomy?), and brings out assumptions about the task domain that are implicit in the model (e.g., the nature of the informant).

6. Conclusions

The driving force in NEOMYCIN's development has been to design a knowledge representation that can be used to model human diagnostic reasoning and explanation capability. The essential (and novel) aspect of the design is representation of the diagnostic procedure as abstract tasks that capture what structural effect the problem solver is trying to have on his evolving model of the problem. These tasks are invoked in a rule-like way that strongly emphasizes the problem solvers' use of relational knowledge about the domain for choosing his next move.

What is the nature of reasoning that such a model of expertise would work? First, there must be relatively more stereotypical situations (tasks and metarule conditions) than special case rules. It must be possible for problem solving to proceed step-by-step in a principled way (even if this would be unnecessary for the experienced problem solver), without encountering combinatorial problems. Second, it must be possible to richly structure knowledge about possible solutions and problem

features. These relations provide means for multiple, orthogonal hierarchical indexes that greatly facilitate search. Note that these constraints are general; they are what enables us to form *any* abstract model of strategy.

One purpose of NEOMYCIN has been to develop a language for representing abstract strategies. Follow-on work is concerned with using them in explanation (Hasling, 1984) and constructing a student model (London and Clancey, 1982). There are many advantages that can be useful in building any expert system (Clancey, 1983b) In our continuing development, we are slowly, but constantly, adding to the strategic model. We are still at the point where a carefully chosen case will reveal one or two important limitations in the model. In short, we are following an "enumeration methodology": Writing what we want to study in some language, organizing the collection to find *underlying themes, and further developing the language to express important distinctions.*

How applicable is the diagnostic procedure to other domains? The limitations described in Section 5.4 suggest that the model is far from complete. For example, electronic diagnosis often requires low-level causal analysis, working backwards from symptoms to component failures (Davis, 1983). However, at a higher, functional level, particularly for an expert who has debugged a particular device such as a given television or automobile model many times, we can expect that stereotypical matching as in infectious disease diagnosis will occur. *In this sense, NEOMYCIN's diagnostic* procedure will carry over to other domains. It should be viewed as a subset of a complete procedure, rather than as a specialized or over-simplified model.

What is the relation of NEOMYCIN to what the expert does? The model can be used to explain his behavior in the sense that it can generate it, but above the level of finding requests and hypotheses, the procedure is an abstraction, not steps he always consciously considers. In this sense, the diagnostic procedure is a *grammar* for parsing a series of information-gathering questions. By analogy with the grammar of natural language, it may reflect the innate nature of human reasoning, specifically how knowledge is remembered. Given that the procedure we have formalized operates entirely upon stereotypic knowledge of disorders, it can be characterized as *a procedure for searching classification knowledge*. Or since all knowledge may be in some sense compiled (e.g., encoded hierarchically as differences from patterns), the diagnostic procedure is analogous to Kolodner's "executive strategies" for remembering (Kolodner, 1983). However, the NEOMYCIN model pertains to the entire information-gathering procedure of diagnosis, not just a single probe of memory.

As a matter of practice, the diagnostic procedure has some of the same value to an expert that

knowledge of English grammar provides for a writer. Like English grammar some elements must be taught or at least enforced early on. The orientation towards "things to think about" is directly useful for teaching. Particularly, the idea of thinking in a hypothesis-directed way must be encouraged (but is this because students simply lack the automatic associations?). Perhaps the grammar or logic of diagnosis need not be conveyed explicitly, but certainly it is useful for a teacher of medicine to know it. How often have teachers criticized students, when they were following the procedure used by experts for coping with limited knowledge?

The idea of teaching students strategies or "how to think" has received considerable attention from AI researchers. Papert's work with LOGO (Papert, 1980) is perhaps the most well-known experiment in applying computational ideas to help problem solving in general. Our work raises interesting questions in this regard. For example, could someone familiar with our description of EXPLORE-AND-REFINE in terms of "looking up and looking down" and viewing diagnosis as a set-construction activity provide *better* explanations than those given by our expert-teacher? That is, having studied the constraints of the task more systematically than the expert, can we give students a better idea of what they should be trying to do?

A teacher using NEOMYCIN's model could go a step beyond Polya (Polya, 1957) and others (e.g., (Schoenfeld, 1981)) who have tried to teach *reasoning strategy to students*. *In contrast with other* research in teaching general strategies, we emphasize the role of domain relations ("structural knowledge") in selecting among different operators that affect the hypothesis space. From our perspective, Polya's heuristics might seem vague and unworkable (Newell, 1983) because:

1. They are not presented as parts of a comprehensive task structure or meta-strategy (as pointed out by Schoenfeld).

2. They lack a premise part that refers to working memory, the situation in which the problem solver will find them to be useful for something he is trying to do; that is, they are not stated as conditional operators.

3. The way in which they index particular mathematical solution methods is not clearly worked out; that is, the domain relation vocabulary is missing.

NEOMYCIN's relational vocabulary consists of causal, subtype, and process relations that classify and link findings and hypotheses. Some of the specific terms considered in this paper are: finding, soft-finding, red-flag finding, substance, and process location. These terms are like parts of speech and syntactic units that classify and organize the problem-solver's domain lexicon. This is *knowledge for organizing knowledge*: a means for expressing and using knowledge. A diagnostic strategy says

in effect, "To accomplish a certain task, think about some finding (or hypothesis) that is related to your current hypotheses (or known findings) by the X relation." "To refine a hypothesis, consider *common causes*. What are the common causes of a sore throat?" As a self-directive, this is an example of meta-cognition. Strategies orient the problem solver towards constructing and refining an appropriate problem space. They constitute the *managerial knowledge* by which the problem solver directs his attention and so brings his expertise to bear on the problem. Having gone beyond MYCIN's single-layer, "quick association" model of thinking (as Schoenfeld has characterized traditional expert systems), we are poised to experiment with teaching strategic reasoning.

Indeed, we have now entered a strange sort of loop in our research. We are teaching the diagnostic strategy to research assistants to make them better computer program debuggers. (The general question, "Has the patient undergone surgery?" becomes "Has this program been edited since it last worked?") This experience suggests ways to generalize the model, helps us to develop ways to teach it, and may enable us to implement the teaching program itself more efficiently. And so again we find ourselves amid the complex web of learning, teaching, and problem solving.

I. Basic terminology of diagnosis

- DIAGNOSTIC PROBLEM: A situation in which a device exhibits behavior (findings) that suggest that it is malfunctioning. A diagnostic problem has a "cause" that, for our purposes, is one of a set of known processes (hypotheses). Example: A severe headache for a week and double vision in a patient is a diagnostic problem.

- FINDING: An observable problem feature, generally characterizing the problem in a very narrow, non-explanatory way. In medicine, these are signs, symptoms and laboratory data. Example: A headache is a finding.

- HYPOTHESIS: An interpretation of findings in terms of underlying substances and processes that produce them. A hypothesis can be said to "explain" the findings. Example: "Space-occupying substance in the brain" is a hypothesis.

- DIFFERENTIAL: The most specific set of hypotheses that the problem solver is considering. By the "single-fault assumption" these hypotheses are mutually exclusive and therefore competing. Example: A typical differential might be brain-abscess and *chronic-meningitis*.

- DOMAIN KNOWLEDGE: Findings, hypotheses, and relations among them that enable inferences to be drawn about their applicability. Example: Medications "subsumes" antibiotics, analgesics, and steroids. Example: An "evidence relation" links a finding to a

hypothesis that causes or might be caused by it, as viral meningitis is caused by exposure to the disease.

- TASK: What the problem-solver is trying to do with respect to findings, hypotheses, and his domain knowledge. A task is accomplished by a procedure of ordered conditional actions, called metarules. We say that the metarules "achieve" the task. For example, the metarules of the task PURSUE-HYPOTHESIS test and refine a given hypothesis. Primitive tasks are to request information about a finding and to make an inference about a finding or hypothesis.

- FOCUS: The finding, hypothesis, or the differential that is the argument to a task, for example, the hypothesis that the problem solver is trying to test.

- METARULE: A conditional statement that partially accomplishes a task by invoking subtasks. For example, "If the task is to establish the space of hypotheses relevant to this problem and the differential has been reduced and refined, then ask general questions." Metarules are either conditional steps in a procedure or preferentially ordered alternative methods for accomplishing a task.

- CONSTRAINT: Some condition that the problem solver must try to satisfy, such as to solve the diagnostic problem in the shortest amount of time, or some limitation or capability of his ability to reason that he must cope with, such as his ability to remember the extent of his knowledge or the differential.

II. Detailed analysis of a protocol

In the protocol that follows, annotations indicate the NEOMYCIN tasks that would generate the finding requests and hypothesis assertions made by the expert.[14] Numbers in parentheses refer to numbered statements that support the interpretation. *Annotations* precede the expert behavior they are intended to explain. This analysis illustrates the knowledge acquisition technique, the nature of the diagnostic problem, and the model's representation in terms of tasks, focus, and domain relations. Note that the metarules that cause the tasks to be invoked are not indicated here; they are listed in Appendix IV. Figure II-1 shows a parse tree of the physician's five data requests, which appear underlined in the protocol. By comparison with Figure 3-2, you can see that this protocol illustrates the central part of the diagnostic procedure, but not most of the tasks.

1 KE: What I wanted to do different in these cases is to pick cases where I

---

[14] While we have a prototype modeling program that can generate similar annotations, they are still not nearly as good as what we can do by hand. In the interest of making NEOMYCIN's model as comprehensible as possible, it seems best to show here the best interpretations we can supply.

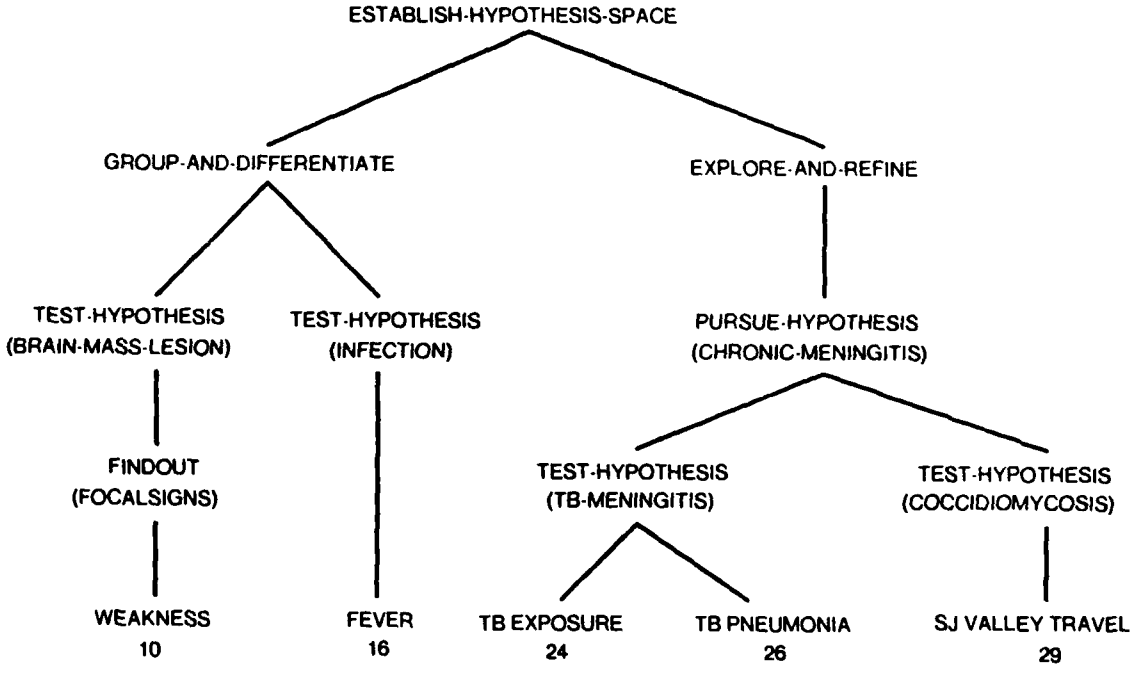


Figure II-1: Parse with respect to the diagnostic model of the five questions asked in the protocol

thought you might have to request more information than what I gave originally so we can look at a little bit of that process. In these cases especially, you can be as complete as possible in telling me what you are thinking.

2 MD: So you just want to give me skeleton data?

3 KE: Yes, we'll see how it goes. I am going to try to follow the general principle we had established, which was to tell you why the person was in the hospital and how they got to the point where the lumbar puncture was done.

4 First example: A 15-year old female. A two-week history of headache, nausea, vomiting; and diplopia one day prior to admission.

task = IDENTIFY-PROBLEM
task = FORWARD-REASON (headache, nausea, vomiting, diplopia, headache-duration, nausea-duration, vomiting-duration, diplopia-duration)
structural knowledge: diplopia is a serious (red flag) CNS finding
task = PROCESS-FINDING (diplopia)
task = APPLY-ANTECEDENT-RULES (causes of diplopia)
evidence rule: diplopia caused-by increased-pressure-in-brain (6)
task = PROCESS-FINDING (diplopia-duration)
task = APPLY-ANTECEDENT-RULES (mentioning diplopia-duration)
definition: max(duration of CNS findings) = CNS-problem-duration (5)

5 MD: (I think this would be a very good case to illustrate whether you should do a lumbar puncture or not.) This is somebody who has evidence of perhaps a pressure build-up in the brain for a two week period of time.

[Causal explanation: how pressure build-up causes diplopia]

6 The diplopia comes because as the pressure builds up in the brain, you can't focus your eyes properly. It is a very sensitive indicator. One of the nerves that enervates the movement of the eyes together is the first one that is impaired as the pressure builds up,

task: PROCESS-HYPOTHESIS (increased-pressure-in-brain) (7)

7 so that I would be concerned in this situation of increased pressure in the brain

task: APPLY-ANTECEDENT-RULES (causes of increased-pressure-in-brain)
evidence rule: increased-pressure-in-brain -> brain-mass-lesion
task: PROCESS-HYPOTHESIS (brain-mass-lesion) (8)
add differential: brain-mass-lesion
task: PURSUE-HYPOTHESIS (brain-mass-lesion)
task: REFINE-HYPOTHESIS (brain-mass-lesion)
structural knowledge: brain-mass-lesion subsumes brain-tumor, hematoma and collection of pus.

8 and worry about tumor--a mass lesion of some type: a collection of blood, a collection of pus.

task: PROCESS-FINDING (serious-CNS-finding)
task: APPLY-ANTECEDENT-RULES (serious-CNS-finding)
evidence rule: serious CNS-finding -> meningitis (9)
task: PROCESS-HYPOTHESIS (meningitis)
add differential: meningitis
task: APPLY-EVIDENCE-RULES (known findings activated by meningitis)
evidence rule: CNS-problem-duration -> chronic-meningitis (9, 22)
replace differential: meningitis -> chronic-meningitis

9 **If it is a meningitis it is clearly a chronic one because we are talking about a two week history.**

task: GROUP-AND-DIFFERENTIATE (brain-mass-lesion, chronic-meningitis)
structural knowledge: brain-mass-lesion is a focal process: (12)
chronic-meningitis is a systemic process.
task: FINDOUT (focal-manifestations) (13)
structural knowledge: focal-manifestations subsumes diplopia (13)
structural knowledge: focal-manifestations subsumes weakness (14)
task: FINDOUT (weakness)

10 **The next historical question that I would want to know: <u>Does she have any weakness anywhere in her body? One side weaker than the other?</u>**

11 KE: **Why do you ask that?**

12 MD: **Since this picture is very suggestive of a focal lesion in the brain,**

13 **I am wondering if there are any focal manifestations other than double vision,**

[Causal explanation: that brain problem affects body extremity]
[Structural knowledge: focal neurological findings subsumes
one-sided hand-weakness and leg-weakness]

14 **e.g. "My hand right has been very weak" and I would wonder if there is something happening in the brain which enervates the right hand. Or, has she been having trouble walking, with one leg being weaker than the other, or is her balance off. Those are what are called focal neurological findings.**

15 KE: **Okay. Focal signs in general... unknown.**

task: GROUP AND DIFFERENTIATE (brain-mass-lesion,
chronic meningitis) (18)
structural knowledge: chronic meningitis is an infection
task: TEST HYPOTHESIS (infection) (18)
evidence rule: fever -> infection (21)
task: FINDOUT (fever)

16 MD: **<u>Has she had</u> fevers?**

17 KE: **Unknown.**

18 **I think that is an important question to help distinguish between an infectious cause versus a non-infectious cause.**

[Structural knowledge: blood clot = hematoma and brain tumor are not infectious causes]

19 **A non-infectious cause being a blood clot or brain tumor.**

20 KE: **So the fact that if there weren't a fever, that would suggest...?**

21 MD: **Not having a fever does not necessarily rule out an infection. But if she had an fever, it would be more suggestive of it.**

22 **The situation we are dealing with is a chronic process.**

task: TEST-HYPOTHESIS (chronic-infection)
evidence rule: low grade fever -> chronic-infection (23)

23 **Sometimes with chronic infections fever can be low or none at all.**

task: PURSUE-HYPOTHESIS (chronic-meningitis)
task: REFINE-HYPOTHESIS (chronic-meningitis)
structural knowledge: chronic-meningitis subsumes TB-meningitis, fungal-meningitis, and partially-rx-bacterial-meningitis (33)
add differential: TB-meningitis, fungal-meningitis, and partially-rx-bacterial-meningitis
task: EXPLORE-AND-REFINE (TB-meningitis, fungal-meningitis, and partially-rx-bacterial-meningitis)
task: PURSUE-HYPOTHESIS (TB-meningitis)
task: TEST-HYPOTHESIS (TB-meningitis)
evidence rule: tuberculosis-exposure -> TB-meningitis
task: FINDOUT (tuberculosis-exposure)

24 **Has she had any exposure to tuberculosis?**

25 KE: **No. No TB risk.**

task: PROCESS-FINDING (negative TB-risk)
task: FINDOUT (TB-risk)
structural knowledge: TB-risk subsumes tuberculosis-pneumonia
task: FINDOUT (tuberculosis-pneumonia)
structural knowledge: pneumonia subsumes tuberculosis-pneumonia (26)
task: FINDOUT (pneumonia)

26 MD: **No recent pneumonia that she knows of? Tuberculosis-pneumonia?**

27 KE: **Let me see how complete "TB risks" is. According to MYCIN, they include one or more of the following: Positive intermediate trans-PPD; history of close contact with person with active TB; household member with past history of active TB; atypical scarring on chest x-ray; history of granulomas on biopsy of liver, lymph nodes or other organs.**

task: FORWARD-REASON
(+ PPD, contact-TB, family-TB, X-ray-TB, granulomas)
structural knowledge: TB-risk subsumes
+ PPD, contact-TB, family-TB, X-ray-TB, granulomas

28 MD: That's pretty solid evidence against a history of TB.

task: EXPLORE-AND-REFINE (fungal-meningitis and partially-rx-bacterial-meningitis)
task: PURSUE-HYPOTHESIS (fungal-meningitis)
task: REFINE-HYPOTHESIS (fungal-meningitis)
structural knowledge: likely fungal-meningitis causes are coccidiomycosis and histoplasmosis (33)
add differential: coccidiomycosis and histoplasmosis
task: PURSUE-HYPOTHESIS (Coccidiomycosis)
task: TEST-HYPOTHESIS (Coccidiomycosis)
evidence rule: San-Joaquin-Valley-travel -> Coccidiomycosis
task: FINDOUT (San-Joaquin-Valley-travel)
structural knowledge: travel subsumes San-Joaquin-Valley-travel (29)
task: FINDOUT (travel)

29 <u>Has she traveled anywhere? Has she been through the Central Valley of California?</u>

30 KE: You asked TB risks because?

31 MD: I asked TB risks because we are dealing here with an indolent (chronic) infection since we have a two week history.

32 I am thinking, even before I have any laboratory data,

33 of infections, chronic infections are most likely. So I'll ask a few questions about TB, cocci, histo and other fungal infections.

34 KE: Histo is a fungal infection?

[structural knowledge: histo location is Midwest]
[structural knowledge: cocci location is Arizona and California]

35 Histoplasmosis is a fungus infection of the Midwest. Cocci is the infection of Arizona and California.

36 KE: So you are focusing now on chronic infections. Why would you look at the history now before doing anything else?

37 MD: I am trying to approach it as a clinician would. Which would be mostly to get a lot of the historical information and do a physical exam, then do a laboratory.

38 A lot of times, people think from the laboratory, whereas I think you should think for the laboratory. People are talking more about that now, especially because the cost of tests are an issue. You can get a lot from just talking with the patient. I could ask for the LP

results, then go back and ask questions. But without knowing the LP results, which would bias me in the way I am going to ask the questions.

39 KE: This helps you...

40 MD: This is the way you approach a patient.

III. Expert-teacher statements of diagnostic strategy

We summarize here the general principles of the model, with excerpts from expert problem-solving and classroom protocols. The tasks of the model are a set of directives for changing focus, testing hypotheses, and gathering information. Note the expert-teacher's method of combining abstract and concrete explanations.

- ESTABLISH-HYPOTHESIS-SPACE -- Establish the breadth of possibilities, then focus.

TEACHER: ... All the cases we have had have fit pretty nicely into trying to establish a breadth of possibilities and then focusing down on the differential within one of the categories.

- GROUP-AND-DIFFERENTIATE -- Ask yourself, "What are the general processes that could be causing this?"

TEACHER: Do you have in mind certain types of sore throats that ... ? Because the types of questions that you ask early on, once you have a sense of the problem, would be to ask a couple of general questions maybe that could lead you into other areas to follow up on, rather than zeroing in.

STUDENT:
Ok.

TEACHER: I was asking that because I think it's important to try to be as economical as possible with the questions so that each question helps you to decide one way or the other. At least with sore throat and my conception of sore throat, I have a hard time thinking of how different types of pain and different types of relief pattern are going to mean different etiologies to the sore throat....

TEACHER (later): Ok, so we think about infectious, but what other things might be running through your mind in terms of broadening out again? We've got a new set of findings now besides fever and sore throat we have...

- EXPLORE-AND-REFINE -- Scan the possibilities and choose one to explore in more detail.

TEACHER: Anything else? Well there are probably a couple of other areas to think about, ... you know, like auto-immune diseases, inflammation of the throat... Why don't we get back to infections now, because we have a story of fever and sore throat, that is a common problem with infectious diseases. So we're talking about strep throat, we're talking about upper-respiratory, viral... Any other type of infectious problem... ?

STUDENT:
... Pneumococcus would give you sore throat too, right?

TEACHER: Pretty rarely.

TEACHER (different case): Well, how about some questions about mononucleosis now. I'd have you zero in on that.

- FORWARD-REASON -- Ask yourself, "What could cause that?" Look for associated symptoms.

TEACHER: Well what's another possibility to think about in terms of weakness? What do a lot of older people think of when they just think of being weak, a common American complaint. Or a common American understanding of weakness. How about tired blood?

STUDENT:
Iron deficiency.

TEACHER: I think of anemias.

TEACHER (different case): Most important is to develop a sense of being reasonably organized in approaching the information base and trying to keep a complete sense of not homing in too quickly. Look for things to grab onto, especially if you have a nonspecific symptom like headache, weakness. Ten million people in the country probably have a headache at this given point in time. What are the serious ones, and what are the benign ones? Look for associated symptoms. Some associated symptoms definitely point to something severe, while others might not.

- REFINE-HYPOTHESIS -- Ask yourself, "What are the common causes and the serious, but treatable causes?"

TEACHER: What anemias do young people get?

TEACHER (different case): What diseases can wind up in congestive heart failure? Congestive heart failure is not a diagnosis, it's kind of an end-stage physiology and there are lots of diseases that lead into congestive heart failure; lots of processes, one is hypertensive. What's the other most common one? There are two that are common in this country. One is hypertensive, what's the other most common one?

STUDENT:
Atherosclerosis?

- TEST-HYPOTHESIS -- Ask yourself, "How can I check this hypothesis?"

TEACHER: How can you check whether someone is anemic? What question might you ask?

- ASK-GENERAL-QUESTIONS -- Ask general questions that might change your thinking.

TEACHER: Well that's an important question I think. Sometimes you can ask it very generally, like, "Is there anything... have you had any major medical problems or are you on any medications?" Then people will come back and tell you. And that's an important issue to establish, whether somebody is a compromised host or a normal host because a normal host... Then you have a sense of what the epidemiology of diseases in a normal host... When you talk about compromised host, you're talking about everything changing around, and you have to consider a much broader spectrum, different diagnoses. So, you might ask that question more specifically, you know, "are you taking any medications or do you have any other medical problems, like asthma," or some times they're taking steroids. Those types of general questions are important to ask early on, because they really tell you how soon you can focus down.

STUDENT:
Are you on any medication right now?

- GENERATE-QUESTIONS -- Try to get some information that suggests hypotheses.

TEACHER: You're jumping around general questions and I think that's useful. I don't know where to go at this point. So this is the appropriate time for a kind of a "buckshot" approach ... every direction till we latch onto something that we can follow up, because right now we just have a very non-specific symptom.

IV. The Diagnostic Procedure

This section describes in detail the content of NEOMYCIN's metarules. The tasks are listed in depth-first calling order, assuming that they are always applicable (refer to Figure 3-2). For each substantial task (FORWARD-REASON, FINDOUT, ESTABLISH-HYPOTHESIS-SPACE and its subtasks), we attempt to list exhaustively all of the implicit assumptions about task and cognitive constraints proceduralized by the metarules. These are an essential part of the model. The model is constantly changing; this is a snapshot as of July 1985. To give an idea of how the program is evolving, metarules now on paper are listed as "<<proposed>>."

IV.1. CONSULT

This is the top level task. A single metarule unconditionally invokes MAKE-DIAGNOSIS and then prints the results of the consultation. (We have disabled MYCIN's therapy routine because the antibiotic information was out of date; it would be invoked here.)

IV.2. MAKE-DIAGNOSIS

A single unconditional metarule invokes the following tasks: IDENTIFY-PROBLEM, REVIEW-DIFFERENTIAL, and COLLECT-INFORMATION. REVIEW-DIFFERENTIAL simply prints out the differential, modeling a physician's periodic restatement of the possibilities he is considering. (In a teaching system, this would be an opportunity to question the student.) Hypothesis-directed reasoning is done by COLLECT-INFORMATION.

IV.3. IDENTIFY-PROBLEM

The purpose of this task is to gather initial information about the case from the informant, particularly to come up with a set of initial hypotheses.

1. The first metarule unconditionally requests "identifying information" (in medicine, the name, age, and sex of the patient) and the "chief complaint" (what abnormal behavior suggests that there is an underlying problem requiring therapy). The task FORWARD-REASON is then invoked.

2. If no diagnoses have been triggered (the differential is empty), the task GENERATE-QUESTIONS is invoked.

IV.4. FORWARD-REASON

The metarules for FORWARD-REASON iterate over the list of new conclusions, first invoking CLARIFY-FINDING for each finding and then PROCESS-FINDING for each serious or "red-flag" finding. PROCESS-FINDING is then invoked for non-specific findings and PROCESS-HYPOTHESIS for each hypothesis. These tasks perform all of the program's forward reasoning.

It is important to "clarify" findings, that is, to make sure that they are well-specified, before doing any forward reasoning. Thus, before considering that the patient has a fever, we first ask what his temperature is. "Red-flag" in contrast with "nonspecific" findings often trigger hypotheses; they are serious, indicative of a real problem to be treated and not just a "functional" imperfection in the

device[15] ; nonspecific findings may very well be explained by the hypotheses that red-flag findings quickly suggest. These considerations are all matters of cognitive economy, means to avoid backtracking and to make a diagnosis with the least search.

IV.5. CLARIFY-FINDING

Using subsumption and process relations among findings, these metarules seek more specific information about a finding, asking two types of questions:

1. Specification questions (e.g., if the finding is "medications," program will ask what drugs the patient is receiving).

2. Process questions (e.g., if the finding is "headache", the program will ask when the headache began).

IV.6. PROCESS-FINDING

The metarules for this task apply the following kinds of domain rules and relations in a forward-directed way:

1. Antecedent rules (causal and definitional rules that use the finding and can be applied now).

2. Generalization (subsumption) relations (e.g., if the finding is "neurosurgery," the program will conclude that "the patient has undergone surgery").

3. Trigger rules (rules that suggest hypotheses; the program will pursue subgoals if necessary to apply these rules). If a nonspecific finding is explained by hypotheses already in the differential, it does not trigger new hypotheses.

4. Ordinary consequent rules that use soft findings to conclude about activated hypotheses (those hypotheses on the differential, plus any ancestor or immediate descendent); no

---

[15] In medicine, a headache usually indicates a functional, as opposed to an "organic," disorder. By analogy, a high load-average in a time-sharing computer often indicates a functional disorder, just a problem of ordinary "life." Though, like a headache, it may signify a serious underlying disorder.

subgoaling is allowed.[16]

5. Ordinary consequent rules that use hard findings, as above, but subgoaling is allowed.

6. (<<Proposed>> Rule out considered hypotheses that do not account for a new red-flag finding.)

7. (<<Proposed>> Refine current hypotheses that can be discriminated into subtypes on the basis of the new finding.[17])

These metarules (and their ordering) conform to the following implicit constraints:

- The associations that will be considered first are those requiring the least additional effort to realize them.

  Effort in forward reasoning, an aspect of what has also been called *cognitive economy*, can be characterized in terms of:

  - *immediacy* (the conclusion need only be stated vs. subgoals must be pursued or the problem solver must perform many intersections of the differential, related hypotheses, and known findings)

  - *relevance* (make conclusions focused with respect to current findings and hypotheses vs. take actions that might broaden the possibilities, require "unrelated" findings, and change the focus).

- The metarules are directed at efficiency by:

  - Drawing inferences in a data-directed way, rather than doing a search when the conclusions are needed. The primary assumption here is that the structure of the problem space makes forward reasoning more efficient.

---

[16] Should the concept of a trigger rule be generalized to allow specification of any arbitrary context? In particular, is the idea of applying rules relevant to children of active hypotheses just a weak form of trigger rule? Perhaps the the "strength" of an association corresponds to the *extent of the context* in which it will come to mind. Trigger rules are simply rules which apply to the entire domain of medical diagnosis. We might associate rules with intermediate contexts as well, for example, "infectious disease diagnosis."

Resolving this issue may make moot the issue of whether trigger rules should be placed before ordinary consequent rules. Their relevance is more directly ascertained, applying consequent rules in a focused, forward way requires intersection of the new finding with specific hypotheses on the differential and their descendents. Trigger rules also have the payoff of indicating new hypotheses. However, if applying a trigger rule requires gathering new findings and then changing the differential, some cost is incurred in returning to consider the ordinary consequent rules afterwards.

[17] This would again promote refocusing, and thus the cost of losing the current context. An agenda model could explain ability to realize these new associations and come back to them later.

- Drawing all possible focused inferences (each metarule is tried once, but executes all inferences of its type) and refining findings to a useful level of detail by asking more questions (not hypothesis-directed).

In summary, the order of forward reasoning is based on cognitive issues, not correctness.

IV.7. PROCESS-HYPOTHESIS

These rules maintain the differential and do forward reasoning.

1. If the belief in the hypothesis is now less than .2, and it is in the differential, it is removed.

2. If the hypothesis is not in the differential and the belief is now greater than or equal to .2, it is added to the differential. The task APPLY-EVIDENCE-RULES is invoked. This task applies rules that support the hypothesis, using previously given findings (the hypothesis might not have been active when the data was processed). Only rules that succeed without setting up new subgoals are considered.

3. (<<Proposed>> If the belief is very high (greater than .8) and the program knows of no evidence that could lower its belief, then the hypothesis is marked as explored, equivalent to completing TEST-HYPOTHESIS.)

4. (<<Proposed>> Apply ordinary consequent rules that use soft findings to conclude about new activated hypotheses.)

5. If the hypothesis has been explored (either because of the previous rule or the task TEST-HYPOTHESIS is complete), then generalization (subsumption) relations and antecedent rules are applied.

Adding a hypothesis to the differential is bookkeeping performed by a LISP function. While NEOMYCIN's differential is a list, it cannot really be separated conceptually from the hierarchical and causal structures that relate hypotheses. The hypothesis is not added if a descendent (causal or subtype) is already in the list. If an ancestor is in the list, it is deleted. If there is no previous ancestor or descendent, the program records that the differential is now "wider"--an event that will effect aborting and triggering of tasks. Thus, the differential is a memory-jogging "cut" through causal and subtype hierarchies.

The ordering of PROCESS-HYPOTHESIS metarules is cognitively based, as for PROCESS-FINDING, but follows a more logical procedural ordering: bookkeeping of the differential, recognition of more evidence, completion of consideration, and drawing more conclusions. The orderliness of

this procedure again reflects the cognitive (and computational) efficiency of locally realizing and recording known information before drawing more conclusions (i.e., returning to the more general search problem).

IV.8. FINDOUT

This task models how the problem solver makes a conclusion about a finding that he wants to know about. (This is a greatly expanded and now explicit version of the original MYCIN routine by the same name (Shortliffe, 1976).) The rules are applied in order until one succeeds.

1. If the finding concerns complex objects (such as cultures, organisms or drugs) then a special Lisp routine is invoked to provide a convenient interface for gathering this information.

2. If the finding is a laboratory test whose source is not available or whose availability is unknown, then the finding is marked as unavailable. (E.g., if it is not known whether the patient had a chest x-ray, nothing can be concluded about what was seen on the chest x-ray.)

3. If the finding is subsumed by any more general finding that is ruled out for this case, then the finding is ruled out also. (E.g., if the patient has not received medications, then he has not received antibiotics.)

4. As a variant on the above rule, if any more general finding can be ruled out that has not been considered before, then the finding can be ruled out.[18]

5. If any more general finding is unknown, then this specific finding is marked as unavailable.

6. If some more specific finding is known to be present, then this finding can be concluded to be present, too. (E.g., if the patient is receiving steroids, then the patient is receiving medications.)

7. If the finding is normally requested from the informant, but shouldn't be asked for this kind of problem, then try to infer the finding from other information.[19]

---

[18] That is, the premise of this metarule invokes FINDOUT recursively. To do this cleanly, we should allow tasks to return "success" or "fail."

[19] "Inferring" means to use backward chaining. Given that source and subsumption relations have already been considered at this point, only definitional rules remain to be considered. That a finding should not be asked is determined by the "don't ask when" relation, requiring the task APPLYRULES to be invoked in the premise of this metarule.

8. If the "finding" is really a disorder hypothesis (we are applying a rule that requires this information), then invoke TEST-HYPOTHESIS (rather than backward chaining through the domain rules in a blind way).

9. If the informant typically expects to be asked about this finding, then request the information, then try to infer it, if necessary.

10. Otherwise, try to infer the finding, then request it.

The constraints that lie behind these rules are:

- Economy: use available information rather than drawing intermediate inference or gathering more information. Keep the number of inferences and requests for data to a minimum. Solve the problem as quickly as possible.

- First requesting more general information attempts to satisfy the economy constraint, but assumes that more than one specific finding in the class will eventually be considered and that the general finding is often negative. Otherwise, the general question would be unnecessary.

- It is assumed that the informant knows and consistently uses the subsumption relations used by the problem solver, so the problem solver is entitled to rule out specific findings on the basis of general categories. For example, knowing that the patient is pregnant, the informant will not say that she is not a compromised host. General questions help ensure completeness. When a more general question is asked, a different specific finding than the one originally of interest could be volunteered. Later forward reasoning could then bring about refocusing.

- Typical of the possible interactions of domain knowledge that must be considered, a finding with a source must not be subsumed by ruled-out findings, otherwise considering the source would be unnecessary, and doing it first would lead to an extra question. Obviously, if there are too many interactions of this sort, the strategic "principles" will be very complex and slow to apply in interpreted form.

Note that we could have added another metarule to rule out a general class if all of its more-specific findings have been ruled out, but the "closed-world assumption" does not make sense with NEOMYCIN's small knowledge base.

IV.9. APPLYRULES

NEOMYCIN has "internal" tasks that control how domain rules are applied: "only if immediate" (antecedent), "with previewing" (looking for a conjunct known to be false), and "with subgoaling." An important aspect of NEOMYCIN as a cognitive model is that new findings, coming from rule invocation, are considered in a depth-first way. That is, the conclusions from new findings are considered before returning to information gathered earlier in the consultation. Implementing this requires "rebinding" the list of new findings (so a "stack" is associated with rule invocations) and marking new findings as "known" if no further reasoning could change what is known about them, thus adding them to the list of findings to be considered in forward reasoning. The basic assumptions are that the informant does not retract findings, that the problem-solver does not retract conclusions, and FORWARD-REASON is done for each new finding.

IV.10. GENERATE-QUESTIONS

This task models the problem solver's attempt to milk the informant for information that will suggest some hypotheses. The program generates one question at a time, stopping when the differential is "adequate" (the end condition of the task). The differential is adequate in the early stage of the consultation if it is not empty, otherwise the belief in some considered hypothesis must be "moderate" (defined as a cumulative CF of .3 or greater, the measure used consistently in domain rules to signify "reasonable evidence").

The metarules generate questions from several sources, invoking auxiliary tasks to pursue different lines of questioning:

1. General questions (ASK-GENERAL-QUESTIONS)

2. Elaboration of previously received data (ELABORATE-DATUM). (The subtask ELABORATE-DATUM asks about subsumed data. For example, if it is known that the patient is immunosuppressed, the program will ask whether the patient is receiving cytotoxic drugs, is an alcoholic, etc. The subtask also requests more "process information." For example, it will ask how a headache has changed over time, its severity, etc.)

3. Any rule using previous data that was not applied before because it required new subgoals to be pursued is now applied.

4. The informant is simply asked to supply more information, if possible.

This task illustrates the importance of record-keeping during the consultation. These metarules

refer to which tasks have been previously completed, which findings have been fully specified and elaborated, and hypothesis relations that have been considered.

IV.11. ASK-GENERAL-QUESTIONS

These questions are the most general indications of abnormal behavior or previously diagnosed disorders, useful for determining if this is a "typical" case that is what it appears to be, or an "unusual" problem, as described in Section 3. These are of course domain-specific questions. They generalize to: Has this problem ever occurred before? What previous diagnoses and treatments have been applied to this device? When was the device last working properly? Are there similar findings manifested in another part of the device? Are there associated findings (occurring at the same time)? These questions are asked in a fixed order, consistent with the case-independent, "something you do every time," nature of this task.

IV.12. COLLECT-INFORMATION

These rules carry out the main portion of data collection for diagnosis; they are applied iteratively, in sequence, until no rule succeeds.

1. If there are hypotheses appearing on the differential that the program has not yet considered actively, then the differential is reconsidered (ESTABLISH-HYPOTHESIS-SPACE) and reviewed (REVIEW-DIFFERENTIAL).[20] If the differential is not "adequate" (maximum CF below .3), an attempt is made to generate more hypotheses (GENERATE-QUESTIONS).

2. If the hypotheses on the differential have all been actively explored (ESTABLISH-HYPOTHESIS-SPACE completed), then laboratory data is requested (PROCESS-HARD-DATA).

---

[20] To avoid recomputation, the function for modifying the differential sets a flag when new hypotheses are added. It is reset each time the task ESTABLISH-HYPOTHESIS-SPACE completes. Generally, the goal of each task (e.g., GENERAL-QUESTIONS-ASKED) is used for history keeping, but tasks like ESTABLISH-HYPOTHESIS-SPACE are invoked conditionally, multiple times during a consultation, as the program loops through the COLLECT-INFORMATION metarules. The use of flags brings up questions about the mind's "register" or "stack" capabilities, whether NEOMYCIN should use an agenda, and so on. In our breadth-first approach to constructing a model, we hold questions like this aside until they become relevant to our performance goals.

IV.13. ESTABLISH-HYPOTHESIS-SPACE

This task iterates among three ordered metarules:

1. If there are ancestors of hypotheses on the differential that haven't been *explored* by TEST-HYPOTHESIS, then these are considered (GROUP-AND-DIFFERENTIATE). (For computational efficiency, the records *parents-explored* and *descendents-explored* are maintained for each hypothesis.)

2. If there are hypotheses on the differential that haven't been *pursued* by PURSUE-HYPOTHESIS, then these are considered (EXPLORE-AND-REFINE).

3. If all general questions have not been asked, invoke ASK-GENERAL-QUESTIONS.

The constraints satisfied by this task are:

- All hypotheses that are placed on the differential are tested and refined (based on correctness).

- Causal and subtype ancestors are considered before more specific hypotheses (based on efficiency and assuming that the best model for explaining findings is a known stereotype disorder, and these stereotypes can be taxonomically organized).

IV.14. GROUP-AND-DIFFERENTIATE

This task attempts to establish the disorder categories that should be explored

1. If all hypotheses on the differential belong to a single top-level category of disease (appear in one subtree whose root is at the first level of the taxonomy), then this category is tested. Such a differential is called "compact"; the concept and strategy comes from (Rubin, 1975).

2. If two hypotheses on the differential differ according to some process feature (location, time course, spread), then ask a question that discriminates on that basis. (This is the metarule that uses orthogonal indexing to group and then discriminate disorders.)

3. If there is some hypothesis whose top-level category has not been tested, then test that category. (E.g., consider infectious-process when there is evidence for chronic-meningitis.)

The first metarule is not strictly needed since its operation is covered by the third metarule. However, we observed that physicians remarked on the presence of an overlap and pursued the single category first, so we included this metarule in the model.

The second metarule uses process knowledge to compare diseases, as described in Section 3.

To summarize the constraints behind the metarules:

- When examining hypotheses, intersection at the highest level is noticed first. The etiological taxonomy is assumed to be a strict tree.

- Use of process knowledge requires two levels of reasoning: mapping over all descriptors and intersecting disorders based on each descriptor. This is more complicated that a subtype intersection, requiring more effort, so it is done after testing the differential for compactness. For this maneuver to be useful, disorders must share a set of process descriptors.

- Because a stereotype disorder inherits features of all etiological ancestors, these ancestors must be considered as part of the process of confirming the disorder (a matter of correctness). This assumes that knowledge of disorders has been generalized and "moved up" the tree (perhaps an inherent property of learning, the effect is beneficial for search efficiency). Furthermore, circumstantial evidence that specifically confirms a disorder can only be applied if ancestors are confirmed or not ruled out. That is, circumstantial associations are context-sensitive.

IV.15. TEST-HYPOTHESIS

This is the task for directly confirming a hypothesis. The following methods are applied in a pure-production system manner:

1. Preference is first given to findings that trigger the hypothesis.

2. Next, causal precursors to the disease are considered. (For infectious diseases, causal precursors include exposure to the disease and immunosuppression.)

3. Finally, all other evidence is considered.

Each metarule selects the domain rules that mention the selected finding in their premise and conclude about the hypothesis being tested. The MYCIN domain rule interpreter is then invoked to apply these rules (in the task APPLYRULES). (So applying the rule will indirectly cause the program to request the datum.) After the rules are applied, forward reasoning using the findings and new hypothesis conclusions is performed (FORWARD-REASON).

<<Proposed>>: The task aborts if belief is high (CF greater than .8) and no further questioning can make the belief negative. The task also aborts if there is no belief in the hypothesis and only weak

evidence (CF less than .3) remains to be considered after several questions have been asked.

Relevant constraints are:

- Findings bearing a strong relation with the hypothesis are considered first because they will contribute the most weight (a matter of efficiency).

- Disconfirming a hypothesis involves discovering that required or highly probable findings--causal precursors or effects--are missing. NEOMYCIN's domain lacks this kind of certainty. Therefore, the program does not use a "ruleout" strategy.

- The end conditions attempt to minimize the number of questions and shift attention when belief is not likely to change (a matter of efficiency).

IV.16. EXPLORE-AND-REFINE

This is the central task for choosing a focus hypothesis from the differential. The following metarules are applied in the manner of a pure production system.

1. If the current focus (perhaps from GROUP-AND-DIFFERENTIATE) is now less likely than another hypothesis on the differential, then the program pursues the stronger candidate (PURSUE-HYPOTHESIS).

2. If there is a child of the current focus that has not been pursued, then it is pursued (this can only be true after the current focus has just been refined and removed from the differential).

3. If there is a sibling of the current focus that has not been pursued, then it is pursued.

4. If there is any other hypothesis on the differential that has not been pursued, then it is pursued.

This task is aborted if the differential becomes wider (see PROCESS-HYPOTHESIS), a precondition that requires doing the task GROUP-AND-DIFFERENTIATE.

Relevant constraints are:

- All selection of hypotheses is biased by the current belief (a matter of efficiency).

- Focus should change as soon as the focus is no longer the most strongly believed hypothesis (a matter of correctness; perhaps at odds with minimizing effort, due to the cost of returning to this focus).

- Siblings are preferred before other hypotheses (a matter of cognitive effort to remain focused within a class; also a matter of efficiency, in so far as siblings are mutually exclusive diagnoses).

IV.17. PURSUE-HYPOTHESIS

Pursuing a hypothesis has two components, testing it (TEST-HYPOTHESIS), followed by refining it (REFINE-HYPOTHESIS). After these two metarules are tried (in order, once), the hypothesis is marked as *pursued*.

Pursuing self followed by children brings about depth-first search. (Specifically, PURSUE-HYPOTHESIS puts the children in the differential and EXPLORE-AND-REFINE focuses on them.) This plan is based on the need to specialize a diagnosis (correctness), to remain focused (minimizing cognitive effort), and to consider more general disorders first (efficiency).

IV.18. REFINE-HYPOTHESIS

The effect of this task is to put taxonomic children or the causes of a state/category into the differential. If the hypothesis being refined has more than four descendents, a subset of possibilities is considered (REFINE-COMPLEX-HYPOTHESIS). For each child considered, the task APPLY-EVIDENCE-RULES is invoked (see PROCESS-HYPOTHESIS).

In order to reach a diagnosis in the etiologic taxonomy, this task requires that there be causal or subtype links from state/category hypotheses into the taxonomy, allowing them to be "refined" as etiologic hypotheses.

IV.19. REFINE-COMPLEX-HYPOTHESIS

Two metarules are used to select the common and unusual causes of the hypothesis. Ordinary domain rules, marked accordingly, are used to define these sets. The assumption is that, if only a few specializations can be considered (for economy), one should consider the common as well as the serious, unusual causes (for correctness). The less important hypotheses will be covered by the strategies of asking general questions and focused forward reasoning.

IV.20. PROCESS-HARD-DATA

Briefly, special functions are used to assemble set of "hard findings" that support hypotheses on the differential, reduce them to a set of "sources" (a lumbar puncture is the source for the CSF findings), and request the sources from the informant. Subsumption and definition relations are used

to infer the sources. Contraindications (dangerous side-effects) of gathering certain information is also considered. As described in PROCESS-FINDING, rules used by these findings are applied with subgoaling enabled. The program will return to GROUP-AND-DIFFERENTIATE and EXPLORE-AND-REFINE new hypotheses as necessary.

7. Acknowledgements

We are especially grateful to the late Timothy Beckett, MD, for serving as the expert-teacher in this research. Reed Letsinger participated in early discussions and helped implement the program. Bob London, Diane Hasling, Curt Kapsner, MD, David Wilkins, and Mark Richer have also contributed to the development of NEOMYCIN. I would like to thank Lewis Johnson for his careful reading and helpful suggestions. This paper was prepared in September 1983, then revised in February 1984 and August 1985.

This research has been supported in part by joint funding from ONR and ARI, Contract N00014-79C-0302, and more recently by ONR Contract N00014-85K-0305 and a grant from the Josiah Macy, Jr. Foundation. Computational resources are provided by the SUMEX-AIM facility (NIH grant RR 00785). NEOMYCIN is implemented in INTERLISP-D.

References

Aikins J. S. *Representation of control knowledge in expert systems*. in *Proceedings of the First AAAI*. pages 121-123, 1980.

Anderson, J. R. and Bower, G. H. *Human Associative Memory: A brief edition*. Hillsdale. NJ: Lawrence Erlbaum Associates 1980.

Anderson, J. R., Greeno, J. G., Kline, P. J., and Neves. D. M. Acquisition of problem-solving skill. In Anderson (editor), *Cognitive Skills and their Acquisition*. pages 191-230. Lawrence Erlbaum Associates. Hillsdale, NJ. 1981.

Benbassat, J., and Schiffmann, A. An approach to teaching the introduction to clinical medicine. *Annals of Internal Medicine*, 1976, *84*, 477-481.

Brown, J. S., Collins, A., and Harris, G. Artificial intelligence and learning strategies. In O'Neill (editor). *Learning Strategies*. . Academic Press. New York. 1977.

Bruner, J. S., Goodnow, J. J., and Austin, G. A. *A Study of Thinking*. New York: John Wiley & Sons. Inc. 1956.

Chandrasekaran, B., Gomez, F., Mittal, S. et al. *An approach to medical diagnosis based on conceptual schemes*, in *Proceedings of the Sixth International Joint Conference on Artificial Intelligence*, pages 134-142, International Joint Conference on Artificial Intelligence. Tokyo, 1979.

Chi, M. T. H., Feltovich, P. J., Glaser, R. Categorization and representation of physics problems by experts and novices. *Cognitive Science*, 1981, 5, 121-152.

Clancey, W. J. GUIDON. In Barr and Feigenbaum (editors), *The Handbook of Artificial Intelligence*, chapter Applications-oriented AI research: Education. William Kaufmann, Inc., Los Altos, 1982.

Clancey, W. J. The epistemology of a rule-based expert system: A framework for explanation. *Artificial Intelligence*, 1983, *20(3)*, 215-251.

Clancey, W. J. *The advantages of abstract control knowledge in expert system design*, in *Proceedings of the National Conference on AI*, pages 74-78, Washington, D.C., August, 1983.

Clancey, W.J. Methodology for Building an Intelligent Tutoring System. In Kintsch, Miller, and Polson (editors), *Method and Tactics in Cognitive Science*, pages 51-83. Lawrence Erlbaum Associates, Hillsdale, NJ, 1984.

Clancey, W. J. Representing control knowledge as abstract tasks and metarules. (To appear in *Computer Expert Systems*, eds. M. J. Coombs and L. Bolc, Springer-Verlag, in preparation).

Clancey, W. J. *Heuristic Classification*. Working Paper, KSL 85-5, Stanford University, March 1985.

(To appear in *Artificial Intelligence*).

Clancey, W. J. and Letsinger, R. NEOMYCIN: Reconfiguring a rule-based expert system for application to teaching. In Clancey, W. J. and Shortliffe, E. H. (editors), *Readings in Medical Artificial Intelligence: The First Decade*, pages 361-381. Addison-Wesley, Reading, 1984.

Cohen, P. R. and Grinberg, M.R. *A framework for heuristic reasoning about uncertainty*, in *Proceedings of the Eigth International Joint Conference on Artificial Intelligence*, pages 355-357, International Joint Conference on Artificial Intelligence, Karlsruhe, West Germany, August, 1983.

Davis, R. Meta-rules: reasoning about control *Artificial Intelligence*, 1980, *15*, 179-222.

Davis, R. *Diagnosis via causal reasoning: Paths of interaction and the locality principle*, in *Proceedings of the National Conference on AI*, pages 88-94. Washington, D.C., August, 1983.

Davis, R. and Lenat, D. *Knowledge-Based Systems in Artificial Intelligence.* New York: McGraw Hill 1982.

Duda, R. O. and Shortliffe, E. H. Expert systems research. *Science*, 1983, *220*, 261-268.

Elstein, A. S., Shulman, L. S., and Sprafka, S. A. *Medical problem solving: An analysis of clinical reasoning.* Cambridge: Harvard University Press 1978.

Ericsson, K. A. and Simon, H. A. Verbal reports as data. *Psychological Review*, 1980, *87*, 215-251.

Feigenbaum, E. A. *The art of artificial intelligence: I. Themes and case studies of knowledge engineering*, in *Proceedings of the Fifth International Joint Conference on Artificial Intelligence*, pages 1014-1029, August, 1977.

Feltovich, P. J., Johnson, P. E., Moller, J. H., and Swanson, D. B. The role and development of medical knowledge in diagnostic expertise. Presented at the 1980 Annual meeting of the American Educational Research Association; in Clancey and Shortliffe (editors), *Readings in Medical Artificial Intelligence: The First Decade*, Addison-Wesley, 1984).

Genesereth, M. R. The use of design descriptions in automated diagnosis. *Artificial Intelligence*, 1984, *24(1-3)*, 411-436.

Genesereth, M.R., Greiner, R., Smith, D.E. *MRS Manual.* Heuristic Programming Project Memo HPP-80-24, Stanford University, December 1981.

Gentner, D. and Stevens, A. (editors). *Mental models.* Hillsdale, NJ: Erlbaum 1983.

Hasling, D. W., Clancey, W. J., Rennels, G. R. Strategic explanations for a diagnostic consultation system. *The International Journal of Man-Machine Studies*, 1984, *20(1)*, 3-19.

Hayes-Roth, B. and Hayes-Roth, F. A cognitive model of planning. *Cognitive Science*, 1979, *3*,

275-310.

Hayes-Roth, F., Waterman, D., and Lenat, D. (eds.). *Building Expert Systems.* New York: Addison-Wesley 1983.

Kassirer, J. P., and Gorry, G. A. Clinical problem solving: A behavioral analysis. *Annals of Internal Medicine*, 1978, *89*, 245-255.

Kassirer, J. P., Kuipers, B. J., and Gorry, G. A. Toward a theory of clinical expertise. *The American Journal of Medicine*, 1982. *73*, 251-259.

Kolodner, J. Maintaining organization in a dynamic long-term memory. *Cognitive Science*, 1983, *7*, 243-280.

Kuipers B. and Kassirer, J. P. Causal reasoning in medicine: Analysis of a protocol. *Cognitive Science*, 1984, *8(4)*, 363-385.

Larkin, J. H., McDermott, J., Simon, D. P., Simon, H. A. Models of Competence in Solving Physics Problems. *Cognitive Science*, 1980, *4*, 317-348.

London, B. and Clancey, W. J. *Plan recognition strategies in student modeling: prediction and description*, in *Proceedings of the Second AAAI*, pages 335-338, 1982.

Miller, Peter B. *Strategy selection in medical diagnosis.* Technical Report AI-TR-153, Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Sept 1975.

Neves, D. M. and Anderson, J. R. Knowledge compilation: Mechanisms for the automization of cognitive skills. In Anderson (editor), *Cognitive Skills and their Acquisition*, pages 57-84. Lawrence Erlbaum Associates, Hillsdale, NJ, 1981.

Newell, A. The heuristic of George Polya and its relation to artificial intelligence. In R. Groner, M. Groner, and W. F. Bischof (editors), *Methods of Heuristics*, . Lawrence Erlbaum Associates, Hillsdale, NJ, 1983.

Newell, A. and Simon, H. A. *Human Problem Solving.* Englewood Cliffs: Prentice-Hall 1972.

Papert, S. *Mindstorms: Children, Computers, and Powerful Ideas.* : Basic Books, Inc. 1980.

Patil, R. S., Szolovits, P., and Schwartz, W. B. *Causal understanding of patient illness in medical diagnosis*, in *Proceedings of the Seventh International Joint Conference on Artificial Intelligence*, pages 893-899, August, 1981.

Patil, R. S., Szolovits, P., and Schwartz, W. B. *Information acquisition in diagnosis*, in *Proceedings of the National Conference on AI*, pages 345-348, Washington, D.C., August, 1982.

Pauker, S. G. and Szolovits, P. Analyzing and simulating taking the history of the present illness: context formation. In Schneider and Sagvall-Hein (editors), *Computational linguistics in*

medicine, pages 109-118. North-Holland, 1977.

Pauker, S. G., Gorry, G. A., Kassirer, J. P., and Schwartz, W. B. Toward the simulation of clinical cognition: taking a present illness by computer. *AJM*, 1976, *60*, 981-995.

Polya, G. *How to Solve It: a new aspect of mathematical method.* Princeton: Princeton University Press 1957.

Pople, H. Heuristic methods for imposing structure on ill-structured problems: the structuring of medical diagnostics. In P. Szolovits (editor), *Artificial Intelligence in Medicine*, pages 119-190. Westview Press, 1982.

Rubin, A. D. *Hypothesis formation and evaluation in medical diagnosis.* Technical Report AI-TR-316. Artificial Intelligence Laboratory, Massachusetts Institute of Technology, January 1975.

Rumelhart, D. E. and Norman, D. A. *Representation in memory.* Technical Report CHIP-116. Center for Human Information Processing, University of California, June 1983.

Schoenfeld, A. H. *Episodes and executive decisions in mathematical problem solving.* Technical Report, Hamilton College, Mathematics Department, 1981. Presented at the 1981 AERA Annual Meeting, April 1981.

Shortliffe, E. H. *Computer-based medical consultations: MYCIN.* New York: Elsevier 1976.

Simon, H. A. and Lea, G. Problem solving and rule induction. In Simon, H. A. (editor), *Models of Thought*, . Yale University Press, New Haven, 1979.

Swartout W. R. *Explaining and justifying in expert consulting programs*, in *Proceedings of the Seventh International Joint Conference on Artificial Intelligence*, pages 815-823, August, 1981.

Szolovits, P. and Pauker, S. G. Categorical and probabilistic reasoning in medical diagnosis. *Artificial Intelligence*, 1978, *11*, 115-144.

VanLehn, K. *Human procedural skill acquisition: Theory, model, and psychological validation*, in *Proceedings of the National Conference on AI*, pages 420-423, Washington, D.C., August, 1983.

VanLehn, K., Brown, J. S., Greeno, J. Competitive argumentation in computational theories of cognition. In Kintsch, Miller, and Polson (editors), *Method and Tactics in Cognitive Science*, pages 235-262. Lawrence Erlbaum Associates, Hillsdale, NJ, 1984.

VanLehn, K. and Brown, J. S. Planning nets: a representation for formalizing analogies and semantic models of procedural skills. In R. E. Snow, Frederico, P. A., and Montague, W. E. (editor), *Aptitude learning and instruction: Cognitive process and analyses*, . Lawrence Erlbaum Associates, Hillsdale, NJ, 1979.

Wescourt, K. T. and Hemphill, L. *Representing and teaching knowledge for*

troubleshooting/debugging. Technical Report, Institute for Mathematical Studies in the Social Sciences, Stanford University, 1978. Technical Report No. 292.

Yu, V. L. et al. Antimicrobial selection by a computer: a blinded evaluation by infectious disease experts. *Journal of the American Medical Association*, September 1979, *242(12)*, 1279-1282.