In W. J. Clancey and E.H. Shortliffe (Eds.), Readings in Medical Artificial Intelligence: The First Decade (pp. 256–274). Reading, MA: Addison-Wesley, 1984.

11

Intelligent Computer-Aided Instruction for Medical Diagnosis

William J. Clancey, Edward H. Shortliffe, and Bruce G. Buchanan

As AIM researchers began to develop techniques for allowing systems to explain their reasoning, some researchers became intrigued by the potential educational role of the developing methods. It became clear that advanced computer-aided instruction (CAI) programming techniques could be applied and extended in the medical setting. Intelligent computer-aided instruction (ICAI) differs from traditional CAI in its use of AI techniques for representing both subject material and teaching strategies.

Among ICAI programs, Clancey's GUIDON system described in this chapter is one of the largest and most complex. It contains all of the knowledge of MYCIN (Chapter 5) and uses a variety of techniques for mixed-initiative dialogue, student modeling, and response to partial student solutions. As a Stanford graduate student, Clancey had been involved in much of the early work on MYCIN and also became interested in ICAI and the possibility of adapting MYCIN for educational purposes. Thus GUIDON reflects the tremendous effort that went into building MYCIN's knowledge base of infectious disease rules, as well as nearly a decade of research in building ICAI systems. MYCIN's good performance in reaching decisions and giving explanations made a tutoring application of the knowledge base attractive. GUIDON also demonstrates the value of representing knowledge so that it can be applied in multiple settings, here for both consultation and teaching. This is the main advantage of separating

 From *Proceedings of the Third Annual Symposium on Computer Applications in Medical Computing,* Silver Springs, Md., October 1979, pp. 175–183.

the medical knowledge from the inference engine and encoding the medical knowledge in a stylized, program-readable form.

This chapter briefly outlines the difference between traditional instructional programs and ICAI. It then illustrates how GUIDON makes contributions in areas important to medical CAI: interacting with the student in a mixed-initiative dialogue (including the problems of feedback and realism), teaching problem-solving strategies, and assembling a computer-based curriculum.

In evaluating GUIDON's performance, one can see the value in the basic idea of formalizing teaching knowledge in procedures that are separate from the knowledge to be taught. However, the program is inherently limited by the MYCIN knowledge base. The rule set is poorly structured, does not contain pathophysiological knowledge for justifying the diagnostic associations, and does not explicitly state the strategies for gathering information and focusing on hypotheses. Thus the teaching perspective puts MYCIN's rules into sharp relief, revealing how they are crafted for good problem-solving, at the expense of making certain forms of common medical knowledge implicit (Clancey, 1983b).

GUIDON research evolved into a reconsideration of what a medical student needs to be taught about diagnosis. What are the diagnostic strategies of the primary care physician (as opposed to MYCIN's specialized top-down approach)? How are causal and subtype relations used to index medical knowledge during problem solving? This study of expertise (described briefly in Chapter 15) is complementary to Feltovich's psychological experiments, which reveal expert knowledge that is not formalized in medical textbooks (Chapter 12) and Gomez's and Chandrasekaran's emphasis on the interrelation of disease knowledge (Chapter 13). The other side of knowing what to teach is developing techniques for representing procedures in a way that makes explanation possible. Swartout's methods (Chapter 16) nicely complement the analysis and improvements to MYCIN that evolved from GUIDON research.

11.1 Introduction

Computer programs designed as aids for teaching medicine have been under development since the early 1960s. While some programs have been used for managing the use of conventional instructional material and grading tests, the predominant application has involved using the computer as a device that interacts with the student directly (Trzebiakowski and Ferguson, 1973). This application is generally called *computer-aided instruction* (CAI).

The goal of CAI research is to construct instructional programs that incorporate well-prepared course material in lessons that are optimized for each student. Early programs were either electronic "page-turners" that printed prepared text and simple, rote drills or practice monitors that printed problems and responded to the student's solutions using prestored answers and remedial comments. In the intelligent CAI (ICAI) programs of the 1970s, course material is represented independently of teaching procedures so that problems and remedial comments can be generated differently for each student. Research today focuses on the design of programs that can construct a truly insightful model of the student's strengths, weaknesses, and preferred style of learning. It is believed that AI techniques will make possible a new kind of learning environment.

In this paper, we outline traditional CAI techniques and discuss the advantages of ICAI programs. GUIDON, an ICAI program for teaching medical diagnosis, is introduced. We then characterize the design issues of past medical CAI programs and illustrate how GUIDON makes contributions to these areas of concern.

11.1.1 Traditional CAI

In traditional systems (Harless et al., 1971; Weinberg, 1973), a course material author attempts to anticipate every wrong student response and prespecifies branching to specific teaching material based on the underlying misconceptions that he or she associates with each wrong response. Branching on the basis of response was the first step toward individualization of instruction (Crowder, 1962). This style of CAI has been dubbed *ad hoc, frame-oriented* (AFO) CAI by Carbonell (1970) to stress its dependence on author-specified units of information.

11.1.2 Intelligent Computer-Aided Instruction

In spite of the widespread application of AFO CAI to many problem areas, many researchers believe that most AFO courses do not make the best use of computer technology. Carbonell has pointed out that a programmed text can do much of what is required in CAI systems of the AFO type (Carbonell, 1970). In this pioneering paper, Carbonell goes on to define a second type of CAI that is known today as knowledge-based or intelligent CAI. Early CAI systems did, of course, have representations of the subject matter they taught, but ICAI systems also carry on a natural language dialogue with the student and use the student's mistakes to diagnose misunderstandings. ICAI has also been called *generative* CAI (Wexler, 1970) because it is typified by programs that present problems by generating them from a large knowledge base representing the subject material to be taught (Koffman and Blount, 1973).

However, the kind of program that Carbonell was describing in his paper was to be more than just a problem generator. Rather, it was to be a computer-tutor that had the inductive powers of its human counterparts and could offer what Brown et al. (1976) call a *reactive learning environment,* in which the student is actively engaged with the instructional system and his or her interests and misunderstandings drive the tutorial dialogue.

The realization of the computer-tutor has involved increasingly complicated computer programs and has prompted CAI researchers to use artificial intelligence techniques. Artificial intelligence (AI) work in natural language understanding, the representation of knowledge, and methods of inference, as well as specific applications such as algebraic simplification, calculus, and theorem proving, have been applied by various researchers toward making CAI programs that are more intelligent and more effective. Early research on ICAI systems focused on representation of the subject matter (Carbonell, 1970; Suppes and Morningstar, 1972; Brown et al., 1974). The high level of domain expertise in these programs permitted them to be responsive in a wide range of problem-solving interactions.

In the mid-1970s, a second phase in the development of generative tutors has augmented knowledge representation techniques with expertise regarding the student's learning behavior, as well as tutorial strategies (Brown and Goldstein, 1977). AI techniques are used to construct models of the learner that represent his or her knowledge in terms of *issues* (Burton and Brown, 1976) or *skills* (Barr and Atkinson, 1975) that should be learned. These models then control tutoring strategies for presenting the instructional material. Finally, some ICAI programs are now using AI techniques to represent explicitly tutoring strategies themselves, gaining the advantages of flexibility and modularity of representation and control (Brown et al., 1976; Goldstein, 1977).

11.1.3 What Medical CAI Programs Attempt to Teach

Medical problem-solving skills can be categorized into three types: manipulative, interpersonal, and cognitive (Hoffer et al., 1975; Feinstein, 1977a). Manipulative skills involve acquisition of data and treatment by instrumentation. Interpersonal skills are involved in taking a patient history and discussing a diagnosis and alternative therapies. Cognitive skills comprise judgmental knowledge for managing a case: collecting data, reaching and testing hypotheses, and prescribing therapy. Most medical CAI programs are designed to teach cognitive skills. These skills are generally presented in two stages: acquisition of facts (e.g., properties of organisms, typical development of an infection) in preclinical years, and application of this knowledge to solve clinical problems (Hoffer et al., 1975). Most medical CAI programs present specific clinical problems that give the student an opportunity to apply his or her knowledge of facts, while following some diagnostic strategy for collecting data and forming hypotheses.

RULE507

IF: 1) The infection which requires therapy is meningitis,
2) Organisms were not seen on the stain of the culture,
3) The type of the infection is bacterial,
4) The patient does not have a head injury defect, and
5) The age of the patient is between 15 years and 55 years
THEN: The organisms that might be causing the infection are diplococcus-pneumoniae (.75) and neisseria-meningitidis (.74)

FIGURE 11-1 A typical MYCIN rule.

11.2 An Overview of the GUIDON System

The purpose of GUIDON research has been to develop a case method tutorial program that combines knowledge encoded in production rules [rules about infectious disease diagnosis provided by the MYCIN consultation system (Shortliffe, 1976) (see also Chapter 5)] with explicit tutorial discourse knowledge, while keeping the two distinct. GUIDON engages a student in a dialogue about a patient (a case) suspected of having an infection, and helps the student consider the relevant clinical and laboratory data for reaching a hypothesis about the causative organism(s). MYCIN's 450 diagnostic rules, one of which is shown in Figure 11-1, provide the underlying expertise that is used by the tutorial program in selecting topics to be discussed. MYCIN's methods provide a problem-solving approach for understanding the student's behavior and for defining skills to be taught. In addition, GUIDON has 200 tutorial rules, which include methods for guiding the dialogue economically, presenting diagnostic strategies, constructing a student model, and responding to the student's initiative.

A MYCIN rule consists of a set of preconditions (called the *premise*) that, if true, justifies the conclusion made in the *action* part of the rule. Conclusions are modified by *certainty factors* (Shortliffe and Buchanan, 1975), numbers that indicate how certain the rule's author is that the given conclusion is correct when the premise is true.

MYCIN's rules have not been modified for the tutoring application, but they are used in additional ways, for example, for forming quizzes, guiding the dialogue, summarizing evidence, and modeling the student's understanding. Flexible use of the rule set is made possible by the existence of *representational meta-knowledge* (Davis and Buchanan, 1977), which enables a program to take apart rules and reason about the components.

Two formal evaluations of MYCIN's performance have demonstrated that MYCIN's competence in selecting antimicrobial therapy for meningitis and for bacteremia is comparable to that of the infectious disease faculty at Stanford University School of Medicine (where MYCIN was developed)

(Yu et al., 1979a; 1979b). From this we conclude that MYCIN's rules capture a significant part of the knowledge necessary for demonstrably high performance in this domain.

11.3 GUIDON's Capabilities

The literature for medical CAI systems is extensive. Not all of the programs reported have a classic AFO design. For example, some programs use probability tables to generate "cases" (a patient with a specific problem) and use differential diagnosis to analyze the student's response and provide assistance (Entwisle and Entwisle, 1963; Steele et al., 1978). GUIDON is the first medical tutorial program we know of that is based on AI techniques. What contributions does it make to medical CAI? Most researchers address the following set of issues in the setting of GUIDON: (1) the nature of the dialogue interaction (including feedback and realism), (2) pedagogy, and (3) the problem of assembling a variety of cases.

We believe that GUIDON's main contribution lies in its capability to carry on a flexible dialogue with the student, allowing for problem-solving assistance in context, providing feedback for partial solutions at any time, and coping with the student's initiative in choosing topics and detail of discussion. Of secondary interest is the ease with which a library of cases can be assembled with minimal human intervention. Finally, current methods by which GUIDON provides assistance demonstrate that it has the potential for explicitly teaching strategies for doing medical diagnosis and perhaps for detecting which strategy the student is using.

11.3.1 Nature of the Dialogue Interaction

Medical CAI programs vary greatly in the nature of the dialogue that the program has with the student. Relevant issues considered here are

1. the form of input entered by the student,
2. the freedom of the student to direct the dialogue,
3. feedback for partial student solutions,
4. assistance provided for solving the problem, and
5. the realism of the interaction.

Input

Some programs restrict the student to key words or even numerical codes for diagnostic tests (Diamond et al., 1974), and others provide a humanlike interaction (by *ad hoc* means) that would tax the resources of any state-of-

Option type	*Examples*
Get case data	BLOCK, ALLDATA
Information retrieval	PENDING, DETAILS
Dialogue context	RULE, TOPIC
Convey what you know	IKNOW, HYPOTHESIS
Request assistance	HINT, TELLME
Change the topic	DISCUSS, STOP
Special	JUSTIFY, PROFILE

FIGURE 11-2 Some of the 30 options available in GUIDON dialogues.

the-art AI program (Swets and Feurzeig, 1965; Feurzeig et al., 1964). Some programs have borrowed AI techniques, for example, keyword analysis (Harless et al., 1971) and anaphoric resolution (Weber and Hageman, 1972). The main issue here is that it should be easy for the students to express themselves by using constructs that the program will be able to understand. This has been an important concern in ICAI in general. Some of the best results have been achieved by Burton (1976).

GUIDON, like most ICAI programs, accepts student input in the form of simple sentences. However, given the range of initiative we would like to allow (more than just collecting data), we are experimenting with the use of short-form options (Figure 11-2). This has the advantage that input is terse, and there is less chance of entering statements that the program cannot understand. In addition, the student is provided with a hard-copy listing of parameter designations that are recognized by the program, for example, BURNED, ALCOHOLIC, and AGE.

Some researchers believe that providing the student with a list of possible clinical factors to consider is unrealistic (a physician does not carry around a "crib sheet"). It serves as a crutch by cueing the student and, by the organization of the list, artificially distorts the student's approach (Hoffer et al., 1975; Harless et al., 1971). Other researchers claim that the list of relevant factors is helpful to a beginning student who would otherwise not know what questions to ask (Entwisle and Entwisle, 1963). We observed in one GUIDON tutorial that the student went down the list of factors in the hard-copy handbook once he had exhausted his own knowledge. This usage may not be harmful; problems may arise when students use the handbook inappropriately.

Student Initiative

An essential part of tutorial *dialogue management* (Clancey, 1979c) is allowing the student self-expression. In a mixed-initiative program, provision must be made for every potential *kind* of initiative that the student may

**** FACTORS**
The following factors will be useful to determine the type of the infection:
3a. whether the meningitis is partially treated
3b. whether a lumbar puncture has been performed on Pt538
3c. the duration of the neurological signs
3d. the CSF protein
(*)3e. whether Pt538 has a rash consistent with Herpes Zoster

.
.
.

Factors marked with * have already been stated. Based on what we have already mentioned, you should be able to conclude about: 3b.
**** USE 3C**
The duration of the neurological signs is 7.2 hours. Using other information available to you now, the duration of the neurological signs enables you to conclude that the type of the infection is bacterial (.2) viral (.2) fungal (-.4) tb (-.4) [RULE524].
**** IKNOW 3B**
Good. MYCIN has also made a decision.
The site of CULTURE-1 is evidence that a lumbar puncture has been performed on Pt538 [RULE112].
Back to our discussion of the type of the infection . . . .

.
.
.

Summary of evidence that the type of the infection is bacterial (.98) viral (-.76) fungal (-.83) tb (-.83):
29a. The fact that the meningitis is partially treated (.95) is evidence that the type of the infection is bacterial (.19) [RULE148]
29b. The WBC from the CSF and the percent of PMNs in the CSF is evidence that the type of the infection is bacterial (.9) viral (-.6) fungal (-.7) tb (-.7) [RULE501]

.
.
.

**** DETAILS 29A**
For which of the following factors would you like to see a review of the evidence? (select from the following):
1) the infection which requires therapy
2) the fact that the meningitis is partially treated (.95)
**** 2**
Summary of evidence that the meningitis is partially treated (.95):
32a. The infection which requires therapy, the fact that organisms were not seen on the stain of the pending CSF culture and the time since therapy with the cephalothin was started is evidence that the meningitis is partially treated (.95) [RULE145]
Do you want to see RULE148?
**** NO**

.
.
.

FIGURE 11-3 Sample use of options.

want to make. This includes being able to refer back to an earlier topic and provide more details, allowing the student to change the topic, and so on. We might summarize this by saying that we must allow the student to specify what he or she knows, wants to know more about, and wants to ignore. Figure 11-3 illustrates GUIDON's flexibility in responding to a student's initiative. Notice that tutorial remarks are indexed so that the student can easily refer to them later (by using them as arguments to options).

We allow the student to explore the reasoning of the underlying expert program, but we do not want the tutor to be simply a passive information retrieval system. In addition to laying out data and inferences clearly, the tutor has to reason about what constitutes reasonable, expected elaboration on the basis of what has been previously discussed. In the excerpt shown in Figure 11-3, GUIDON provided details for an inference (RULE148) by offering to support necessary preconditions that were not considered in the dialogue up to this point, though they could be inferred from known data.

Similarly, when the student takes the initiative by saying he or she knows something (see Figure 11-3), the tutor needs to determine what response makes sense, based on what it knows about the student's knowledge and shared goals for the tutorial session. The tutor may want to hold a detailed response in abeyance, simply acknowledge the student's remark, or probe for a proof. Selection among these *alternative dialogues* might require determining what the student could have inferred from previous interactions and the current situation. In the excerpt shown here, GUIDON decides that there is sufficient evidence that the student knows the solution to a relevant subproblem, so detailed discussion and probing are not necessary.

In many AFO systems, the flow of the dialogue is permanently fixed by the author of the course material. The student cannot change topics as he or she might wish, discussing subproblems and offering hypotheses to be evaluated. Systems like ATS (Weber and Hageman, 1972) have limited ability to reason with author-provided material (by indexing material with keywords), but it is still necessary for a course author to "sit down and play the role of the student for each major step in his tutorial." Thus it is still necessary to anticipate possible contingencies in each case individually.

Decoupling domain expertise from the dialogue program, an approach used by all ICAI systems, is a powerful way to provide flexible dialogue interaction. In GUIDON, *discourse procedures* (Clancey, 1979a) formalize how the program should behave in general terms, not in terms of the data and outcome of a particular case. A discourse procedure is a sequence of actions to be followed under conditions determined by the complexity of the material, the student's understanding of the material, and tutoring goals for the session. Each option available to the student generally has a discourse procedure associated with it. These procedures invoke other procedures for carrying on the dialogue, depending on circumstances of the particular situation.

For example, the procedure for the IKNOW option invokes the procedure for requesting and evaluating a student's hypothesis if the expert program has not yet made a final decision (so the tutor does not believe that the student can know the result). Otherwise, if the expert program has a final result, the procedure for discussing a completed topic is followed. Whether or not the student will be probed for details will depend

T-RULE5.02 [Directly state single, known rule]

IF: 1) There are rules having a bearing on this goal that have succeeded and have not been discussed, and
2) The number of rules having a bearing on this goal that have succeeded is 1, and
3) There is strong evidence that the student has applied this rule

THEN: Simply state the rule and its conclusion

FIGURE 11-4 T-rule for deciding how to complete discussion of a topic.

on the model that the tutor is building of the student's understanding (considered below).

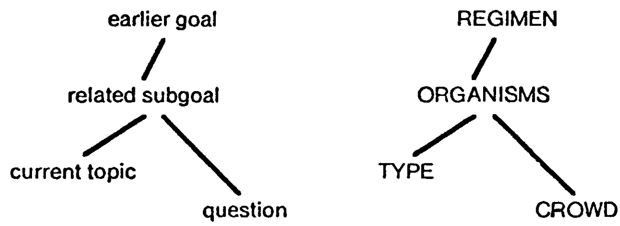
Conditional actions in discourse procedures are expressed as tutoring rules. Figure 11-4 shows the tutoring rule that caused GUIDON to acknowledge the student's statement about what he or she knew, rather than to ask for details.

As a final example of the problem of providing for and coping with the student's preferences, we will briefly consider the problem of focusing on topics during the dialogue. GUIDON allows a student to explicitly change the topic by using the DISCUSS option. However, student requests for data can also (implicitly) change the topic if the datum requested is not relevant to the current topic (cannot be used directly in any inference). In this respect, GUIDON enforces a goal-directed dialogue, so it will tell the student when he or she appears to be changing the topic. For example, if requested information is relevant to a previous, shallower subgoal (in the tree of topics by which the expert structures the problem solution), the tutor states this relation so that it is clear to the student what topic is currently being pursued (Figure 11-5).

Feedback

Nearly every discussion of medical CAI points to the importance of providing feedback to the student—primarily an evaluation of the student's solution, including mention of unnecessary and missed diagnostic questions. Programs vary from providing feedback at the end of the solution (Harless et al., 1971), to a step-by-step report that is inherent in AFO CAI (Feurzeig et al., 1964). Indeed, it is widely believed that the immediate correction of errors is an important capability of CAI (Hoffer et al., 1975). In a more general sense, the feedback that a CAI continuing education program offers provides a valuable tool for experienced physicians to evaluate their practices in light of new techniques (Brandt, 1974).

Providing feedback to the student is one problem that ICAI systems seem directly designed to resolve. A frame-oriented system is inherently unable to deal with unanticipated student errors; this would require that



****DOES THE PATIENT LIVE IN A CROWDED ENVIRONMENT?**
Pt538 does not live in a crowded environment.
Whether the patient does live in a crowded environment is not relevant to determining the type of the infection. It is a consideration we can use later when we return to our discussion of the organisms that might be causing the infection.

FIGURE 11-5 Coping with an indirectly relevant question.

the author prepare for all possible contingencies, a combinatorially unattainable goal. Part of the problem is that these programs cannot flexibly articulate diagnostic reasoning, however realistic or humanlike the responses may appear. That is, they cannot tolerate a random interruption of the dialogue in which the student poses a partial solution to the problem and asks for a response. Programs based on statistical tables (Kirsch, 1963; Entwisle and Entwisle, 1963; Steele et al., 1978) certainly have more flexibility to evaluate a given student's questions, but they have trouble characterizing the typicality of a case or what are the reasonable, important questions to ask. Of course, probabilistic programs have no way to justify correlations or to provide principles for remembering them.

MYCIN's diagnostic rules make it possible to provide feedback to the student at any time, regardless of the order and completeness of the student's collection of case data. In contrast with the frame-oriented approach, we prepare for contingencies abstractly, for example, what to do *in general* when the student says he or she knows something but the expert program does not know it, as opposed to what to do if the student says he or she believes X after questions Q1 and Q2 have been asked but not Q3.

In order to provide feedback, the tutor must be able to account for the student's behavior in terms of problem-solving methods and strategies. This may involve interpreting wrong, but consistent, behavior in terms of misconceptions. In correcting the student, the tutor may need to refer back to past student actions and compare alternative approaches. Allowing for elaboration in a mixed-initiative exchange poses additional technical problems in dialogue management, such as maintaining context while each participant varies the topic.

In the excerpt that follows (Figure 11-6), the student's hypothesis for a subproblem is correct, but after analyzing the previous history of interaction with the student, GUIDON is not sure that the student really knows

What is the type of the infection?
** **BACTERIAL**
What facts about this case tell you that the type of the infection is bacterial? (Enter one factor per line.)
** **PMNS**
The percent of PMNs in the CBC is not relevant for determining the type of the infection.
** **PMNS IN THE CSF**
** **CEREBRAL-FLUID GLUCOSE VALUE**
** **PROTEIN OF THE CSF**
**
You're right about the CSF protein, the percent of PMNs in the CSF and the CSF glucose value, however, you should have mentioned the fact that the meningitis is partially treated, the WBC from the CSF, and the fact that the simultaneous blood glucose value is not known.

.
.
.

MYCIN rejected some possibilities you did not mention:

What facts about this case tell you that the type of the infection is not viral?

.
.
.

FIGURE 11-6 Responding to a student hypothesis.

why this hypothesis is correct. Therefore, instead of accepting this answer, GUIDON requests relevant factors to support the hypothesis and probes further to see if the student knows why an alternative hypothesis is discredited.

Note that GUIDON does more than say which student answers are right and which are wrong: the program actually reasons about the expertise that is required to solve the problem. Referring to the rules that MYCIN uses to solve subproblems (such as determining whether a meningitis infection is bacterial, fungal, viral, or tuberculous), GUIDON decides which of these rules, if any, are known by the student. That is, what inference chains are consistent with the student's behavior?

As the student inquires about the patient and receives more case data, this same information is given, piece by piece, to the MYCIN program. MYCIN applies rules, as far as it can, to carry its reasoning about the case forward. Thus, at any one time, some of the rules MYCIN follows for determining, say, the type of the infection, have made a conclusion, while others will require more information about the patient before they can be applied. Then, when the student offers an hypothesis, GUIDON compares the student's partial solution to MYCIN's, and, referring to what is known about the student's level of expertise and records from past interactions with him or her, the program attempts to explain the student's solution in terms of MYCIN rules. This analysis is complicated by the fact that a particular hypothesis about the problem may be indicated by more than one rule or that negative evidence may outweigh positive evidence.

A student model constructed in this way is called an *overlay model* because the tutorial program attempts to explain the student's behavior solely in terms of a fixed, single set of "expert knowledge," in this case, MYCIN's

** HINT
Hint: Look for strong, positive evidence that Diplococcus-pneumoniae is one of the organisms that might be causing the infection.

.
.
.

** HELP
Let's consider whether J.Smith has a head injury.

We already know that the patient has not had an injury to the central nervous system; this is evidence that he does not have a head injury [RULE509].

We now have strong evidence that Diplococcus-pneumoniae and Neisseria-meningitidis are organisms that might be causing the infection (considering the age of J.Smith and the fact that he does not have a head injury [RULE507]).

It remains for us to consider other factors for determining the organisms that might be causing the infection.

.
.
.

FIGURE 11-7 Providing assistance in context.

rules. Overlay models were first used by Burton and Brown (1976); the technique was elaborated further by Carr and Goldstein (1977). Limitations of this approach are considered in Section 11.4.

Assistance

Another basic property of a tutorial dialogue is the extent to which the program is able to provide assistance for solving the problem. Ideally, the tutor's guidance should be based on the student's partial solution. In general, this is a difficult problem because it requires that the tutor be sensitive to the student's current problem-solving strategy and the kind of advice he or she prefers (a hint? full details?). It must also be able to articulate problem-solving methods that might be applied (a problem of knowledge representation).

Using its overlay model of the student, GUIDON is able to provide assistance by once again reasoning about the rules that MYCIN has been able to apply at the time that the student requests help. In the example shown here (Figure 11-7), GUIDON provides assistance by applying a solution method (RULE507) that suggests evidence contrary to that which has been discussed to this point of the dialogue. In this case the selected method was alluded to in an earlier hint.

The program has many ways to present a rule to the student, such as forming a question or discussing each clause of the rule explicitly. Here GUIDON demonstrates the applicability of the solution method by showing how the truth of the single precondition that remains to be considered can be inferred from known evidence (RULE509). The inference is trivial, so it is given directly rather than opened up for discussion. GUIDON then applies the original method (RULE507) and comments about the status of the current subproblem.

Thus providing assistance can involve applying a teaching strategy that carries the solution of the problem forward. This in turn requires being able to articulate reasoning on the basis of what the student knows, according to principles of economical presentation.

Observe that, to provide feedback and assistance, it is not sufficient simply to have a model of what the student knows: the program needs methods for presenting new material to the student. In a knowledge-based tutor, presentations are generated solely from the knowledge base of rules and facts. This requires that the tutor have presentation methods that opportunistically *adapt material to the needs of the dialogue.* In particular, the tutor has to be sensitive to how a tutorial dialogue fits together, including what kinds of interruptions and probing are reasonable and expected in this kind of discourse. GUIDON demonstrates its sensitivity to these concerns when it corrects the student before quizzing him or her about "missing hypotheses," chooses between terse and lengthy discussions of inferences, follows up on previous hints, and comments on the status of a subproblem after an inference has been discussed ("other factors remain to be considered . . .").

Realism of Course Material

Implicit in the design of most medical CAI programs is the assumption that similarity of the tutorial problem-solving environment to actual conditions in actual practice (e.g., the timing and sequence of events, interactions with assistants) is important to ensure transferability of learning to the clinical setting. Furthermore, when the purpose of the tutorial is to make the student familiar with his or her responsibilities on the ward, realism is an integral part of the course material.

Some medical CAI systems attempt to present the student with a "simulated patient" who can be interviewed and given therapy (Harless et al., 1971). Others place the student in a simulated hospital setting in which the student, as attending physician, orders tests, comes back "the next day" to reevaluate the patient, etc. (Feurzeig et al., 1964). The majority of programs, like GUIDON, simulate the kind of tutorial discussion that the student might have on the hospital wards with a resident physician or classroom instructor (Diamond et al., 1974; Weber and Hageman, 1972).

Compared to the investigation of discourse, modeling, and pedagogy, the simulation of a particular real-world problem-solving environment has not been a major focus of ICAI research. However, it seems probable that AI research dealing with the importance of knowledge about prototypical problem situations in everyday reasoning will be useful for generating realistic cases to be solved by the student, as well as for simulating moment-by-moment patient events.

11.3.2 Pedagogy

The main pedagogical question in CAI programs concerns what diagnostic strategy, if any, should be conveyed to the student and how this should be done. For example, one program is specifically designed to teach Weed's "problem-oriented approach" (Benbassat and Schiffmann, 1976); it imposes a fixed logical order on the kinds of questions that the student asks. Other researchers believe that a completely uninterrupted, "free-form" style is an essential part of teaching independent thinking and responsible problem solving (Harless et al., 1971).

GUIDON attempts to allow for a free-form style while still conveying problem-solving strategies. The student is free to gather case data in any order, but is told when he or she is wandering from the topic under consideration. Hints and help are based on a problem-solving strategy (Figure 11-7) that could be altered (nontrivially) to reflect Weed's approach.

CAI programs, including ICAI ones, have generally not focused on teaching problem-solving strategies because it is difficult to represent them internally in a way that allows the program to use them for teaching material (e.g., mentioning the strategy when posing a hint based on it) as well as for modeling the student (i.e., knowing that the student is following a particular strategy). Technical problems aside, medical CAI programs have probably focused on teaching facts and decision rules rather than strategies because "there is little agreement among medical educators about an explicit and detailed model of clinical competence" (Hoffer et al., 1975). Only recently have physicians developed scientific descriptions of alternative problem-solving strategies (Kassirer and Gorry, 1978), which, interestingly enough, have been based on AI research.

It is possible that the expert modules of ICAI systems (for example, the role MYCIN plays in the GUIDON program) will provide useful testbeds for formalizing and experimenting with problem-solving strategies. Meta-rules (Davis and Buchanan, 1977) and strategies for revising hypotheses provide a language by which GUIDON can be used to formalize and measure diagnostic competency. AI alone cannot provide the missing physiological, chemical, and physical knowledge that will provide a deeper understanding of medical problems, but AI approaches to search and hypothesis confirmation may provide suitable information-processing models for talking about different approaches to diagnosis.

11.3.3 Case Generation

A major advantage of CAI over other forms of medical instruction is that it has the potential to expose a student to a variety of cases that might far exceed what actual hospital experience would provide. However, to achieve this potential, it has been necessary in traditional medical CAI to spend many days designing and debugging each case. Various estimates are given for the ratio of design time to course time, and 1 week of design for a 20-minute course is not atypical (Bitzer and Bitzer, 1973). Researchers emphasize the ease with which their frame-oriented systems may be changed, but it must be remembered that each clever addition in one case must be repeated in others, a clearly untenable situation if the science of instruction is to advance. GUIDON offers an improvement over the traditional approach: experience is cumulative, so that modifications made on the basis of one tutorial interaction will automatically show up in similar situations during discussion of any other case.

By coupling GUIDON to the patient library that has been accumulated during the testing of the MYCIN consultation program, formal course preparation is unnecessary. Given that MYCIN can work out the reasoning for solving a case and GUIDON can selectively discuss it with a student, preparation time for a new case is reduced to less than an hour for each hour of course time, allowing for the provision of some annotations that point out the pedagogical value of the case. (Patient cases are entered into the MYCIN system for the purpose of receiving a consultation or for testing the program, so the case library is available to GUIDON at no cost.) Eventually, given case-selection strategies (based on knowledge about the spectrum of cases), even these annotations would be unnecessary, and course preparation time would be eliminated.

11.4 Limitations of the Approach

A potential weakness of the GUIDON program is that it attempts to explain the student's behavior solely in terms of MYCIN's rules. When the student is basing his or her questions and hypotheses on incorrect rules, GUIDON is not able to detect these rules and correct them directly. It is also possible that the student's concepts are different from MYCIN's, so the student's conclusions might be correct, but he or she will want to support them using reasoning that is different from MYCIN's. This could involve something as simple as wanting to refer to the patient's age in general terms (infant, adolescent), while MYCIN only recognizes precise numerical ages.

Representing a hierarchy of common concepts is a relatively straightforward change to MYCIN. However, modeling medical reasoning in terms of an alternative rule set (not just a subset of MYCIN's) is a theory-formation problem that goes beyond the current capabilities of AI. Brown has investigated methods for modeling improper procedures that students follow when doing subtraction (Brown and Burton, 1978). Unfortunately, this technique does not generalize easily to semantically complex domains like medicine. The approach followed by Stevens for meteorology is to collect data about kinds of student misconceptions and then incorporate

**** TOPIC**
We are discussing the degree of reversibility of airway obstruction of PATIENT-7. Hypotheses you should be considering are: none, insignificant, slight, good and significant.

**** WHAT IS THE CHANGE IN FORCED EXPIRATORY VOLUME?**
The change in Forced-Expiratory-Volume (1sec.) after bronchodilation of M.Smith is 31.

**** WHAT IS THE CHANGE IN RESISTANCE?**
The change in Resistance after bronchodilation of M.Smith is 20.

We have substantially completed our discussion of the degree of reversibility of airway obstruction of M.Smith. Would you like to make a hypothesis now (as opposed to gathering more information)?

**** NO**

We will now consider evidence that leads to a conclusion about the degree of reversibility of airway obstruction of M.Smith.

Here is a conclusion that you may not have considered . . . .

The fact that the change in Forced-Expiratory-Volume (1sec.) after bronchodilation of M.Smith and the change in Resistance after bronchodilation are known is evidence that there are post bronchodilation test results [RULE028].

Given the post bronchodilation test results, does this tell you that the degree of reversibility of airway obstruction is slight?

**** YES**
No, given this fact, we can conclude that the degree of reversibility of airway obstruction of M. Smith is significant [RULE020].

The degree of reversibility of airway obstruction would be slight when the change in Forced-Expiratory-Volume (1sec.) after bronchodilation of the patient is between 1 and 5.

FIGURE 11-8 Excerpt from a PUFF tutorial.

these variations in the modeling process (Stevens et al., 1978). We believe that GUIDON tutorials will provide the opportunity for furthering this study.

11.5 Experimentation with Other Domains

Besides being able to use a fixed set of teaching strategies (the discourse procedures) to tutor different cases, GUIDON is able to provide tutorials in any problem area for which a MYCIN-like knowledge base of decision rules and fact tables have been formalized (van Melle, 1979). This affords an important perspective on the generality of the discourse and pedagogical rules. At this time two other medical consultation programs have been developed using MYCIN's rule formalism: PUFF (Kunz et al., 1978) provides diagnoses about pulmonary disease; HEADMED (Heiser and Brooks, 1978) advises about the use of psychopharmaceuticals.

The example shown in Figure 11-8 is taken from a GUIDON tutorial that uses PUFF's knowledge base for the problem of pulmonary function analysis. This example shows the program taking the initiative to present new information to the student. GUIDON first interrupts the student's data collection to suggest that the student make an hypothesis; but the student does not do so. The program then observes that there is a particular problem-solving method that can be applied and that is probably known to the student (RULE020). However, the student probably cannot apply the method to this case because he or she does not know how to verify a necessary precondition. GUIDON presents the inference that it believes is unknown to the student (RULE028), and then asks him or her to take this evidence forward.

Experimental tutorials with other knowledge bases have revealed that the effectiveness of discourse strategies for carrying on a dialogue economically is determined in part by the depth and breadth of the reasoning tree for solving the problem. When a solution involves many rules at a given level (for example, when there are many rules to determine the organism causing the infection), the tutor and student will not have time to discuss each rule in the same degree of detail. Similarly, when inference chains are long, then an effective discourse strategy will entail summarizing evidence on a high level, rather than considering each subgoal in the chain.

11.6 Conclusions

In traditional medical CAI, as well as in some ICAI programs, teaching expertise is "compiled" into the program, combining all kinds of problem-solving, communication, and pedagogical strategies. In GUIDON we make the important step of explicitly codifying teaching expertise within the program as a body of rules to follow in various situations. In fact, the rules *are* the program. By decoupling medical expertise from dialogue strategies, we are able to focus more directly on rules of conversation and communication or "kibitzing" strategies (Burton, 1979). This is one of the special advantages of GUIDON's framework of discourse knowledge. GUIDON's tutoring rules never mention cultures or disease or any application area. Instead, the rules state how to teach, how to reply to a student, and how to guide a student. With these explicit principles before us, we are in a much better position to say what we are evaluating when we test the program.

The key to GUIDON's contributions lies in the flexibility of its representation of teaching and problem-solving knowledge. MYCIN's domain rules can be reasoned about to construct a student model, to provide assistance, and to select presentation methods. GUIDON's tutoring rules, wholly separated from the domain rules, constitute general procedures that can be followed any time in a dialogue, giving the program the capability to cope with arbitrary student initiative within the considerable range of expression the program's options allow. Finally, these tutoring rules are problem- and domain-independent, allowing flexibility for teaching any case formalized in a MYCIN-like consultation system.

With respect to the issues of dialogue interaction, pedagogy, and case generation, GUIDON's primary contributions to medical CAI are greater individualization of tutorials, a framework for expressing and accumulating tutorial dialogue expertise, and a language for diagnostic problem-solving strategies. By constructing a model of problem-solving strategies in a student model, something not possible with traditional technology, ICAI systems could provide a basis for critiquing and teaching diagnosis in terms that even go beyond classroom or clinical experience.

ACKNOWLEDGMENTS

This research was supported in part by an ONR contract (N00014-79C-0302) and a National Library of Medicine career development award (LM00048) to E. H. Shortliffe.