In the early stages of the development of any science different men confronting the same range of phenomena, but not usually all the same particular phenomena, describe and interpret them in different ways. What is surprising, and perhaps also unique in its degree to the fields we call science, is that such initial divergences should ever largely disappear.

T. S. Kuhn, *The Structure of Scientific Revolutions (International Encyclopedia of Unified Science,* vol. II, no. 2). Chicago: University of Chicago Press, 1962.

The philosopher's treatment of a question is like the treatment of an illness.

L. Wittgenstein, *Philosophical Investigations,* para. 255 (trans. G. E. M. Anscombe). New York: Macmillan, 1953.

Every one then who hears these words of mine and does them will be like a wise man who built his house upon the rock; and the rain fell, and the floods came, and the winds blew and beat upon that house, but it did not fall, because it had been founded on the rock. And every one who hears these words of mine and does not do them will be like a foolish man who built his house upon the sand; and the rain fell, and the floods came, and the winds blew and beat against that house, and it fell; and great was the fall of it.

Matthew 7:24–27
(Revised Standard Version)

PART ONE

# Background

# 1

# The Context of the MYCIN Experiments

Artificial Intelligence (AI) is that branch of computer science dealing with symbolic, nonalgorithmic methods of problem solving. Several aspects of this statement are important for understanding MYCIN and the issues discussed in this book. First, most uses of computers over the last 40 years have been in numerical or data-processing applications, but most of a person's knowledge of a subject like medicine is not mathematical or quantitative. It is symbolic knowledge, and it is used in a variety of ways in problem solving. Also, the problem-solving methods themselves are usually not mathematical or data-processing procedures but qualitative reasoning techniques that relate items through judgmental rules, or heuristics, as well as through theoretical laws and definitions. An algorithm is a procedure that is guaranteed either to find the correct solution to a problem in a finite time or to tell you there is no solution. For example, an algorithm for opening a safe with three dials is to set the dials on every combination of numbers and try the lock after each one. Heuristic methods, on the other hand, are not guaranteed to work, but will often find solutions in much shorter times than will exhaustive trial and error or other algorithms. For the example of the safe, one heuristic is to listen for tumblers to drop into place. Few problems in medicine have algorithmic solutions that are both practical and valid. Physicians are forced to reason about an illness using judgmental rules and empirical associations along with definitive truths of physiology.

MYCIN is an expert system (Duda and Shortliffe, 1983). By that we mean that it is an AI program designed (a) to provide expert-level solutions to complex problems, (b) to be understandable, and (c) to be flexible enough to accommodate new knowledge easily. Because we have designed MYCIN to provide advice through a consultative dialogue, we sometimes refer to it as a consultation system.

There are two main parts to an expert system like MYCIN: a knowledge base and an inference mechanism, or engine (Figure 1-1). In addition, there are often subprograms designed to facilitate interaction with users,

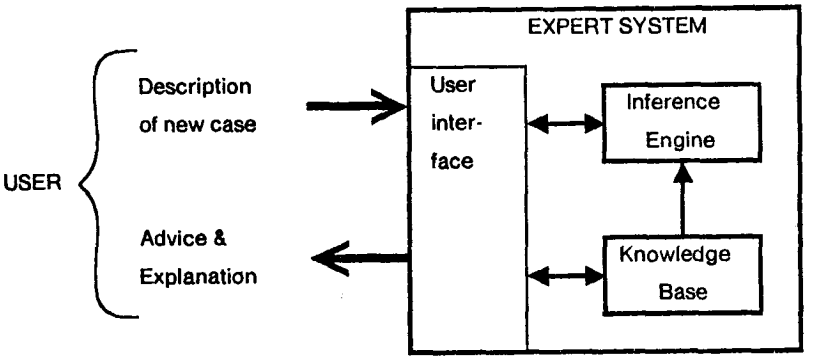


**FIGURE 1-1 Major parts of an expert system. Arrows indicate information flow.**

to help build a knowledge base, to explain a line of reasoning, and so forth.

The knowledge base is the program's store of facts and associations it "knows" about a subject area such as medicine. A critical design decision is how such knowledge is to be represented within the program. There are many choices, in general. For MYCIN, we chose to represent knowledge mostly as conditional statements, or rules, of the following form:

IF: There is evidence that A and B are true,

THEN: Conclude there is evidence that C is true.

This form is often abbreviated to one of the following:

If A and B, then C

A & B → C

We refer to the antecedent of a rule as the premise or left-hand side (LHS) and to the consequent as the action or right-hand side (RHS).

The inference mechanism can take many forms. We often speak of the control structure or control of inference to reflect the fact that there are different controlling strategies for the system. For example, a set of rules may be chained together, as in this example:

| | |
|---|---|
| If A, then B | (Rule 1) |
| If B, then C | (Rule 2) |
| A | (Data) |
| ∴C | (Conclusion) |

This is sometimes called forward chaining, or data-directed inference, because the data that are known (in this case A) drive the inferences from left to right in rules, with rules chaining together to deduce a conclusion (C).

MYCIN primarily uses backward chaining, or a goal-directed control strategy. The deductive validity of the argument is established in the same way, but the system's behavior is quite different. In goal-directed reasoning a system starts with a statement of the goal to achieve and works "backward" through inference rules, i.e., from right to left, to find the data that establish that goal, for example:

| | |
|---|---|
| Find out about C | (Goal) |
| If B, then C | (Rule 1) |
| If A, then B | (Rule 2) |
| ∴If A, then C | (Implicit rule) |
| Question: Is A true? | (Data) |

Since there are many rule chains and many pieces of data about which the system needs to inquire, we sometimes say that MYCIN is an evidence-gathering program.

The whole expert system is used to perform a task, in MYCIN's case to provide diagnostic and therapeutic advice about a patient with an infection as described in Section 1.2. We sometimes refer to the whole system, shown in Figure 1-1, as the *performance system* to contrast it with other subsystems not so directly related to giving advice. MYCIN contains an explanation subsystem, for example, which explains the reasoning of the performance system (see Part Six).

Several of the chapters in this book deal with the problems of constructing a performance system in the first place. We have experimented with different kinds of software tools that aid in the construction of a new system, mostly by helping with the formulation and understanding of a new knowledge base. We refer to the process of mapping an expert's knowledge into a program's knowledge base as *knowledge engineering*.[1] The intended users of these kinds of tools are either (a) the so-called knowledge engineers who help an expert formulate and represent domain-specific knowledge for the performance system or (b) the experts themselves. Al-

[1]The term *knowledge engineering* was, to the best of our knowledge, coined by Edward Feigenbaum after Donald Michie's phrase *epistemological engineering*. Like the phrases *expert system* and *knowledge-based system*, however, it did not come into general use until about 1975. For more discussion of expert systems, see Buchanan and Duda (1983).

though either group might also run the performance system to test it, neither overlaps with the intended routine users of the performance system. Our model is that engineers help experts build a system that others later use to get advice. Elaborating on the previous diagrams, we show this model in Figure 1-2.

### Choice of Programming Language

LISP has been the programming language of choice for AI programs for nearly two decades (McCarthy et al., 1962). It is a symbol manipulation language of extreme flexibility based on a small number of simple constructs.[2] We are often asked why we chose LISP for work on MYCIN, so a brief answer is included here. Above all, we needed a language and programming environment that would allow rapid modification and testing and in which it was easy and natural to separate medical rules in the knowledge base from the inference procedures that *use* the rules. LISP is an interpretive language and thus does not require that programs be recompiled after they have been modified in order to test them. Moreover, LISP removes the distinction between programs and data and thus allows us to *use* rules as parts of the program and to *examine* and *edit* them as data structures. The editing and debugging facilities of Interlisp also aided our research greatly.

Successful AI programs have been written in many languages. Until recently LISP was considered to be too slow and too large for important applications. Thus there were reasons to consider other languages. But for a research effort, such as this one, we were much more concerned with saving days during program development than with saving seconds at run time. We needed the flexibility that LISP offered. When Interlisp became available, we began using it because it promised still more convenience than other versions. Now that additional tools, such as EMYCIN, have been built on top of Interlisp, more savings can be realized by building new systems using those tools (when appropriate) than by building from the base-level LISP system. At the time we began work on MYCIN, however, we had no choice.

## 1.1 Historical Perspective on MYCIN

As best as we can tell, production rules were brought into artificial intelligence (AI) by Allen Newell, who had seen their power and simplicity demonstrated in Robert Floyd's work on formal languages and compilers

[2]See Winston and Horn (1981), Charniak et al. (1980), and Allen (1978) for more information about the language itself.

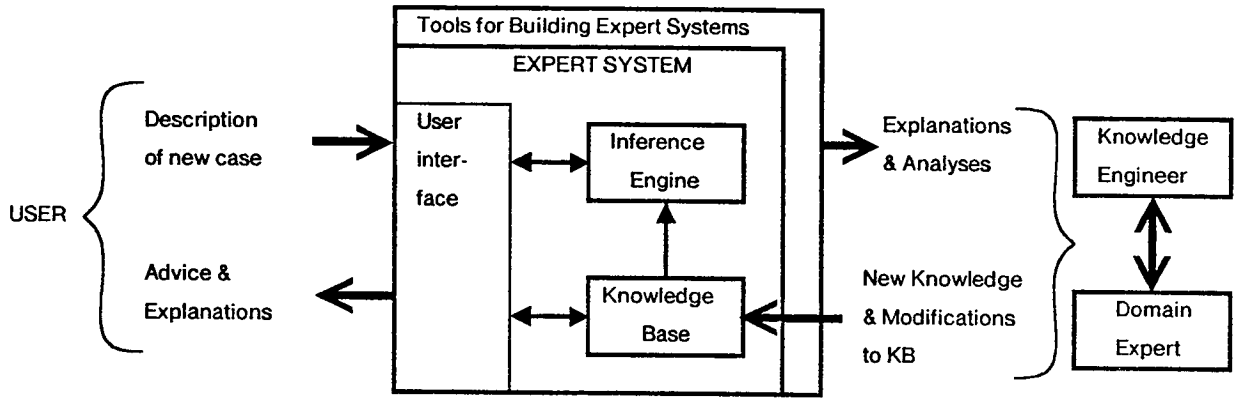


**FIGURE 1-2** Interaction of a knowledge engineer and domain expert with software tools that aid in building an expert system. Arrows indicate information flow.

(Floyd, 1961) at Carnegie-Mellon University. Newell saw in production systems an elegant formalism for psychological modeling, a theme still pursued at Carnegie-Mellon University and elsewhere. Through conversations between Newell and himself at Stanford in the 1960s (see Newell, 1966), Edward Feigenbaum began advocating the use of production rules to encode domain-specific knowledge in DENDRAL. Don Waterman picked up on the suggestion, but decided to work with rules and heuristics of the game of poker (Waterman, 1970) rather than of mass spectrometry. His success, and Feigenbaum's continued advocacy, led to recoding much of DENDRAL's knowledge into rules (Lindsay et al., 1980).

The DENDRAL program was the first AI program to emphasize the power of specialized knowledge over generalized problem-solving methods (see Feigenbaum et al., 1971). It was started in the mid-1960s by Joshua Lederberg and Feigenbaum as an investigation of the use of AI techniques for hypothesis formation. It constructed explanations of empirical data in organic chemistry, specifically, explanations of analytic data about the molecular structure of an unknown organic chemical compound.[3] By the mid-1970s there were several large programs, collectively called DENDRAL, which interacted to help organic chemists elucidate molecular structures. The programs are knowledge-intensive; that is, they require very specialized knowledge of chemistry in order to produce plausible explanations of the data. Thus a major concern in research on DENDRAL was how to represent specialized knowledge of a domain like chemistry so that a computer program could use it for complex problem solving.

MYCIN was an outgrowth of DENDRAL in the sense that many of the lessons learned in the construction of DENDRAL were used in the design and implementation of MYCIN. Foremost among these was the newfound power of production rules, as discussed in Chapter 2. The senior members of the DENDRAL team, Lederberg and Feigenbaum, had convinced themselves and Bruce Buchanan that the AI ideas that made DENDRAL work could be applied to a problem of medical import. At about that time, Edward Shortliffe had just discovered AI as a medical student enrolled in a Computer Science Department course entitled "Models of Thought Processes," taught at the time by Jerome Feldman. Also, Stanley Cohen, then Chief of Clinical Pharmacology at the Stanford University Medical School, had been working on a medical computing project, the MEDIPHOR drug interaction warning system (Cohen et al., 1974). He had sought Buchanan's involvement and had also just accepted Shortliffe as a research assistant on the project. In addition, the late George Forsythe, then Chairman of the Computer Science Department, was strongly supportive of this kind of interdisciplinary research project and encouraged

[3]Even more specifically, the data about the unknown compound were data from a mass spectrometer, an instrument that bombards a small sample of a compound with high-energy electrons and produces data on the resulting fragments.

Shortliffe in his efforts to obtain formal training in the field. Thus the scene was set for a collaborative effort involving Cohen, Buchanan, and Shortliffe—an effort that ultimately grew into Shortliffe's dissertation.

After six months of collaborative effort on MEDIPHOR, our discussions began to focus on a computer program that would monitor physicians' prescriptions for antibiotics and generate warnings on inappropriate prescriptions in the same way that MEDIPHOR produced warnings regarding potential drug-drug interactions. Such a program would have needed to access data bases on three Stanford computers: the pharmacy, clinical laboratory, and bacteriology systems. It would also have required considerable knowledge about the general and specific conditions that make one antibiotic, or combination of antibiotics, a better choice than another. Cohen interested Thomas Merigan, Chief of the Infectious Disease Division at Stanford, in lending both his expertise and that of Stanton Axline, a physician in his division. In discussing this new kind of monitoring system, however, we quickly realized that it would require much more medical knowledge than had been the case for MEDIPHOR. Before a system could monitor for inappropriate therapeutic decisions, it would need to be an "expert" in the field of antimicrobial selection. Thus, with minor modifications for direct data entry from a terminal rather than from patient data bases, a monitoring system could be modified to provide consultations to physicians. Another appeal of focusing on an interactive system was that it provided us with a short-term means to avoid the difficulty of linking three computers together to provide data to a monitoring system. Thus our concept of a computer-based consultant was born, and we began to model MYCIN after infectious disease consultants. This model also conformed with Cohen's strong belief that a computer-based aid for medical decision making should suggest therapy as well as diagnosis.

Shortliffe synthesized medical knowledge from Cohen and Axline and AI ideas from Buchanan and Cordell Green. Green suggested using Interlisp (then known as BBN-LISP), which was running at SRI International (then Stanford Research Institute) but was not yet available at the university. Conversations with him also led to the idea of using Carbonell's program, SCHOLAR (Carbonell, 1970a), as a model for MYCIN. SCHOLAR represented facts about the geography of South America in a large semantic network and answered questions by making inferences over the net. However, this model was not well enough developed for us to see how a long dialogue with a physician could be focused on one line of reasoning at a time. We also found it difficult to construct semantic networks for the ill-structured knowledge of infectious disease. We turned instead to a rule-based approach that Cohen and Axline found easier to understand, particularly because chained rules led to lines of reasoning that they could understand and critique.

One important reason for the success of our early efforts was Shortliffe's ability to provide quickly a working prototype program that would show Cohen and Axline the consequences of the rules they had stated at

each meeting. The modularity of the rules was an important benefit in providing rapid feedback on changes. Focusing early on a working program not only kept the experts interested but also allowed us to design the emerging program in response to real problems instead of trying to imagine the shape of the problems entirely in advance of their manifestations in context.

Green recommended hiring Carli Scott as our first full-time employee, and the MYCIN research began to take shape as a coordinated project. Axline subsequently enlisted help from infectious disease fellows to complement the expertise of Cohen's clinical pharmacology fellow. Graduate students from the Computer Science Department were also attracted to the work, partly because of its social relevance and partly because it was new and exciting. Randall Davis, for example, had been working on vision understanding at the Stanford AI Lab and had been accepted for medical school when he heard about MYCIN and decided to invest his research talents with us.

In our first grant application (October, 1973), we described the goals of the project.

> For the past year and a half the Divisions of Clinical Pharmacology and Infectious Disease plus members of the Department of Computer Science have collaborated on initial development of a computer-based system (termed MYCIN) that will be capable of using both clinical data and judgmental decisions regarding infectious disease therapy. The proposed research involves development and acceptable implementation of the following:
>
> A. CONSULTATION PROGRAM. The central component of the MYCIN system is an interactive computer program to provide physicians with consultative advice regarding an appropriate choice of antimicrobial therapy as determined from data available from the microbiology and clinical chemistry laboratories and from direct clinical observations entered by the physician in response to computer-generated questions;
>
> B. INTERACTIVE EXPLANATION CAPABILITIES. Another important component of the system permits the consultation program to explain its knowledge of infectious disease therapy and to justify specific therapeutic recommendations;
>
> C. COMPUTER ACQUISITION OF JUDGMENTAL KNOWLEDGE. The third aspect of this work seeks to permit experts in the field of infectious disease therapy to teach the MYCIN system the therapeutic decision rules that they find useful in their clinical practice.

The submission of our initial grant application encouraged us to choose a name for the project on which we had already been working for two years. After failing to find a suitable acronym, we selected the name MYCIN at Axline's suggestion. This name is simply the common suffix associated with many antimicrobial agents.

Although we were aiming at a program that would help physicians, we also realized that there were many computer science problems with

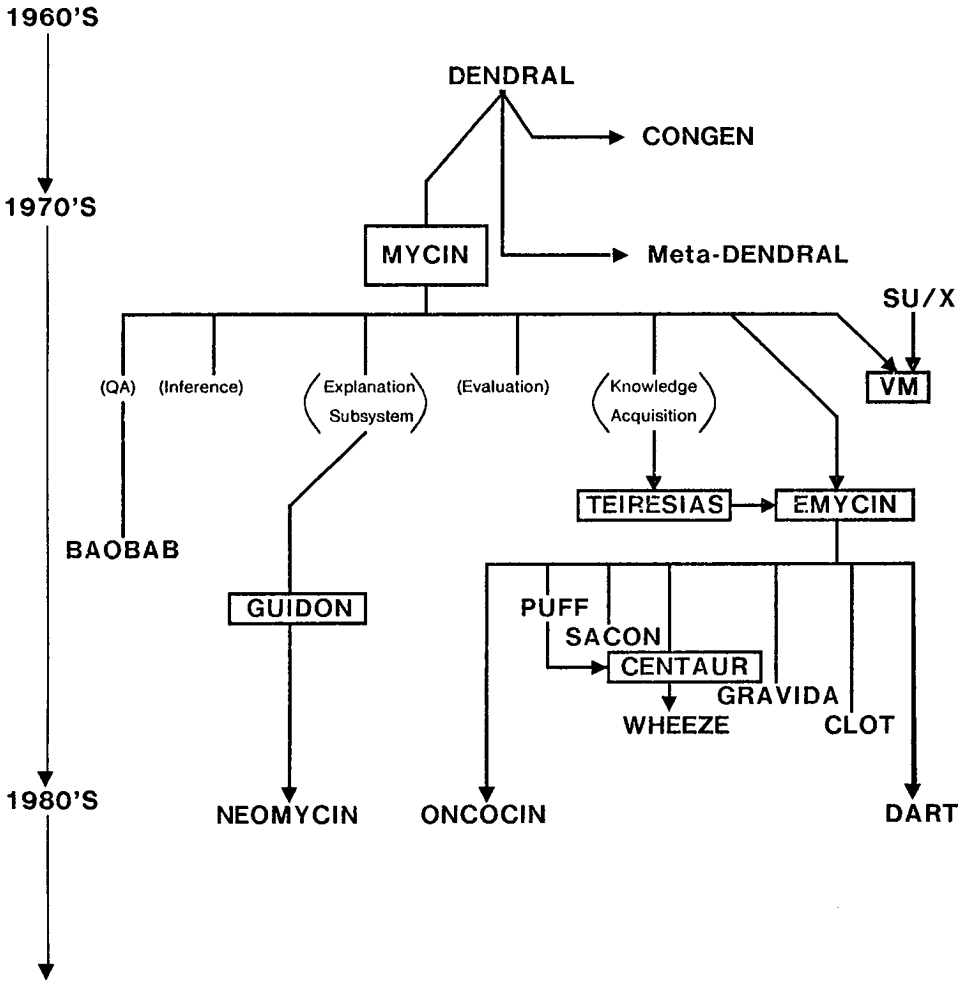


**FIGURE 1-3 HPP programs relating to MYCIN. (Program names in boxes were Ph.D. dissertation research programs.)**

which we had to grapple. No other AI program, including DENDRAL, had been built using so much domain-specific knowledge so clearly separated from the inference procedures.

A schematic review of the history of the work on MYCIN and related projects is shown in Figure 1-3. MYCIN was one of several projects in the Stanford Heuristic Programming Project (HPP); others were DENDRAL, CONGEN, Meta-DENDRAL, and SU/X.[4] There was much interaction

[4]Later renamed HASP/SIAP (Nii and Feigenbaum, 1978; Nii et al., 1982).

among the individuals working in HPP that is not shown in this simplified diagram, of course. Within the MYCIN project individuals were working on several nearly separable subprojects, some of which are shown: Question Answering (QA), Inference (including certainty factors, or CF's, and the therapy recommendation code), Explanation, Evaluation, and Knowledge Acquisition. These subprojects formed the basis of several of the experiments reported in this volume. All were well-focused projects since we were undertaking them partly to improve the knowledge base and the performance of MYCIN. Figure 1-3 shows roughly the chronology of work; however, in the organization of this book chronology is not emphasized.

## Ancient History

Jaynes (1976) refers to a collection of 20,000–30,000 Babylonian tablets, about 20% of which contain sets of production rules ("omens") for governing everyday affairs.[5] These were already written and catalogued by about 650 B.C. He describes the form of each entry as "an if-clause or protasis followed by a then-clause or apodosis." For example,

> "If a horse enters a man's house and bites either an ass or a man,
> the owner of the house will die and his household will be scattered."
>
> "If a man unwittingly treads on a lizard and kills it,
> he will prevail over his adversary."

Included in these are medical rules, correlating symptoms with prognoses. According to one of Jaynes' sources (Wilson, 1956; 1962), these tablets of scientific teachings were catalogued by subject matter around 700 B.C. Among the left-hand sides quoted from the medical tablets are the following (Wilson, 1956):

> "If, after a day's illness, he begins to suffer from headache . . ."
>
> "If, at the onset of his illness, he had prickly heat . . ."
>
> "If he is hot (in one place) and cold (in another) . . ."
>
> "If the affected area is clammy with sweat . . ."

Each clause is catalogued as appearing in 60–150 entries on the tablets. One right-hand side for the medical rules cited by Wilson is the following:

> ". . . he will die suddenly."

[5]We are indebted to James Bennett for pointing out this reference.

Thus we see that large collections of simple rules were used for medical diagnosis long before MYCIN and that some thought had been given to the organization of the knowledge base.[6]

## 1.2 MYCIN's Task Domain—Antimicrobial Selection

Because a basic understanding of MYCIN's task domain is important for understanding much of what follows, we include here a brief description of infectious disease diagnosis and therapy.[7]

### 1.2.1 The Nature of the Decision Problem

An antimicrobial agent is any drug designed to kill bacteria or to arrest their growth. Thus the selection of antimicrobial therapy refers to the problem of choosing an agent (or combination of agents) for use in treating a patient with a bacterial infection. The terms *antimicrobial* and *antibiotic* are often used interchangeably, even though the latter actually refers to any one of a number of drugs that are isolated as naturally occurring products of bacteria or fungi. Thus the well-known penicillin mold is the source of an antibiotic, penicillin, that is used as an antimicrobial. Some antibiotics are too toxic for use in treating infectious diseases but are still used in research laboratories (e.g., dactinomycin) or in cancer chemotherapy (e.g., daunomycin). Furthermore, some antimicrobials (such as the sulfonamides) are synthetic drugs and are therefore not antibiotics. There are also semisynthetic antibiotics (e.g., methicillin) that are produced in chemical laboratories by manipulating a naturally occurring antibiotic molecule. In writing about MYCIN we have tended not to rely on this formal distinction between antimicrobial and antibiotic and have used the terms as though they were synonymous.

Antimicrobial selection would be a trivial problem if there were a single nontoxic agent effective against all bacteria capable of causing human disease. However, drugs that are highly useful against certain organisms are often not the most effective against others. The identity (genus) of the organism causing an infection is therefore an important clue for deciding

[6]The fact that the rules on the tablets were themselves indexed by premise clauses would suggest that they were used in data-directed fashion. Yet the global organization of rules on tablets was by subject matter, so that medical rules were together, house-building rules together, and so on. This "big switch" organization of the knowledge base is an early instance of using rule groups to focus the attention of the problem solver, a pressing problem, especially in large, data-directed systems such as the Babylonian omens.

[7]This section is based on a similar discussion by Shortliffe (1974).

what drugs are apt to be beneficial for the patient. Initially, MYCIN did not consider infections caused by viruses or pathogenic fungi, but since these other kinds of organisms are particularly significant as causes of meningitis, they were later added when we began to work with that domain.

Selection of therapy is a four-part decision process. First, the physician must decide whether or not the patient has a significant infection requiring treatment. If there is significant disease, the organism must be identified or the range of possible identities must be inferred. The third step is to select a set of drugs that may be appropriate. Finally, the most appropriate drug or combination of drugs must be selected from the list of possibilities. Each step in this decision process is described below.

## Is the Infection Significant?

The human body is normally populated by a wide variety of bacteria. Organisms can invariably be cultured from samples taken from a patient's skin, throat, or stool. These normal flora are not associated with disease in most patients and are, in fact, often important to the body's homeostatic balance. The isolation of bacteria from a patient is therefore not presumptive evidence of significant infectious disease.

Another complication is the possibility that samples obtained from normally sterile sites (such as the blood, cerebrospinal fluid, or urinary tract) will be contaminated with external organisms either during the collection process itself or in the microbiology laboratory where the cultures are grown. It is therefore often wise to obtain several samples and to see how many contain organisms that may be associated with significant disease.

Because the patient does have a normal bacterial flora and contamination of cultures may occur, determination of the significance of an infection is usually based on clinical criteria. Does the patient have a fever? Is he or she coughing up sputum filled with bacteria? Does the patient have skin or blood findings suggestive of serious infection? Is his or her chest x-ray normal? Does the patient have pain or inflammation? These and similar questions allow the physician to judge the seriousness of the patient's condition and often demonstrate why the possibility of infection was considered in the first place.

## What Is the Organism's Identity?

There are several laboratory tests that allow an organism to be identified. The physician first obtains a sample from the site of suspected infection (e.g., a blood sample, an aspirate from an abscess, a throat swabbing, or a urine specimen) and sends it to the microbiology laboratory for culture.

There the technicians first attempt to grow organisms from the sample on an appropriate nutritional medium. Early evidence of growth may allow them to report the morphological and staining characteristics of the organism. However, complete testing of the organism to determine a definite identity usually requires 24–48 hours or more.

The problem with this identification process is that the patient may be so ill at the time when the culture is first obtained that the physician cannot wait two days before beginning antimicrobial therapy. Early data regarding the organism's staining characteristics, morphology, growth conformation, and ability to grow with or without oxygen may therefore become crucially important for narrowing down the range of possible identities. Furthermore, historical information about the patient and details regarding his or her clinical status may provide additional useful clues as to the organism's identity.

### What Are the Potentially Useful Drugs?

Even once the identity of an organism is known with certainty, its range of antimicrobial sensitivities may be unknown. For example, although a *Pseudomonas* is usually sensitive to gentamicin, an increasing number of gentamicin-resistant *Pseudomonae* are being isolated. For this reason the microbiology technicians will often run *in vitro* sensitivity tests on an organism they are growing, exposing the bacterium to several commonly used antimicrobial agents. This sensitivity information is reported to the physician so that he or she will know those drugs that are likely to be effective *in vivo* (i.e., in the patient).

Sensitivity data do not become available until one or two days after the culture is obtained, however. The physician must therefore often select a drug on the basis of the list of possible identities plus the antimicrobial agents that are statistically likely to be effective against each of the identities. These statistical data are available from many hospital laboratories (e.g., 82% of *E. coli* isolated at Stanford Hospital are sensitive *in vitro* to gentamicin), although, in practice, physicians seldom use the probabilistic information except in a rather intuitive sense (e.g., "Most of the *E. coli* infections I have treated recently have responded to gentamicin.").

### Which Drug Is Best for This Patient?

Once a list of drugs that may be useful has been considered, the best regimen is selected on the basis of a variety of factors. These include the likelihood that the drug will be effective against the organism, as well as a number of clinical considerations. For example, it is important to know whether or not the patient has any drug allergies and whether or not the drug is contraindicated because of age, sex, or kidney status. If the patient

has meningitis or brain involvement, whether or not the drug crosses the blood-brain barrier is an important question. Since some drugs can be given only orally, intravenously (IV), or intramuscularly (IM), the desired route of administration may become an important consideration. The severity of the patient's disease may also be important, particularly for those drugs whose use is restricted on ecological grounds or which are particularly likely to cause toxic complications. Furthermore, as the patient's clinical status varies over time and more definitive information becomes available from the microbiology laboratory, it may be wise to change the drug of choice or to modify the recommended dosage.

### 1.2.2 Evidence That Assistance Is Needed

The "antimicrobial revolution" began with the introduction of the sulfonamides in the 1930s and penicillin in 1943. The beneficial effects that these and subsequent drugs have had on humanity cannot be overstated. However, as early as the 1950s it became clear that antibiotics were being misused. A study of office practice involving 87 general practitioners (Peterson et al., 1956) revealed that antibiotics were given indiscriminately to all patients with upper respiratory infections by 67% of the physicians, while only 33% ever tried to separate viral from bacterial etiologies. Despite attempts to educate physicians regarding this kind of inappropriate therapy, similar data have continued to be reported (Kunin, 1973).

At the time we began work on MYCIN, antibiotic misuse was receiving wide attention (Scheckler and Bennett, 1970; Roberts and Visconti, 1972; Kunin, 1973; Simmons and Stolley, 1974; Carden, 1974). The studies showed that very few physicians go through the methodical decision process that was described above. In the outpatient environment antibiotics are often prescribed without the physician's having identified or even cultured the offending organism (Kunin, 1973). In 1972 the FDA certified enough (2,400,000 kg) of the commonly used antibiotics to treat two illnesses of average duration in every man, woman, and child in the country. Yet it has been estimated that the average person has an illness requiring antibiotic treatment no more often than once every five to ten years (Kunin, 1973). Part of the reason for such overprescribing is the patient's demand for some kind of prescription with every office visit (Muller, 1972). It is difficult for many physicians to resist such demands; thus improved public education is one step toward lessening the problem.

However, antibiotic use is widespread among hospitalized patients as well. Studies have shown that, on any given day, one-third of the patients in a general hospital are receiving at least one systemic antimicrobial agent (Roberts and Visconti, 1972; Scheckler and Bennett, 1970; Resztak and Williams, 1972). The monetary cost to both patients and hospitals is enormous (Reimann and D'ambola, 1966; Kunin, 1973). Simmons and Stolley (1974) have summarized the issues as follows:

1. Has the wide use of antibiotics led to the emergence of new resistant bacterial strains?
2. Has the ecology of "natural" or "hospital" bacterial flora been shifted because of antibiotic use?
3. Have nosocomial (i.e., hospital-acquired) infections changed in incidence or severity due to antibiotic use?
4. What are the trends of antibiotic use?
5. Are antibiotics properly used in practice?
   - Is there evidence that prophylactic use of antibiotics is harmful, and how common is it?
   - Are antibiotics often prescribed without prior bacterial culture?
   - When cultures are taken, is the appropriate antibiotic usually prescribed and correctly used?
6. Is the increasingly more frequent use of antibiotics presenting the medical community and the public with a new set of hazards that should be approached by some new administrative or educational measures?

Having stated the issues, these authors proceed to cite evidence that indicates that each of these questions has frightening answers—that the effects of antibiotic misuse are so far-reaching that the consequences may often be worse than the disease (real or imagined) being treated!

Our principal concern has been with the fifth question: are physicians rational in their prescribing habits and, if not, why not? Roberts and Visconti examined these issues in 1,035 patients consecutively admitted to a 500-bed community hospital (Roberts and Visconti, 1972). Of 340 patients receiving systemic antimicrobials, only 35% were treated for infection. The rest received either prophylactic therapy (55%) or treatment for symptoms without verified infection (10%). A panel of expert physicians and pharmacists evaluated these therapeutic decisions, and only 13% were judged to be rational, while 66% were assessed as clearly irrational. The remainder were said to be questionable.

Of particular interest were the reasons why therapy was judged to be irrational in those patients for whom some kind of antimicrobial therapy was warranted. This group consisted of 112 patients, or 50.2% of the 223 patients who were treated irrationally. It is instructive to list the reasons that were cited, along with the percentages indicating how many of the 112 patients were involved:

| | |
|---|---|
| Antimicrobial contraindicated in patient | 7.1% |
| Patient allergic | 2.7 |
| Inappropriate sequence of antimicrobials | 26.8 |
| Inappropriate combination of antimicrobials | 24.1 |
| Inappropriate antimicrobial used to treat condition | 62.5 |
| Inappropriate dose | 18.7 |

| | |
|---|---|
| Inappropriate duration of therapy | 9.8 |
| Inappropriate route | 3.6 |
| Culture and sensitivity needed | 17.0 |
| Culture and sensitivity indicate wrong antibiotic being used | 16.1 |

The percentages add up to more than 100% because a given therapy may have been judged inappropriate for more than one reason. Thus 62.5% of the 112 patients who required antimicrobial therapy but were treated irrationally were given a drug that was inappropriate for their clinical condition. This observation reflects the need for improved therapy selection for patients requiring therapy—precisely the decision task that MYCIN was designed to assist.

Once a need for improved continuing medical education in antimicrobial selection was recognized, there were several valid ways to respond. One was to offer appropriate post-graduate courses for physicians. Another was to introduce surveillance systems for the monitoring and approval of antibiotic prescriptions within hospitals (Edwards, 1968; Kunin, 1973). In addition, physicians were encouraged to seek consultations with infectious disease experts when they were uncertain how best to proceed with the treatment of a bacterial infection. Finally, we concluded that an automated consultation system that could substitute for infectious disease experts when they are unavailable or inaccessible could provide a valuable partial solution to the therapy selection problem. MYCIN was conceived and developed in an attempt to fill that need.

## 1.3 Organization of the Book

This volume is organized into twelve parts of two to four chapters, each highlighting a fundamental theme in the development and evolution of MYCIN. This introductory part closes with a classic review paper that outlines the production rule methodology.

The design and implementation of MYCIN are discussed in Part Two. Shortliffe's thesis was the beginning, but the original system he developed was modified as required.

In Part Three we focus on the problems of building a knowledge base and on knowledge acquisition in general. TEIRESIAS, the program resulting from Randy Davis' dissertation research, is described.

In Part Four we address the problems of reasoning under uncertainty. The certainty factor model, one answer to the question of how to propagate uncertainty in an inference mechanism, forms the basis of this part.

Part Five discusses the generality of the MYCIN formalism. The EMYCIN system, written largely by William van Melle as part of his dissertation

work, is a strongly positive answer to the question of whether MYCIN could be generalized.

Work on explanation is reviewed in Part Six. Explanation was a major design requirement from the start, and many persons contributed to MYCIN's explanation capabilities.

In Part Seven we discuss some of the experimentation we were doing with alternative representations. Jan Aikins' thesis work on CENTAUR examined the advantages of combining frames and production rules. Larry Fagan's work on VM examined the augmentations to a production rule system that are needed to reason effectively with data monitored over time.

As an outgrowth of the explanation work, we came to believe that MYCIN had some pedagogical value to students trying to learn about infectious disease diagnosis and therapy. William Clancey took this idea one step further in his research on the GUIDON system, described in Part Eight. GUIDON is an intelligent tutor that we initially believed could tutor students about the contents of any knowledge base for an EMYCIN system. There is now strong evidence that this hypothesis was false because more knowledge is needed for tutoring than for advising.

In Part Nine we discuss the concept of meta-level knowledge, some of which we found to be necessary for intelligent tutoring. We first examined rules of strategy and control, called meta-rules, in the context of the TEIRESIAS program. One working hypothesis was that meta-rules could be encoded as production rules similar to those at the object level (medical rules) and that the same inference and explanation routines could work with them as well.

From the start of the project, we had been concerned about performance evaluation, as described in Part Ten. We undertook three different evaluation experiments, each simpler and more realistic but somewhat more limited than the last.

Another primary design consideration was human engineering, the subject of Part Eleven. We knew that a useful system had to be well enough engineered to make people want to use it; high performance alone was not sufficient. The chapters in this part discuss experiments with both natural language interfaces and customized hardware and system architectures.

Finally, in Part Twelve, we attempt to summarize the lessons about rule-based expert systems that we have learned in nearly a decade of research on the programs named in Figure 1-3. We believe that AI is largely an experimental science in which ideas are tested in working programs. Although there are many experiments we neglected to perform, we believe the descriptions of several that we did undertake will allow others to build on our experience and to compare their results with ours.