Chapter 6

Future Directions for MYCIN

6.1 Introduction

There are several questions regarding MYCIN's performance that are currently unanswered. Most of these involve issues that cannot be adequately analyzed until the program has been introduced for ongoing use in the clinical setting. This chapter introduces our plans for clinical implementation and evaluation of MYCIN. It also discusses some immediate and long range goals for expansion of MYCIN's capabilities.

Section 6.2 discusses the current status of the MYCIN project, the short range goals, and the way in which the research group is currently organized. Section 6.3 then briefly describes one of MYCIN's goals that has already received considerable attention, namely the problem of rule-acquisition (Subprogram 3 shown in Figure 1-1). I first explain the current operation of the Rule-Acquisition System and then proceed to a discussion of what additional capabilities will be needed. I also discuss the way in which MYCIN can automatically identify and correct inconsistencies or contradictions as new knowledge is added to the corpus of system rules, and conclude with an assessment of how a growing rule corpus will affect system performance.

Section 6.4 discusses evaluation questions that must be answered both before and after MYCIN is implemented in the ward setting. The remainder of the chapter then deals with issues that are not immediate concerns but which reveal the potential for eventual wide influence of a program like MYCIN. Section 6.5 deals with how MYCIN could efficiently be implemented as a module in a total

Hospital Information System (HIS) or in any environment where computer-based patient data could be shared. Section 6.6 takes the HIS example one step further, pointing out ways in which MYCIN could be instituted as a nonpunitive peer review mechanism for prospective monitoring of physician prescribing habits. Section 6.7 then discusses the potential educational applications of MYCIN, and I conclude with brief mention of other task domains in which the MYCIN formalisms can perhaps be applied.

6.2 Plans for Immediate Future

The work described in this text has involved the combined efforts of several collaborating physicians and computer scientists. After a two year growing period, during which the program gradually took shape, MYCIN began to interest other individuals who were able to devote time to the project. Research funding also became available and, as a result, MYCIN currently involves the full time efforts of at least eight individuals. This infusion of people with diverse interests, but united by a common fascination with applications of AI in medicine, has enabled MYCIN to begin to expand in a number of new directions. In this section, I shall describe some of these projects.

The primary concern at present is to introduce MYCIN in the clinical setting at Stanford Hospital. This initially involves developing the program's knowledge base for bacteremia until we are convinced that MYCIN does indeed give expert advice for patients with that subset of bacterial infections. Clinical fellows in infectious diseases and clinical pharmacology are currently analyzing MYCIN's rules and exercising the program with actual patient cases in an effort to identify additional rules, both for bacteremia and other infectious disease problems, that will help to improve the program's performance.

Once the knowledge-base is deemed adequate, interactive terminals will be placed on appropriate wards at Stanford Hospital and the affiliated Veterans Administration Hospital in Palo Alto. Since users often need to refer back to parts of a consultation, quiet but fast hard-copy terminals will probably be utilized. After physicians have been educated regarding MYCIN's availability and how it is used, a

formal evaluation of the program's clinical impact and acceptability will be undertaken. Current prescribing habits will be monitored prior to introduction of the program so that valid control data will be available.

Chapters 3 and 5 both closed with discussions of some of the recognized improvements needed for Subprograms 1 and 2. Work on some of these problems is already underway. In particular, one project member is studying the problem of transferring function-based knowledge about drug selection to rules. A second investigator is examining the current design of the Explanation System to see whether the IQ prefix can be dropped from informational questions (§ 5.2.2-1) without introducing so much syntactic or semantic processing that the QA-module becomes unworkably slow.

Finally, one project member is examining several issues related to computer programs that "understand" their own operation. MYCIN provides an interesting practical environment for this kind of theoretical study because its goal-oriented control structure and formalized rules provide generalized data structures that do let the program analyze itself. The WHY option to which we have alluded (§ 3.3.2-2) is the first result of this work, but attention is also being paid to the semantics of certainty factors, rule-acquisition, and problems resulting from the interaction of new rules with a large corpus of pre-existing rules.

6.3 Knowledge Acquisition

Although we have already spent much time studying mechanisms for acquisition of new rules and have also undertaken some preliminary programming, so many of the problems in this domain remain unsolved that we have postponed discussing the current status until this chapter. Rule-acquisition is accomplished via Subprogram 3 (shown in Figure 1-1). As indicated in the figure, this subprogram may be entered from Subprogram 2 if the user is an infectious disease expert who is recognized by the system (see the RA option, § 5.2.3). The expert enters a new rule in English, it is translated into LISP, and the rule is then added to the knowledge base so that it will be available for future consultations.

It might seem reasonable to call rule-acquisition either teaching

(by the expert) or learning (by the machine). Both terms are potentially misleading, however, because "teaching" may lead to confusion with Computer-Aided Instruction (CAI) and "learning" has a rather special meaning in the AI field. When a program "learns," the term usually means that experience has allowed an intelligent program to infer a truth or strategy and to incorporate the fact or heuristic into its knowledge-base. For example, I used the word when describing Waterman's poker program [Waterman, 1970] and its ability to "learn" heuristics. The classic example of a learning program is Samuel's checker-playing system that modifies its evaluation function in response to experience playing the game and has thereby improved so that it regularly beats its creator [Samuel, 1959, 1967]. Winston described a program that "learns" how to identify geometric objects from examples and counterexamples [Winston, 1970].

As currently envisioned, Subprogram 3 differs from these examples of "learning" programs in that it waits to be told what it needs to know. Thus the expert must deduce exactly what information is missing from the system or what previous rule is incorrect. Although the Explanation System simplifies this task, the expert is the primary problem solver for improving MYCIN's knowledge base. Possible mechanisms for changing this emphasis are discussed in § 6.5.

6.3.1 CURRENT STATUS OF RULE-ACQUISITION

The current version of Subprogram 3, although it is limited in usefulness, does serve to demonstrate both the generality of MYCIN's natural language capabilities and a potential methodology for powerful interactive knowledge acquisition. Limited effort has been spent on this capability to date, and the speed with which a mechanism for learning simple rules was developed suggests that more concentrated efforts in this area may well prove fruitful in a relatively short period of time.

6.3.1-1 Subprogram 3

Subprogram 3 allows an expert either to enter a new decision rule or to change a pre-existing rule that is in some way inadequate. Both tasks require similar computer processing, so I shall first discuss

acquisition of new rules and then explain the necessary modifications for altering old rules.

Subprogram 3 acquires new rules using the following ten-step procedure:

(1) Tells expert name of rule he is creating;
(2) Acquires PREMISE conditions one-by-one, translating each from English into the corresponding LISP representation;
(3) Acquires ACTION clauses one-by-one, translating each into its LISP representation and requesting an associated certainty factor (CF) when necessary;
(4) Displays English translation of rule using standard LISP-to-English routines (§ 3.2.7);
(5) Asks user to approve translated version; if rule is not correct, allows him to make changes and then goes back to Step (4);
(6) Searches for contradictions, inconsistencies, or subsumptions involving the new rule and other rules that are already part of the knowledge base; interacts with user as necessary in order to clarify any problems that are noted;
(7) Asks for assistance classifying rule, if necessary (§ 3.2.2-2);
(8) Adds rule to LOOKAHEAD list for all clinical parameters referenced in PREMISE (§ 3.2.3-2);
(9) Adds rule to either CONTAINED-IN list or UPDATED-BY list for all clinical parameters referenced in ACTION (§ 3.2.3-2);
(10) Tells expert that new rule is part of MYCIN's knowledge-base.

Note that Step (9) insures that the rule will automatically be invoked when appropriate because FINDOUT will find the rule on the UPDATED-BY list of the parameter for which the new decision criterion is useful (§ 3.3.2-1). Note also that the current procedure ignores ELSE clauses since these are seldom needed; the ability to add an ELSE clause to a rule may eventually be added.

Section 6.3.1-2 discusses our initial plans regarding the implementation of Step (6). The impact of a new rule on the pre-existing corpus is an important theoretical problem, particularly in light of the complexities introduced by our certainty factor scheme. In Section 6.3.1-3, I close with a discussion of the way a growing rule corpus may affect the speed and efficiency of consultation sessions.

This subsection will first conclude, however, with a sample interactive session illustrating the current capabilities of the Rule-

Acquisition System. As usual, words entered by the physician follow the double asterisk:

```
**NR
The new rule will be called RULE200

     IF:    1 ** ORGANISM IS A GRAM NEGATIVE ROD
        and 2 ** IT IS ANAEROBIC
        and 3 ** IT ENTERED THE BLOOD FROM THE GI TRACT
        and 4 **
     THEN:  1 ** IT IS PROBABLE A BACTEROIDES
                 On a scale of 1 to 10, how much certainty would you
                 affix to this conclusion
                 **9
        and 2**

This is my understanding of your rule:
   RULE200

     IF:    1) THE SITE OF THE CULTURE IS BLOOD, AND
            2) THE SITE OF THE CULTURE IS GI, AND
            3) THE STAIN OF THE ORGANISM IS GRAMNEG, AND
            4) THE MORPHOLOGY OF THE ORGANISM IS ROD, AND
            5) THE AEROBICITY OF THE ORGANISM IS ANAEROBIC
     THEN:  THERE IS STRONGLY SUGGESTIVE EVIDENCE (.9)
            THAT THE IDENTITY OF THE ORGANISM IS
            BACTEROIDES

Okay? (YES or NO)
**NO
```

In this example MYCIN has incorrectly interpreted the expert's statement regarding the gastrointestinal (GI) tract. The physician has therefore indicated that the program's understanding of the rule is not yet correct. The session continues:

```
Change PREMISE? (YES or NO)
**YES
Want to see the current version? (YES or NO)
**NO
Change (C), Delete (D), or Add (A) a component:
**C
Which Component? (Enter number)
**2
```

```
Enter the new component:
   New: **PORTAL OF ENTRY IS THE GI TRACT
More changes to this PREMISE? (YES or NO)
**NO
Change ACTION? (YES or NO)
**NO

This is my understanding of your rule:
   RULE200

      IF:   1) THE SITE OF THE CULTURE IS BLOOD, AND
            2) THE STAIN OF THE ORGANISM IS GRAMNEG, AND
            3) THE MORPHOLOGY OF THE ORGANISM IS ROD, AND
            4) THE AEROBICITY OF THE ORGANISM IS
            ANAEROBIC, AND
            5) THE PORTAL OF ENTRY OF THE ORGANISM IS GI
      THEN: THERE IS STRONGLY SUGGESTIVE EVIDENCE (.9)
            THAT THE IDENTITY OF THE ORGANISM IS
            BACTEROIDES

Okay? (YES or NO)
**YES
```

This time the rule correctly reflects the intention of the expert. Note that the PREMISE conditions have been reordered for efficiency; clauses referencing parameters which are most likely to have been referenced before (and whose values are thus apt to be known) are placed near the beginning of the PREMISE. The algorithm used for reordering is described elsewhere [Shortliffe, 1974b].

Steps (1) through (5) of the acquisition procedure have been completed in the sample session above. Of the remaining steps, only Step (7) currently requires further interaction with the user. MYCIN can easily infer that the new rule is some kind of organism rule, but it is not obvious whether it should be classified as an ORGRULE, a CURORGRULE, or a PRORGRULE (§3.2.2-2). Therefore, MYCIN concludes with the following question:

```
This rule may best be described as a rule which:
   1 - Applies to all organisms
   2 - Applies to prior organisms only
   3 - Applies to current organisms only
**1
```

Thank you for your assistance.
RULE200 is now part of the Consultation System.

RULE200 is thus classfied as an ORGRULE and the rule-acquisition procedure is complete. The rule has been added to the LOOKAHEAD list for SITE, GRAM, MORPH, AIR, and PORTAL and to the UPDATED-BY list for IDENT. Thus the rule will be invoked whenever MYCIN is trying to infer the identity of an organism. The internal representation of the rule created by the above interaction is as follows:

RULE200

```
PREMISE: ($AND  (SAME CNTXT SITE BLOOD)
                (SAME CNTXT GRAM GRAMNEG)
                (SAME CNTXT MORPH ROD)
                (SAME CNTXT AIR ANAEROBIC)
                (SAME CNTXT PORTAL GI))
ACTION:  (CONCLUDE CNTXT IDENT BACTEROIDES TALLY .9)
```

MYCIN's mechanism for changing rules parallels the above procedure, starting at the point where the expert was asked if he wanted to change the PREMISE of RULE200. Thus when the physician indicates that he wants to change a rule, he is asked for the name of the rule requiring alteration and is then permitted to modify only that portion of the rule which is faulty. It is not necessary to delete the erroneous rule and to re-enter it from the beginning as though it were new.

Although we are eager to permit experts to teach the system new rules, there are potential dangers in letting anyone have uncontrolled access to MYCIN's knowledge base. This observation is particularly worrisome while the Step (6) consistency check is in rather rudimentary form. We therefore do not yet automatically store new rules as part of the permanent Consultation System. Instead they are stored temporarily in a file assigned specifically to the expert from whom the rules were acquired. Whenever that expert uses the system he may load his personal rules and they are temporarily added to MYCIN's knowledge base. MYCIN project members have an opportunity to examine both the new rules and the English text from

which they were derived, however, before the new knowledge is transferred from the expert's personal file to the permanent Consultation System.

6.3.1-2 (*) Interaction of New and Old Rules

Step (6) of the rule-acquisition procedure (§ 6.3.1-1) requires a screening process to see if the new rule improperly interacts with other rules in the knowledge-base. Although we have given considerable thought to this problem, Subprogram 3 does not yet undertake this consistency check. Programs to accomplish the necessary screening will be written in the near future, however, and I present here some preliminary observations.

Subsumption: I mentioned the problem of subsumption several times in Chapter 4. Of all the aberrant interactions of new rules with the pre-existing corpus, subsumption is perhaps the easiest to handle in an automated fashion. Suppose, for example, there were already a rule in the corpus as follows (see § 3.2.4 for an explanation of the notation):

[a] A & B & C –x→ D

If an expert now entered the following new rule, a problem of subsumption would arise:

[b] A & B & C & E & F –y→ D

Clearly any time rule [b] is satisfied, rule [a] will also be satisfied since the PREMISE of [b] subsumes the PREMISE of [a]. Yet rule [a] adds nothing to [b] and it would be improper to use both rules in the same context. On the other hand, eliminating [a] is not an adequate solution because [a] may apply in contexts where [b] does not and in those cases the knowledge inherent in [a] is needed.

The solution to the problem is to modify [a] so that it is no longer subsumed by [b] but so that it still will hold for all contexts that fail for [b] but would succeed for [a]. Namely, we propose replacing [a] with a new rule [c]:

[c] A & B & C & (not.E or not.F) –z→ D

Now any context that would have satisfied [a] will succeed either for [b] or [c] but not for both. Negation here implies a predicate's complement over the certainty factor range. Thus not.SAME is NOTSAME, not THOUGHTNOT (§ 3.2.5). The transition from [a] to [c] may be accomplished automatically except for the possible change in CF (from x to z). We therefore propose displaying [c] for the expert and asking for the CF he would assign.

Finding old rules such as [a] that are subsumed by new rules such as [b] does not require a search through the entire rule corpus. MYCIN merely uses the LOOKAHEAD and UPDATED-BY lists for the clinical parameters in [b] to find rules that use all or some of the same parameters to deduce values of the same parameter. These rules may then be checked for subsumption.

Single-Rule Contradictions: It is also easy to find single rule contradictions using LOOKAHEAD and UPDATED-BY lists. However, the discrepancies cannot be handled in an automated fashion and the inconsistencies must be "discussed" with the user. Two rules contradict each other if they use the same conditions to reach the same conclusion but with different certainty factors. Clearly the extreme case occurs when one CF is positive and the other is negative; in such instances the experts disagree not only on the degree of evidence but also on the direction of evidence! Although such contradictions have not yet arisen during the development of MYCIN, Subprogram 3 must be prepared to identify and handle such problems if they do occur. Hopefully the expert will usually suggest a compromise CF that is also acceptable to the expert from whom the old rule was acquired. Expert clinicians often disagree on clinical questions, however, and we must be willing to accept this fact during the design of MYCIN's knowledge acquisition capabilities. If no acceptable compromise can be found, it may be necessary to store both of the rules and later to ask the user whether he wishes advice based upon the rules acquired from Dr. X or those from Dr. Y. This solution does not seem unreasonable since physicians commonly do have to choose among consultants.

Multiple-Rule Contradictions: The most complex interactions between a new rule and the pre-existing corpus occur when the new rule is inconsistent not with a single old rule but with a reasoning chain of old rules. Not only are such inconsistencies difficult to find, but it is also difficult to judge the severity of contradictions because of the interaction of reasoning chains with the CF's of the compo-

nent rules. In fact, unless the new rule has CF=1 or the reasoning chain is comprised only of rules with CF=1 (a situation for which the descriptive term "unity path" has been coined), it may perhaps be argued that no true contradiction exists. We are currently examining the nature of such inconsistencies in order to decide both how to find them using an automated mechanism and also under what conditions they may be ignored.

6.3.1-4 Impact of Knowledge Growth on System Performance

A question we are often asked is whether rule-acquisition will lead to an exponential growth problem. If each new rule permitted an entire new pathway to sprout in the reasoning network below it, we would have to expect exponential growth of search time as the number of rules increased. Indeed, if each new rule referenced several clinical parameters with which the system was not already familiar, and if each of these attributes in turn required a series of rules for use in inferring its value, both the size of the network and the time required for a consultation would grow unmanageably large. Our experience has so far indicated, however, that most new rules reference only the 65 clinical parameters with which the system is already familiar. Since each of these attributes is traced by the FINDOUT mechanism at most once during a consultation session, a new rule referencing parameters already traced for other reasons will generate no additional search time (except for that required to evaluate the single rule itself). Thus, growth in the size of the reasoning network and in search time is at most linear for a new rule that references only clinical parameters that are already recognized and traced by MYCIN. Furthermore, the new rule will have no effect whatsoever on search time in consultations where it is not invoked by the dynamic FINDOUT mechanism. Since we expect that the number of clinical parameters will not increase in proportion to the number of rules, we do not anticipate exponential growth problems.

6.3.2 FUTURE EXTENSIONS

The current rule-acquisition mechanism is limited in scope and applicability for a variety of reasons. Although the current approach may perhaps be adjusted so that it will accept all well-formed rules referencing clinical parameters known to the system, its dependence

upon MYCIN's dictionary (see Chapter 8 in [Shortliffe, 1974b]) results in an inability to learn rules that relate to unfamiliar concepts. Suppose, for example, that an expert wishes to enter a rule in which a condition involves a clinical parameter that has not been used by any existing rule. The system will miss the relevant vocabulary clues since there will be no dictionary entries related to that parameter.

Furthermore, suppose the new parameter refers to a context-type that has never before been created. For example, the length of time a Foley catheter has been in place might be a new parameter (TIME-IN-PLACE) for a new kind of context (CATHETER or PROCEDURE). Learning about new kinds of nodes for the context tree is a problem for which the current approach is clearly ill-equipped.

Finally, the most serious problem arises if the expert references a predicate (function) which has not previously been programmed. An attempt to automate this kind of rule-acquisition will perhaps be dependent upon the AI field known as automatic programming.

Future extensions to Subprogram 3 will therefore attempt to handle unrecognized parameters, contexts, and functions. We have not yet defined how best to approach these problems. We are currently relying on a variety of specialized "service" functions for defining new parameters explicitly and for easily entering and editing LISP versions of rules when Subprogram 3 has failed to acquire their English versions correctly.

In closing this section, I should also mention an entirely different approach to knowledge acquisition that was recently suggested. MYCIN could ask an expert how to treat a specific patient and, if the advice differed from that which the program would have recommended, could seek explanations from the expert. For example:

MYCIN: What would you prescribe?
EXPERT: I would give X.
M: Oh? Why would you give X?
E: It is the first choice drug for Y.
M: True. But how did you know the organism was Y?
E: Because its PARAM was Z.
M: I see. But how did you know its PARAM was Z?
.
.
.

The natural language problems inherent in this approach appear to be somewhat formidable, but the idea is rather enticing and may appeal to researchers with an interest in the theoretical problems underlying this kind of AI task.

6.4 Evaluation of MYCIN

In Chapter 2, I pointed out that the primary design consideration for MYCIN was that it be useful. The other five acceptability criteria discussed in that chapter (i.e., an educational capability, explanation, natural language understanding, knowledge acquisition, and modularity of knowledge) were justified in terms of their ability to enhance the system's usefulness. It was also explained that a consultation program's usefulness can be measured along three scales:

(1) the need for the assistance which the program provides;
(2) the reliability of the advice;
(3) useability; i.e., the mechanics for accessing the machine and retrieving the desired information.

Evaluating the MYCIN system thus requires an assessment of the program's performance along all three of these dimensions.

Section 1.4.2 addressed itself to the first of these three usefulness scales. There is ample evidence that antimicrobial agents are misused and that physicians would benefit from a mechanism that could improve the basis for antimicrobial therapy selection. An implied second component to this question is whether MYCIN is actually able to encourage more rational antimicrobial prescribing habits. Clearly this question cannot be answered until the program has been implemented for ongoing use in the clinical setting.

The reliability of MYCIN's advice, on the other hand, can be assessed even before the program becomes generally available. In fact, as has previously been stated, we do not plan to implement MYCIN on the hospital wards until we are convinced that the program does give reliable advice for patients with bacteremia. We have therefore devised an experimental method for judging the system's validity and have undertaken preliminary studies using this technique. Section 6.4.1 describes the first such study.

The third usefulness scale (i.e., the system's useability) has been

considered throughout MYCIN's development. The success of attempts to make the program easy-to-use cannot be rigorously evaluated, however, until MYCIN is generally available. At that time we will be able to talk to physicians who have interacted with the system and to compile data indicating whether they consult MYCIN regularly or lose interest after one or two encounters.

Evaluation of MYCIN will therefore be a continuing process occurring in stages. The first phase involves validation of the program's advice and will thus predate implementation. Subsequent stages will assess acceptability, clinical impact, and other questions that can be adequately answered only after MYCIN is generally available. In § 6.4.2 through § 6.4.6, I discuss some of these questions and our plans for analyzing them.

6.4.1 RELIABILITY OF MYCIN'S ADVICE

It is our belief that MYCIN should not be made available clinically until its advice has been shown to be reliable; a program that demonstrates its usefulness from the outset is more apt to gain the acceptance of the physicians for whom it is designed. Evaluation of MYCIN's reliability thus requires a pre-implementation testing procedure with which we are currently involved. A semiformal study was undertaken in mid-1974 [Shortliffe, 1974b] when the program appeared to be performing well on the basis of rules provided by the clinical collaborators. Five infectious disease experts (not associated with the project) were asked to evaluate the program's performance with fifteen cases of bacteremia selected without bias from current inpatients. The program's principal problem was found to be a lack of sufficient rules to allow it to consider adequately the severity of a patient's illness. Nevertheless, the experts approved MYCIN's therapy recommendation in 72% of the 75 evaluations (5 experts reviewing 15 patients each). Differences of opinion regarding the best therapy as judged by the experts themselves were considerable however, and this lack of unaminity introduces some challenging problems to an evaluation study of this kind.

Since the experts have helped to identify errors and gaps in the program's knowledge-base, additional rules are currently being added. This process of formal evaluation followed by new rule addition and re-evaluation will continue until the independent experts agree with MYCIN's reasoning as much as they agree with each

other's. We expect that the program will be ready for introduction on the wards once the experts approve its performance in at least 90% of cases.

6.4.2 MYCIN'S ACCEPTABILITY TO PHYSICIANS

Unless MYCIN is accepted by the physicians who must use it, its ability to give valid advice will be of little value. Chapter 2 emphasized those features of the system that were designed to heighten its acceptability. Once the system is generally available, however, new requirements may become evident. We must therefore implement ongoing mechanisms for identifying those aspects of the program that interfere with the willingness of physicians to use it.

One approach will be to keep a record of physicians who have tried the system and to interview them in order to assess their reactions. It is inevitable that MYCIN's interactive capabilities will have to be constantly modified and improved as feedback from physician users is obtained. Another feedback mechanism will be to permit the physician to type in comments at any time during a consultation. Such remarks can be stored in the computer and regularly reviewed by MYCIN project members.

A second tactic is to identify those physicians who have never tried the system and to find out why. If they are simply unaware of the program's existence, that failing can be easily rectified by an appropriate publicity campaign. If their failure to consult MYCIN results from a basic aversion to interacting with a computer, on the other hand, or if they have heard negative comments about the program from their colleagues, it is important to determine whether changes in the system or its mode of interaction will help to make it more attractive. Although physicians have been involved in the design of MYCIN from the outset, it is unlikely that all the concerns of potential users will have been taken into account. We must therefore be prepared to modify the program, or even radically to overhaul it, in an effort to maximize MYCIN's use by those physicians who may need it when they prescribe for an infectious disease problem.

6.4.3 MYCIN'S IMPACT ON PRESCRIBING HABITS

A second important set of questions to be answered once the system is implemented involves its effect on physician prescribing

habits. This can be adequately assessed only if control data regarding current prescribing practices are obtained *before* MYCIN becomes available. It will then be possible to judge whether antimicrobials are used more appropriately after MYCIN has begun to exert its influence.

It will also be important to assess whether physicians who use MYCIN actually follow its advice. When they do not, we should find out why since that may help with the specification of missing decision rules. If, on the other hand, they reject MYCIN's advice and prescribe less appropriately, an attempt must be made to understand why MYCIN failed to influence them. For example, there may be problems with the Explanation System that prevent it from convincing the user that the program's reasoning is valid.

The educational impact of MYCIN can also be judged by monitoring prescribing habits before and after the system is available. It is possible that MYCIN will result in a new awareness of antibiotic prescribing habits throughout the hospital staff so that even physicians who have never used the program will prescribe more appropriately. Furthermore, clinicians who use the program extensively at first may grow to depend upon it less as they become more familiar with the important therapeutic considerations.

6.4.4 MYCIN'S IMPACT ON PATIENT CARE

Influencing physician prescribing habits is not a sufficient goal for MYCIN unless it also has demonstrably beneficial effects upon patient care. It will therefore be necessary to develop mechanisms for measuring MYCIN's effect on the quality of care for patients with bacterial infections.

A number of approaches are possible. One is merely to monitor the response of a patient's disease when he is treated with the regimen suggested by MYCIN. Not only may such monitoring provide evidence that MYCIN is suggesting appropriate therapy but, in cases where the patient does not respond as desired, it may also help identify inadequacies in the decision rules that have been given to MYCIN by experts.

Monitoring individual patients provides information that is more anecdotal than statistically significant, however. It may therefore be wise to gather data reflecting trends in length-of-stay for hospitalized

patients, incidence of adverse reactions to antimicrobial agents, or pharmacy costs to the patient. All these parameters may reflect beneficial effects of MYCIN that can be verified statistically.

6.4.5 (*) SPEED, EFFICIENCY, AND STORAGE REQUIREMENTS

Descriptions of MYCIN often lead to questions regarding the potential difficulty in implementing a completed system without the program proving too large and slow. The final answers concerning these issues will not be available until we get a better feel for how many new rules and system changes will be necessary before MYCIN can become an effective and acceptable clinical tool. We have devoted considerable thought and discussion, however, to the running time and storage requirements of a high performance consultation program such as the one we hope MYCIN will eventually become. Although economic considerations may eventually require that the program be translated for use on a small computer (see § 6.4.6), we are convinced that response time or computer storage limitations are unlikely to present difficulties in implementing a completed version of MYCIN under the present TENEX operating system [Myer, 1971]. Some of the considerations involved in this conclusion are:

Space: The TENEX system that we currently use allocates up to 256 thousand virtual words of memory (512 pages) to each user. Of the 490 pages that we currently use, approximately 320 pages are used by the INTERLISP system, which includes such features as the spelling corrector, CLISP (Conversational LISP), and the LISP compiler. Of the remaining 170 pages, approximately 100 pages (50 K) are for the compiled MYCIN program. The other 70 pages contain MYCIN's rules, clinical parameters, knowledge tables, and working space. The current program appears to operate adequately within these space limitations. As noted in § 1.6.2, that lengthy sample consultation required approximately 20 minutes at a computer terminal, including the time devoted to the optional use of the Explanation System. Moreover, the following options are available to accommodate future growth of the system:

(1) Smaller LISP: Many INTERLISP features are useful for developing a new program but are not essential for running a performance system. For

example, the LISP compiler, LISP editor, and CLISP are all unnecessary for MYCIN's purposes. In response to the demand of many INTERLISP users that the language dispense with certain features in return for increased memory availability, the language will soon have an "overlay" feature that will permit INTERLISP users to customize versions of LISP in accordance with their individual requirements. When implemented, the "overlay" capability will permit us to create a much smaller version of LISP containing only those features needed by MYCIN.

(2) Modular Programs: The three major components of the MYCIN system (Subprograms 1, 2, and 3; Figure 1-1) are currently loaded into core for every run of the program. However, this is not necessary. For a consultation session only Subprogram 1 needs to be used. At the end of an advice-giving session (or in response to the QA command, § 3.3.2-2), the Explanation System can be added to the Consultation System. The Rule-Acquisition System will not be used at all during standard consultations. Since Subprogram 3 depends upon the expert being able to run Subprogram 1 and 2 as well, however, space considerations may be most important during rule-acquisition sessions. The "overlay" feature mentioned above should alleviate some of these space problems by permitting the three subprograms to be loaded when needed and then deleted programmatically.

(3) The Rule Corpus: By far the fastest growing part of the system is the rule corpus. Although the rest of MYCIN is continually being modified, its size has not increased substantially for several months. Relative to the rest of the program, MYCIN's 200 current rules take up only a small amount of space (16 pages = 8 K). Thus, we believe that the system can easily accommodate the many additional rules which we recognize will be needed.

(4) Recoding For Efficiency: In the initial stages of this work, less attention was paid to space considerations than to major design considerations. As we proceed further with development of the program, we expect to be able to recode parts to enable them to make more efficient use of working space and to take up less space themselves.

Running Time: Because MYCIN requires substantial interaction at the terminal, it is, to a large extent, input–output bound. However, at times the system becomes compute bound, such as when it must chain through a large number of rules that do not generate questions, or when it is garbage collecting the working space. Except for a few lapses during these compute bound activities, the program's running time is currently acceptable. We are therefore developing ways to

further optimize our rule searching strategy (§ 3.7) and to reuse active core locations so that fewer garbage collections will occur.

The number of users in a time sharing environment is also a major consideration. To alleviate this potential problem once MYCIN is implemented on the wards, it is possible to arrange for changes to the scheduling algorithm during periods of peak use, and we can at least alert the physician to a potential slow-down when the number of other users is large. It is also worth noting that the times when consultants in infectious disease therapy are least apt to be available (i.e., late at night and on weekends) are precisely those periods when time-sharing systems are most apt to have a low number of users. Thus, the system becomes a particularly viable alternative to the human consultant when he is unavailable.

Since the efficiency of MYCIN is another important consideration, we have accomplished a substantial improvement in execution time by compiling our code for service use of the program. The INTERLISP block compiler may appropriately be used for portions of the code and will give us extra efficiency not attainable by compiling each function individually.

We believe that the present organization of the knowledge base makes for efficient processing of the set of rules. When the number of rules increases substantially, we expect that the present organization will continue to cope successfully for three reasons. First, the rules are divided by context-type so that potentially useful rules are eliminated from consideration if their classification is inappropriate for the context being examined. Second, the rules are linked together in such a way that determining the truth or falsity of the PREMISE of one rule does not require a search of all other rules. Finally, since we have devised a strategy for recognizing those branches of the reasoning network that have already been searched, new rules that reference clinical parameters with which the system is already familiar will not result in exponential growth of the search space.

6.4.6 COST OF MYCIN'S CONSULTATIONS

An important topic that has previously been ignored in this volume is the cost of a system like MYCIN. The present system was developed on a large computer (Digital Equipment Corporation

PDP-10) that is seldom found in hospitals. Furthermore, the operating system and the INTERLISP language [Teitelman, 1974] are designed primarily for AI applications and are therefore mostly found in university or government research environments. Before MYCIN can become generally available outside the university environment, therefore, it will probably need to be rewritten for a computing system that is more accessible to those hospitals most in need of the program's services. As a result, any attempt to evaluate the cost of a consultation with MYCIN would be premature at present. Research and development expenses naturally bear little resemblance to the costs that will be incurred once MYCIN is an ongoing service system on an in-hospital computer. INTERLISP has been a powerful development tool, but it is slow and demands more computing power than most hospitals can afford.

6.5 MYCIN and Shared Data Bases

Section 1.2.2-6 described Hospital Information Systems (HIS) and their potential for assisting with information handling chores in the clinical environment. Regardless of whether such systems are implemented as a single large installation, or as a set of integrated but independently developed submodules, they are characterized by large amounts of diverse patient data that can be shared among the system components.

Let us consider what MYCIN's role might be in an HIS that contains up-to-date patient information in the following categories:

(1) chemistry laboratory data (including hematology)
(2) pharmacy data
(3) microbiology laboratory data
(4) clinical data traditionally found in the patient chart

It should be clear that most of the clinical parameters used by MYCIN may be classified in one of these categories. Thus, if MYCIN were a component in a comprehensive HIS and could reference the patient's information from the above four data bases, several of the questions currently asked of the physician would no longer be necessary. For example, information regarding current and prior

cultures would be available from data base (3) and the patient's recent drug history could be found in data base (2). In fact, any piece of information currently classified as LABDATA (§ 3.2.3-2) would presumably be available from one of the four data bases. The user would therefore be asked to interact with MYCIN only for consideration of those non-LABDATA parameters for which the rule corpus was unable to infer values (Figure 3-8). This corresponds to the observation that ASK1 questions would no longer be necessary and that only ASK2 questions would need to be displayed for the physician (see Figure 3-9).

As was pointed out in § 3.3.2-1, however, one of the goals in the future development of MYCIN's knowledge-base is to acquire enough rules allowing the values of non-LABDATA parameters to be inferred so that ASK2 questions need no longer occur. One of the impediments to this goal has been the tendency for such rules to generate large numbers of highly specific questions that make MYCIN appear to be groping for ideas and that are thus annoying for the user. Consider, for example, the non-LABDATA parameter COMPROMISED that is a "yes–no" parameter indicating whether the patient is a compromised host. There are currently no rules for inferring the value of this parameter, so an ASK2 question is generated whenever FINDOUT tries to find its value (Figure 3-8). If MYCIN were to make the conclusion on its own, rather than to leave the decision up to the judgment of the user, the program would require a series of rules itemizing disease categories that suggest that a patient's immune response system is not functioning normally. Such rules would in turn generate a series of apparently groping questions such as "Does the patient have leukemia?," "Is the patient an alcoholic?," etc. If a series of questions regarding diagnoses could be answered via queries sent to other HIS data bases, however, the more basic rules regarding compromised-host status could be added to MYCIN's knowledge base without generating annoying questions for the physician.

The discussion of the preceding paragraphs indicates the way in which access to shared clinical data bases could reduce the number of questions asked of the physician by MYCIN. Since much of MYCIN's current time requirement is bound by the terminal-based interaction with the physician, an efficient linkage between MYCIN and other data bases might well decrease the time from sign-on until MYCIN's

recommendation becomes available. In the extreme case, one can imagine a user simply giving MYCIN the name of his patient and answering no additional questions. MYCIN would evaluate the patient on the basis of primitive data (LABDATA) obtainable directly from the clinical laboratory, microbiology, pharmacy, and medical record data bases. After a variable length of time (depending upon the complexity of the patient's infectious disease problem), a therapeutic recommendation would be printed by MYCIN and the physician would be able to use the Explanation System (Chapter 5) to query the program regarding the reasoning behind the suggested regimen.

A formally constituted HIS is not a prerequisite for the shared data base application of MYCIN just described. All that is really necessary is the up-to-date data bases plus communication links between the computers in which the information is stored. Stanford Hospital already has all four of the required data bases: pharmacy [Cohen, 1974], microbiology [Petralli, 1970], clinical chemistry [Sussman, unpublished], and medical records [Fries, 1972]. Unfortunately, all four systems were developed independently and currently operate on separate computers. Since all the programs would benefit from access to one another's patient data, however, communication links between the machines are being contemplated. As soon as these are available, we hope to connect MYCIN to the network and to develop the mechanisms for direct access to patient data in accordance with the model that I described above.

If the four clinical data bases are effectively linked, as is planned, another potential addition to MYCIN would be an ability to monitor a patient's response to the recommended therapy. In this way, it could perhaps acquire statistics that would enable it to alter its drug selection strategy or first-choice drugs. If this capability were implemented, it would resemble the kind of machine "learning" discussed at the beginning of § 6.3.

6.6 Prospective Monitoring of Prescribing Habits

Of all the issues currently involving American organized medicine, there is perhaps none more emotion-laden than the question of peer review. Known euphemistically as patient care appraisal,

quality-of-care assessment, or quality assurance, peer review has entered the political arena since a Social Security amendment was signed into law in 1972. Known as Public Law 92-603, the legislation requires that Professional Standards Review Organizations (PSRO) be set up to monitor medical practice, to identify problems, and to take steps to correct them. PSRO's are to be instituted locally in all parts of the country, and physician organizations initially have priority in establishing them.

Although physicians had begun to participate in peer review activities prior to passage of the new legislation, until recently emphasis has been on assessing those parameters of practice that are most easily measured. Thus utilization review committees and tissue review boards have traditionally taken on the primary peer review responsibilities. PL 92-603 has sparked new interest in peer review issues, however, both with regard to how review should be undertaken and whether the government should be able to interfere in an area that had previously been the concern solely of the medical practitioners themselves. Organized medicine has many reservations regarding PSRO [Watts, 1973], and the strengths and weaknesses of the legislation have been much analyzed [Welch, 1973].

As mentioned in § 1.4.6, it is my conviction that the primary reasons for physician opposition to peer review legislation result from the fact that medicine is one of the few professions in which individuals have traditionally been free from close observation and criticism. Legislation to promote government influence on medical care delivery, whether it be Medicare or PSRO, is thus met with widespread opposition and, in some cases, fear [Gottesman, 1972]. What is particularly worrisome to physicians is the potential for being punished when they make decisions that are judged by others to be mistakes.

Regardless of whether PSRO deserves opposition, the bill has been signed into law and is not apt to be repealed. It is therefore appropriate to look for ways to insure that the new peer review mechanisms will both accomplish the goals of the legislation and will be at least mildly acceptable to physicians. I therefore cite the following proposed criteria for acceptability of the developing peer review mechanisms:

(1) They should be able to judge questions of medical care, not merely parameters such as length-of-stay data;

(2) They should emphasize educational benefits rather than punitive actions when errors are noted;
(3) They should ideally inform the physician of a possible error before it is too late to rectify matters;
(4) They should encourage feedback from physicians regarding strengths and weaknesses of the approach.

The importance of the second point cannot be overstated. There has already been experience to indicate that patient care monitoring can be made acceptable to physicians if they are not led to believe that they will be punished when errors are observed [Alper, 1974].

With criteria such as those above in mind, authors have begun to suggest ways to choose peer review methods [Brook, 1973]. For several years there have been efforts to assess quality of care by reviewing patient charts [Fessel, 1972]. Medical audit of this variety is difficult, however, because the task is arduous, it requires a time-commitment from the reviewing physicians, and the criteria for judging care are, in general, ill-defined. One innovation has been the institution of departmental medical audit workshops at which physicians attempt to delineate what should be the criteria for quality care at their hospital. These criteria can then be used for assessment of care when medical records are reviewed.

The above discussion has been an attempt to lay the groundwork for justifying the claim that MYCIN provides a useful model for a peer review mechanism satisfying the cited acceptability criteria. I shall explain this model by describing an existing system and discussing how MYCIN could be adapted in a similar way.

The MEDIPHOR System [Cohen, 1974] was briefly mentioned in § 1.2.2-5. This is a large computer program developed at Stanford Medical School for the prospective control and study of drug interactions in hospitalized patients. Using a comprehensive and documented data base of drug interaction information, the system generates warnings to pharmacists, nursing personnel, and physicians when potentially interacting drug combinations have been prescribed. Drug profiles for patients are available to the system because it also serves as a label printing machine in the hospital pharmacy. Whenever a label is printed, the computer records the information in the patient's drug profile. Thus, whenever a new drug is prescribed, the machine can use its drug interaction data base to search for interactions between the new prescription and drugs the patient is

already receiving. If a potential interaction is found, a warning is printed in the pharmacy and sent to the ward along with the drug. There the physician and nursing staff may consider the interaction information before the interacting drug is administered. If the physician decides to give the drug, he at least knows about the potential for adverse effects and is therefore careful to monitor the appropriate clinical parameters of the patient.

The MEDIPHOR system offers many of the advantages of the peer review acceptability criteria I described. Clearly it addresses itself to an important clinical practice question that is difficult to assess even by chart review. Furthermore, it points out potential problems before they occur and thus reveals its educational emphasis. Physicians are more apt to be defensive about their decisions if possible errors are not pointed out until two or three months after the incident. By that time, notification appears to be a scolding since it is too late for corrective action to be taken. Finally, a system like MEDIPHOR can also be used to accumulate the data necessary for judging trends in the quality of care, at least for the topic of drug interactions.

Suppose, now, that the various computer-based data banks at Stanford Hospital were joined by communication links as discussed in § 6.5. In that section I explained how MYCIN could provide consultations without asking questions of the physician so long as all pertinent data were available in one of the Stanford data bases. Under those circumstances, the physician seeking advice is needed only to initiate the consultation. Consider, then, the potential for initiating the consultation program not in response to a request from a physician seeking advice but instead whenever an antimicrobial agent is prescribed in the hospital pharmacy. The MEDIPHOR system could notify MYCIN regarding the patient, drug, and dose. MYCIN could then use its knowledge base to decide how it would treat the patient and whether the drug actually prescribed is appropriate. If a prescription were clearly inappropriate, MYCIN could send the relevant information back to MEDIPHOR and a warning could in turn be generated in the pharmacy. This warning would then be returned to the ward with the prescribed drug where the physician could consider MYCIN's recommendations before deciding whether to administer the drug he had originally prescribed. The physician would, in effect, receive a consultation from MYCIN when he needed it rather than when he asked for it.

This approach to peer review provides an exciting potential for impacting the antimicrobial prescribing habits of physicians, and for monitoring other clinical practice questions as MYCIN-like knowledge bases are developed for additional problem areas. In this sense peer review may be considered as "covert consultation" in much the same sense that human consultations may be looked upon as "overt peer review." This model for prospective monitoring of prescribing habits is particularly appealing because it satisfies our proposed acceptability criteria for a peer review mechanism.

This section concludes with an example of a situation in which the monitoring model we have described would have been highly useful. During early development of the MYCIN system, we reviewed several patient charts in an effort to identify decision rules needed by the program. In one such chart we found that a patient had been treated with streptomycin as a single agent to combat an organism which was known to be resistant to streptomycin *in vitro*. Furthermore, the patient who was given the drug (which is toxic to the kidney) had chemistry laboratory values for BUN and creatinine indicating that he was in renal failure. In short, the streptomycin therapy was highly inappropriate. If MYCIN had been monitoring antimicrobial prescriptions in the hospital pharmacy, it would have automatically evaluated the streptomycin prescription. The lab values for BUN and creatinine would have been available from the clinical laboratory data base, and the microbiology data base would have revealed the organism's resistance to the drug. MYCIN would therefore have concluded that the streptomycin was inappropriate and a warning would have been generated. It is possible, in turn, that the warning would have had a beneficial educational impact on the physician who made the improper therapeutic decision. As was discussed in § 1.5.2, there is much evidence that this kind of inappropriate prescribing of an antibiotic is not an isolated incident, although the above example is, perhaps, somewhat extreme.

6.7 Educational Applications

As I have 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 once it is available on the wards.

It would be possible, however, 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. A similar instructional session might be generated even for actual patient cases provided by the student. Although there is great potential for this kind of educational use of MYCIN's knowledge base, we have no plans to pursue this application in the near future.

6.7 Other Applications of MYCIN Formalism

In § 3.8.3 I noted that one of the principal advantages of the MYCIN approach is its domain independent control structure. Attempts have also been made to preserve generality in Subprograms 2 and 3. We have not yet tested this claim with a second data base, however. As explained in Chapter 3, acquiring rules and defining parameters are such complex and time-consuming tasks that we have so far been unable to experiment with alternate clinical problem areas.

Our current plan is gradually to broaden MYCIN's knowledge base into other infectious disease topics (i.e., in addition to bacteremia). We feel it is important, however, eventually to test the approach in medical decision areas that have nothing to do with antimicrobial therapy. Not only will this assist in determining the generality of the MYCIN formalism, but it will also help us define which clinical problems are best suited for a rule-based system rather than for Bayesian or model-based approaches. As I have stated before, MYCIN's formalism seems to be most appropriate for applications in

which informal judgmental knowledge is the basis for decisions. If good statistical information is available or a problem is suited to physiological modeling, an alternate approach may be preferable. Until MYCIN is tested in new arenas, we will be unable to reach justifiable decisions regarding these issues.

It is also interesting to ask whether MYCIN's approach can be usefully applied to nonmedical problems. Although we have no current intention to investigate such questions ourselves, other AI researchers have begun to indicate an interest in pursuing this rule-based approach for nonmedical applications. Of particular relevance, of course, are those problems that can benefit from a technique for coding the heuristics of an individual.