PART EIGHT

Tutoring

25

Intelligent Computer-Aided Instruction

The idea of directly teaching students "how to think" goes back at least to Polya (1957), if not to Socrates, but it reached a new stage of development in Papert's laboratory (Papert, 1970). In the LOGO lab, young students were taught AI concepts such as hierarchical decomposition, opening up a new dimension by which they could take apart a problem and reason about its solution. In part, Polya's heuristics have seemed vague and too general, too hard to follow in real problems (Newell, 1983). But progress in AI programming, particularly expert system design, has suggested a vocabulary of *structural concepts* that we now see must be conveyed along with the heuristics to make them intelligible (see Chapter 29).

Developing in parallel with Papert's educational experiments and capitalizing even more directly on AI technology, programs called *intelligent tutoring systems* (ITS) were constructed in the 1970s. In contrast with the computer-aided instruction (CAI) programs of the 1960s, these programs used new AI formalisms to separate out the subject matter they teach from the programs that control interactions with students. This is called intelligent computer-aided instruction (ICAI). This approach has several advantages: it becomes possible to keep records of what the student knows; the logic of teaching can be generalized and applied to multiple problems in multiple problem domains; and a model of student knowledge can be inferred from student behavior and used as a basis for tutoring. The well-known milestones in ITS research include:

- interacting with the student in a mixed-initiative dialogue[1] (Carbonell, 1970b) and tutoring by the Socratic method (Collins, 1976)

Parts of this chapter are taken from the final report to the Office of Naval Research for the first period of GUIDON research (1979–1982). That report appeared as a technical memo (HPP-82-2) written by William J. Clancey and Bruce G. Buchanan from the Heuristic Programming Project, Department of Computer Science, Stanford University.

[1]In a mixed-initiative dialogue between a student and a program, either party can initiate questions and expect reasonable responses from the other party. This contrasts sharply with drill and practice programs or MYCIN's dialogue, in which users cannot volunteer information or direct the program's reasoning.

- evaluating student hypotheses for consistency with measurements taken (Brown et al., 1975)
- enumerating bugs in causal reasoning (Stevens et al., 1978)
- interpreting student behavior in terms of expert knowledge ("overlay model") (Burton, 1979; Carr and Goldstein, 1977; Clancey, 1979b)
- codifying discourse procedures for teaching (Clancey, 1979c)
- constructing models of incorrect plans or procedures (Genesereth, 1981; Brown and Burton, 1978)
- relating incorrect procedures to a generative theory (Brown and Van-Lehn, 1980)

The record of ITS research reveals a few recurring questions:

1. *Nature of expertise:* What is the knowledge we want to teach a student?
2. *Modeling:* How can we determine what the student knows?
3. *Tutoring:* How can we improve the student's performance?

Almost invariably, researchers have backed off from initially focusing on the last question—"How shall we teach?"—to reconsider the second question, that of building a model of the student's knowledge. This follows from the assumption that student errors are not random but reflect misconceptions about the procedure to be followed or facts in the problem domain and that the best teaching strategy is to address directly the student's misconceptions.

In order to extend the research in building models of misconceptions in well-understood domains such as subtraction to more complex domains such as physics, medicine, and electronic troubleshooting, we need a sounder understanding of the nature of knowledge and expertise. Comparison studies of experts and novices (Chi et al., 1980; Feltovich et al., 1980; Lesgold, 1983) reveal that how the expert structures a problem, the very concepts he or she uses for thinking about the problem, distinguishes an expert's reasoning from a student's often formal, bottom-up approach. These studies suggest that we might directly convey to the student the kinds of quick associations, patterns, and reasoning strategies that experts build up tediously over long exposure to many kinds of problems—the kind of knowledge that tends not to be written down in basic textbooks.

It is with this premise—that we will be better teachers by better understanding expertise—that research on expert systems becomes of keen interest to the educator. These *knowledge-based programs* contain within them a large number of facts and rulelike associations for solving problems in restricted domains of medicine, science, and engineering. While these programs were developed originally just for the sake of building systems that could solve difficult problems, they have special interest to research in cognitive science as simulation models that can be used as a "laboratory workbench" for experimenting with knowledge structures and control strategies. By altering the "program as a model," one can test hypotheses about human performance [for example, see Johnson et al. (1981)].

Another natural application for expert systems in education is to use them as the "knowledge foundation" for an intelligent tutoring system. Brown pioneered this technology in the SOPHIE3 system (Brown et al., 1974), which took a student through the paces of debugging a circuit. Brown, Collins (1978), and Goldstein (1978) pioneered the use of production rules to express knowledge about how to interact with a student and how to interpret his or her behavior. The first tutor built on top of a complex expert system was GUIDON (Clancey, 1979a), using MYCIN's 450 production rules and tables for teaching medical diagnosis by the case method. GUIDON's teaching expertise is represented cleanly and independently of the domain rules; it has been demonstrated for both medical and engineering domains.[2]

25.1 Tutoring from MYCIN's Knowledge Base

Early in the course of building MYCIN, we observed that a program with enough medical knowledge for consulting had high potential for educating physicians and medical students. Physicians who seek advice from a consultant—human or machine—do so because they are uncertain whether or not they are ignoring important possibilities or making conclusions that are correct. Along with confirmation and advice, a consultant provides reasons, answers questions, and cites related issues. The educational component of a computer-based consultant was too obvious for us to ignore.

MYCIN's *conclusions* alone would not help a physician understand the medical context of the case he or she presents to the program. But the *dialogue* with MYCIN already begins to illuminate what are the key factors for reaching those conclusions. Because MYCIN asks whether or not the patient has been burned, for example, a physician is reminded that this factor is relevant in this context. This is very passive instruction, however, and does not approach the Socratic dialogue we expect from good teachers.

MYCIN's *explanation capabilities* were introduced to give a physician an opportunity to examine parts of the dialogue he or she found puzzling. When the program asks whether or not the patient has burns, the user can inquire why that information is relevant. As described in Part Six, answers to such inquiries elucidate MYCIN's line of reasoning on the case at hand and thus provide brief instructional interchanges in the course of a consultation. Similarly, the *question-answering capabilities* give a physician

[2]In GUIDON teaching knowledge is treated as a form of expertise. That is, GUIDON has a knowledge base of teaching rules that is distinct from MYCIN's knowledge base of infectious disease rules.

instructional access to the static knowledge base. Although we now understand better the difference between MYCIN's explanation capabilities and an active tutor, we enthusiastically wrote in 1974 (Shortliffe, 1974, pp. 230–231):

> As . . . emphasized throughout this report, an ability to instruct the user was an important consideration during the design of MYCIN. We believe it is possible to learn a great deal simply by asking MYCIN for consultative advice and taking advantage of the program's explanation capabilities. It is quite likely, in fact, that medical students in their clinical years will comprise a large percentage of MYCIN's regular users.

We were also aware of the need to make an instructional program more active, as others in AI were doing. In 1974 we noted (Shortliffe, 1974, p. 231):

> It would be possible . . . to adapt MYCIN so that its emphasis became primarily educational rather than consultative. This could be accomplished in a number of ways. In one scenario, MYCIN would present a sample patient to a student. The program would then judge the student's ability to ask important questions and to reach valid conclusions regarding both the identity of the organism(s) and the most appropriate therapeutic regimen. By comparing the student's questions and decisions to its own, MYCIN could infer inadequacies in the user's knowledge and enter into a tutorial discourse customized for the student. . . . We have no plans to pursue this application in the near future.

It was within this intellectual context that Clancey began asking about the adequacy of MYCIN's knowledge base for education. We initially believed that the rules and tables MYCIN used for diagnosing causes of infections would be a sufficient instructional base for an ICAI program. We felt that the only missing intelligence was pedagogical knowledge: how to carry on a mixed-initiative dialogue, how to select and present information, how to build and use a model of the student, and so on. Clancey began work on a tutorial program, called GUIDON, within two years after the material quoted above was written. The initial model of interaction between MYCIN and GUIDON is shown schematically in Figure 25-1.

GUIDON was first conceived as an extension of the explanation system of the MYCIN consultation program. This previous research provided the building blocks for a teaching program:

- modular representation of knowledge in production rules
- English translation of the internal rule representation
- a developed "history trace" facility for recording reasoning steps
- representation in the system of the grammar of its rules, so they can be parsed and reasoned about by the system itself
- an explanation subsystem with a well-developed vocabulary for the log-

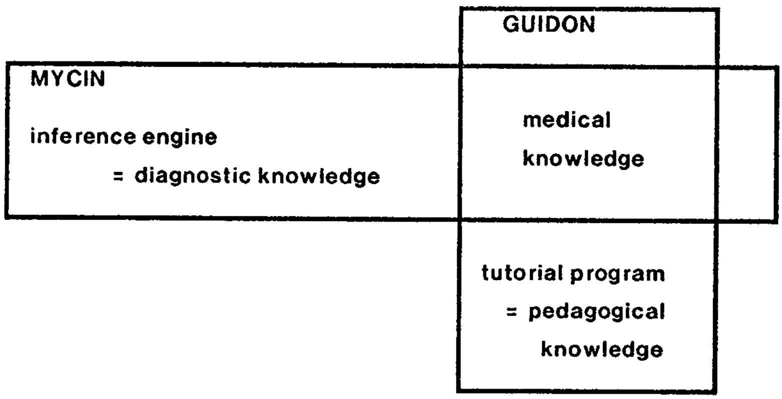

FIGURE 25-1 Model of interaction between MYCIN and GUIDON.

ical kinds of questions that can be asked about MYCIN's reasoning ("Why didn't you ask X?" or "How did you use X to conclude about Y?")

With this foundation, we constructed a tutoring program that would take MYCIN's solution to a problem, analyze it, and use it as the basis for a dialogue with a student trying to solve the same problem. About two hundred tutoring rules were developed, organized into "discourse procedures" for carrying on the dialogue (offering advice, deciding whether and how to interrupt, etc.) (Clancey, 1979b). Student modeling rules were used to interpret a student's partial problem solutions in terms of MYCIN's knowledge, and the resulting model was used to decide how much to tell the student and when to test his or her understanding.

Our 1978 proposal to the Office of Naval Research (ONR) for GUIDON research outlined investigation of both problem-solving and teaching strategies. With the program so well developed, it was expected that early experimentation could be done with alternative teaching approaches. However, during preliminary discussions with other researchers in this field, a key question was repeatedly raised. To paraphrase John Brown (August 2, 1978, at Stanford University):

> What is the nature of the expertise to be transmitted by this system [GUIDON]? You are not just unfolding a chain of inferences; there is also glue or a model of process. . . . What makes a rule click?

Following this lead, we began to concentrate on the nature of the expertise to be taught. GUIDON's interactions were studied, particularly the kind of feedback it was able to provide in response to incorrect partial solutions. The inability of the program to provide strategical guidance—

advice about what to do next—revealed that the "glue" that was missing had something to do with the system of rules as a whole. With over 400 rules to learn, there had to be some kind of underlying logic that made them fit together; the idea of teaching a set of weakly structured rules was now seriously in question. Significantly, this issue had not arisen in the years of developing MYCIN but was now apparently critical for teaching, and probably had important implications for MYCIN's explanation and knowledge acquisition capabilities as well.

It soon became clear that GUIDON needed to know more than MYCIN knows about diagnosis. MYCIN's route from goal to specific questions is not the only acceptable line of reasoning or strategy for gathering evidence. The order in which MYCIN asks for test results, for example, is often arbitrary. Thus a student is not necessarily wrong if he or she deviates from that order. Moreover, MYCIN's explicit knowledge about medicine is often less complete than what a tutor needs to convey to a student. It is associational knowledge and does not represent causal relationships explicitly. The causal models have been "compiled into" the associations. Thus MYCIN cannot justify an inference from A to B in terms of a causal chain, $A \rightarrow A_1 \rightarrow A_2 \rightarrow B$. A student, therefore, is left with an incomplete, and easily forgotten, model of the disease process. These two major shortcomings are discussed at length in Chapters 26 and 29.

25.2 Recent Work

Complementing the studies of differences between experts and novices, as well as our own work at Stanford on systems that explain their reasoning, our recent work has shown that expert systems must represent knowledge in a special way if it is to be used for teaching (Chapter 29). First, *the program must convey organizations and approaches that are useful to the student;* this argues for a knowledge base that reflects ways of thinking used by people (the hypothesis formation approach). Second, *various kinds of knowledge must be separated out and made explicit so reasoning steps can be carefully articulated*—the expert's associations must be decomposed into structural and strategic components. Under our current contract with ONR, such an expert system, called NEOMYCIN, has been constructed (Clancey and Letsinger, 1981). It is being readied for use with students through both active development of its knowledge base and construction of modeling programs that will use it as a basis for interpreting student behavior.

The ultimate goal of our work in the past few years has been to use NEOMYCIN for directly teaching diagnostic problem solving to students. Students will have the usual classroom background but will be exposed in this tutoring system to a way of thinking about and organizing their textbook knowledge that is usually taught only informally in apprenticeship settings. That is, we are beginning to capture in an expert system what we deem to be the essential knowledge that separates the expert from the novice and teaching it to the novice in practice sessions in which its value for getting a handle on difficult, confusing problems will be readily apparent. Empirical studies are a key part of this research.

We view our work as the logical "next step" in knowledge-based tutoring. Just as representing expert knowledge in a simulation program provides a vehicle for testing hypotheses about how people reason, using this knowledge in a tutoring system will enable us to see how the knowledge might be explained and recognized in student behavior. The experience with the first version of GUIDON, as detailed further in Chapter 26, illustrates how the tutoring framework provides a "forcing function" that requires us to clarify what we want to teach and how we want to teach it.

During 1979–1980 a study was undertaken to determine how an expert remembered MYCIN's rules (the "model of process" glue) and how he or she remembered *to use them.* This study utilized several common AI methods for knowledge acquisition but built upon them significantly through the development of an epistemological framework for characterizing kinds of knowledge, detailed in Chapter 29. The expert's explanations were characterized in terms of: strategy, structure, inference rule, and support. With this kind of framework, discussions with the expert were more easily focused, and experiments were devised for filling in the gaps in what we were told.

By the end of 1980, we had formulated and implemented a new, comprehensive psychological model of medical diagnosis (Clancey and Letsinger, 1981) based on extensive discussions with Dr. Tim Beckett. NEOMYCIN is a consultation program in which MYCIN's rules are reconfigured according to our epistemological framework. That is, the knowledge representation separates out the inference rules (simple associations among data and hypotheses) from the structural and strategic knowledge: *we separate out what a heuristic is from when it is to be applied.* Moreover, the strategies and structure we have chosen model how an expert reasons. We have attempted to capture the expert's forward-directed inferences, "diagnostic task structure," and the types of focusing strategies he or she uses. This explicit formulation of diagnostic strategy in the form of meta-rules is exactly the material that our original proposal only mentioned as a hopeful aside. Recently, we have been fine-tuning NEOMYCIN, investigating its applicability to other domains, and exploiting it as the foundation of a student model.

25.3 Multiple Uses of the Same Knowledge

From a slightly different perspective, we were also interested in exploring the question of whether or not one knowledge base could be used for multiple purposes. From the DENDRAL and MYCIN experiences, we

Predictive Rule for DENDRAL:

IF the molecular structure contains the subgraph

$$\begin{array}{c} \mathrm{O} \\ \| \\ \mathrm{R_1{-}C{-}R_2} \end{array} \qquad \text{(where } R_1 \text{ and } R_2 \text{ represent any substructures)}$$

THEN predict that the molecule will fragment in the mass spectrometer at either side of the carbon atom, retaining the positive charge on the C=O group.

Corresponding Interpretive Rule for DENDRAL:

IF the mass spectrum shows data points at masses x_1 and x_2 such that the sum of x_1 and x_2 is the molecular weight plus 28 mass units (the overlapping C=O group) and at least one of the two peaks is high (because the fragmentation is favorable)

THEN infer that the molecular structure contains the subgraph

$$\begin{array}{c} \mathrm{O} \\ \| \\ \mathrm{R_1{-}C{-}R_2} \end{array}$$

where the masses of R_1 and R_2 are just $(x_1 - 28)$ and $(x_2 - 28)$.

FIGURE 25-2 Two forms of the same knowledge in DENDRAL.

were painfully aware of how difficult it is for experts to build a single knowledge base capable of supporting high performance in reasoning. Yet there are many related reasoning tasks in any domain for which one knowledge base would be important. We had been troubled, for example, by the fact that DENDRAL's predictive rules of mass spectrometry had to be recast to serve as interpretive rules.[3] Prediction is from cause to effect; interpretation depends on inferences from effects to causes. An example from DENDRAL is shown in Figure 25-2. When we began working on MYCIN, we were thus already sensitized to the issue of avoiding the work of recasting MYCIN's interpretive rules in a form suitable for teaching or other purposes.

The GUIDON program discussed in the next chapter has at least three important facets. First, GUIDON can be seen as an expert system in its own right. Its expertise is in pedagogy, but it obviously needs a knowledge base of medicine to teach from as well as a knowledge base about pedagogy. Second, we had hoped that GUIDON would help us understand the problem of transfer of expertise. We believe there is some symmetry between GUIDON's transferring medical knowledge to a student and an expert's

[3] We experimented with two ways of using predictive rules for interpretation in DENDRAL: (a) generate the interpretive rules automatically from the predictive model (Delfino et al., 1970), and (b) simulate the behavior of a skeletal structure under all plausible substitutions of substructures for the unnamed radicals in order to infer the structure and location of substituents around the skeleton (Smith et al., 1972).

transferring his or her medical knowledge to MYCIN. We need to do much more work here. And third, because professional educators cannot yet provide a firm set of pedagogical rules and heuristics, GUIDON can also be seen as a laboratory for experimenting with alternative teaching strategies. In all three of these areas, the possibilities are exciting because of the newness of the territory and frightening because of the expanse of uncharted waters.