CONSULTATION SYSTEMS FOR PHYSICIANS: The Role of Artificial Intelligence Techniques

Edward H. Shortliffe

Departments of Medicine and Computer Science
Heuristic Programming Project
Stanford University School of Medicine
Stanford, California 94305

ABSTRACT

Computer systems for use by physicians have had limited impact on clinical medicine. When one examines the most common reasons for poor acceptance of medical computing systems, the potential relevance of artificial intelligence techniques becomes evident. This paper proposes design criteria for clinical computing systems and demonstrates their relationship to current research in knowledge engineering. The MYCIN System is used to illustrate the ways in which our research group has attempted to respond to the design criteria cited.

1. INTRODUCTION

Although computers have had an increasing impact on the practice of medicine, the successful applications have tended to be in domains where physicians have not been asked to interact at the terminal. Few potential user populations are as demanding of computer-based decision aids. This is due to a variety of factors which include their traditional independence as lone decision makers, the seriousness with which they view actions that may have life and death significance, and the overwhelming time demands that tend to make them impatient with any innovation that breaks up the flow of their daily routine.

This paper examines some of the issues that have limited the acceptance of programs for use by physicians, particularly programs intended to give advice in clinical settings. My goal is to present design criteria which may encourage the use of computer programs by physicians, and to show that AI offers some particularly pertinent methods for responding to the design criteria outlined. Although the emphasis is medical throughout, many of the issues occur in other user communities where the introduction of computer methods must confront similar barriers. After presenting the design considerations and their relationship to AI research, I will use our work with MYCIN to illustrate some of the ways in which we have attempted to respond to the acceptability criteria I have outlined.

1.1. The Nature Of Medical Reasoning

It is frequently observed that clinical medicine is more an "art" than a "science". This statement reflects the varied factors that are typically considered in medical decision making; any practitioner knows that well-trained experts with considerable specialized experience may still reach very different conclusions about how to treat a patient or proceed with a diagnostic workup.

One factor which may contribute to observed discrepancies, even among experts, is the tendency of medical education to emphasize the teaching of facts, with little formal advice regarding the reasoning processes that are most appropriate for decision making. There has been a traditional assumption that future physicians should learn to make decisions by observing other doctors in action and by acquiring as much basic knowledge as possible. More recently, however, there has been interest in studying the ways in which expert physicians reach decisions in hopes that a more structured approach to the teaching of medical decision making can be developed [Kassirer 1978, Elstein 1978].

---

[1] This article is based on a longer paper to be published as a book chapter by Academic Press [Shortliffe 1980].

[2] Dr. Shortliffe is recipient of research career development award LM00048 from the National Library of Medicine.

Computer programs for assisting with medical decision making have tended not to emphasize models of clinical reasoning. Instead they have commonly assigned structure to a domain using statistical techniques such as Bayes' Theorem [deDombal 1972] or formal decision analysis [Gorry 1973]. More recently a number of programs have attempted to draw lessons from analyses of actual human reasoning in clinical settings [Wortman 1972, Pauker 1976]. Although the other methodologies may lead to excellent decisions in the clinical areas to which they have been applied, many believe that programs with greater dependence on models of expert clinical reasoning will have heightened acceptance by the physicians for whom they are designed.

1.2. The Consultation Process

Accelerated growth in medical knowledge has necessitated greater sub-specialization and more dependence upon assistance from others when a patient presents with a complex problem outside one's own area of expertise. Such consultations are acceptable to doctors in part because they maintain the primary physician's role as ultimate decision maker. The consultation generally involves a dialog between the two physicians, with the expert explaining the basis for advice that is given and the nonexpert seeking justification of points found puzzling or questionable. Consultants who offered dogmatic advice they were unwilling to discuss or defend would find that their opinions were seldom sought. After a recommendation is given, the primary physician generally makes the decision whether to follow the consultant's advice, seek a second opinion, or proceed in some other fashion. When the consultant's advice is followed, it is frequently because the patient's doctor has been genuinely educated about the particular complex problem for which assistance was sought.

Since such consultations are accepted largely because they allow the primary physician to make the final management decision, it can be argued that medical consultation programs must mimic this human process. Computer-based decision aids have typically emphasized only the accumulation of patient data and the generation of advice [Shortliffe 1979]. On the other hand, an ability to explain decisions may be incorporated into computer-based decision aids if the system is given an adequate internal model of the logic that it uses and can convey this intelligibly to the physician-user. The addition of explanation capabilities may be an important step towards effectively encouraging a system's use.

2. ACCEPTABILITY ISSUES

Studies have shown that many physicians are inherently reluctant to use computers in their practice [Startsman 1972]. Some researchers fear that the psychological barriers are insurmountable, but we are beginning to see systems that have had considerable success in encouraging terminal use by physicians [Watson 1974]. The key seems to be to provide adequate benefits while creating an environment in which the physician can feel comfortable and efficient.

Physicians tend to ask at least seven questions when a new system is presented to them:

(1) Is its performance reliable?

(2) Do I need this system?

(3) Is it fast and easy to use?

(4) Does it help me without being dogmatic?

(5) Does it justify its recommendations so that I can decide for myself what to do?

(6) Does use of the system fit naturally into my daily routine?

(7) Is it designed to make me feel comfortable when I use it?

Experience has shown that reliability alone may not be enough to insure system acceptance [Shortliffe 1979]; the additional issues cited here are also central to the question of how to design consultation systems that doctors will be willing to use.

3. DESIGN CRITERIA

The design considerations for systems to be used by physicians can be divided into three main categories: mechanical, epistemological, and psychological.

3.1. Mechanical Issues

It is clear that the best of systems will eventually fail if the process for getting information in or out of the machine is too arduous, frustrating, or complicated. Someday physician-computer interaction may involve voice communication by telephone or microphone, but technology is likely to require manual interaction for years to come. Thus, careful attention to the mechanics of the interaction, the simplicity of the displays, response tim^ accessibility of terminals, and self-do...mentation, are all essential for the successful implementation of clinical computing systems.

3.2. Epistemological Issues

As has been discussed, the quality of a program's performance at its decision making task is a basic acceptability criterion. A variety of approaches to automated advice systems have been developed, and many perform admirably [Shortliffe 1979]. Thus the capturing of knowledge and data, plus a system for using them in a coherent and consistent manner, are the design considerations that have traditionally received the most attention.

Other potential uses of system knowledge must also be recognized, however. As has been noted, physicians often expect to be educated when they request a human consultation, and a computer-based consultant should also be an effective teaching tool. On the other hand, physicians would quickly reject a pedantic program that attempted to convey every pertinent fact in its knowledge base. Thus it is appropriate to design programs that convey knowledge as well as advice, but which serve this educational function only when asked to do so by the physician-user.

As has been mentioned, physicians also prefer to understand the basis for a consultant's advice so that they can decide for themselves whether to follow the recommendation. Hence the educational role of the consultation program can also be seen as providing an explanation or justification capability. When asked to do so, the system should be able to retrieve and display any relevant fact or reasoning step that was brought to bear in considering a given case. It is also important that such explanations be expressed in terms that are easily comprehensible to the physician.

Since it would be unacceptable for a consultation program to explain <u>every</u> relevant reasoning step or fact, it is important that the user be able to request justification for points found to be puzzling. Yet an ability to ask for explanations generally requires that the program be able to understand free-form queries entered by the user. A reasonable design consideration, then, is to attempt to develop an interface whereby simple questions expressed in English can be understood by the system and appropriately answered.

It is perhaps inevitable that consultation programs dealing with complex clinical problems will occasionally reveal errors or knowledge gaps, even after they have been implemented for ongoing use. A common source of frustration is the inability to correct such errors quickly so that they will not recur in subsequent consultation sessions. There is often a lapse of several months between "releases" of a system, with an annoying error recurring persistently in the meantime. It is therefore ideal to design systems in which knowledge is easily modified and integrated; then errors can be rapidly rectified once the missing or erroneous knowledge is identified. This requires a flexible knowledge representation and powerful methods for assessing the interactions of new knowledge with other facts already in the system.

Finally, the acquisition of knowledge can be an arduous task for system developers. In some applications the knowledge may be based largely on statistical data, but in others it may be necessary to extract judgmental information from the minds of experts. Thus another design consideration is the development of interactive techniques to permit acquisition of knowledge from primary data or directly from an expert without requiring that a computer programmer function as an intermediary.

3.3. Psychological Issues

The most difficult problems in designing consultation programs may be the frequently encountered psychological barriers to the use of computers among physicians [Startsman 1972, Croft 1972]. Many of these barriers are reflected in the mechanical and epistemological design criteria mentioned above. However, there are several other pertinent observations:

(1) It is probably a mistake to expect the physician to adapt to changes imposed by a consultation system.

(2) A system's acceptance may be greatly heightened if ways are identified to permit physicians to perform tasks that they have wanted to do but had previously been unable to do [Mesel 1976, Watson 1974].

(3) It is important to avoid premature introduction of a system while it is still "experimental".

(4) System acceptance may be heightened if physicians know that a human expert is available to back up the program when problems arise.

(5) Physicians are used to assessing research and new techniques on the basis of rigorous evaluations; hence novel approaches to assessing both the performance and the clinical impact of medical systems are required.

4. KNOWLEDGE ENGINEERING

In recent years the terms "expert systems" and "knowledge-based systems" have been coined to describe AI programs that contain large amounts of specialized expertise that they convey to system users in the form of consultative advice. The phrase "knowledge engineering" has been devised [Michie 1973] to describe the basic AI problem areas that support the development of expert systems. There are several associated research themes:

(1) Representation of Knowledge. A variety of methods for computer-based representation of human knowledge have been devised, each of which is directed at facilitating the associated symbolic reasoning and at permitting the codification and application of "common sense" as well as expert knowledge of the domain.

(2) Acquisition of Knowledge. Obtaining the knowledge needed by an expert program is often a complex task. In certain domains programs may be able to "learn" through experience or from examples, but typically the system designers and the experts being modelled must work closely together to identify and verify the knowledge of the domain. Recently there has been some early experience devising programs that actually bring the expert to the computer terminal where a "teaching session" can result in direct transfer of knowledge from the expert to the system itself [Davis 1979].

(3) Methods of Inference. Closely linked to the issue of knowledge representation is the mechanism for devising a line of reasoning for a given consultation. Techniques for hypothesis generation and testing are required, as are focusing techniques. A particularly challenging associated problem is the development of techniques for quantitating and manipulating uncertainty. Although inferences can sometimes be based on established techniques such as Bayes' Theorem or decision analysis, utilization of expert judgmental knowledge typically leads to the development of alternate methods for symbolically manipulating inexact knowledge [Shortliffe 1975].

(4) Explanation Capabilities. For reasons I have explained in the medical context above, knowledge engineering has come to include the development of techniques for making explicit the basis for recommendations or decisions. This requirement tends to constrain the methods of inference and the knowledge representation that is used by a complex reasoning program.

(5) The Knowledge Interface. There are a variety of issues that fall in this general category. One is the mechanical interface between the expert program and the individual who is using it; this problem has been mentioned for the medical user, and many of the observations there can be applied directly to the users in other knowledge engineering application domains. Researchers on these systems also are looking for ways to combine AI techniques with more traditional numerical approaches to produce enhanced system performance. There is growing recognition that the greatest power in knowledge-based expert systems may lie in the melding of AI techniques and other computer science methodologies [Shortliffe 1979].

Thus it should be clear that artificial intelligence, and specifically knowledge engineering, are inherently involved with several of the design considerations that have been suggested for medical consultation systems. In the next section I will discuss how our medical AI program has attempted to respond to the design criteria that have been cited.

5. AN EXAMPLE: THE MYCIN SYSTEM

Since 1972 our research group at Stanford University[1] has been involved with the development of computer-based consultation systems. The first was designed to assist physicians with the selection of antibiotics for patients with

---

[1]Several computer scientists, physicians, and a pharmacist have been involved in the development of the MYCIN System. These include J. Aikins, S. Axline, J. Bennett, A. Bonnet, B. Buchanan, W. Clancey, S. Cohen, R. Davis, L. Fagan, F. Rhame, C. Scott, W. vanMelle, S. Wraith, and V. Yu.

serious infections. That program has been termed MYCIN after the suffix utilized in the names of many common antimicrobial agents. MYCIN is still a research tool, but it has been designed largely in response to issues such as those I have described. The details of the system have been discussed in several publications [Shortliffe 1976, Davis 1977, Scott 1977] and may already be well known to many readers. Technical details will therefore be omitted here, but I will briefly describe the program to illustrate the ways in which its structure reflects the design considerations outlined above.

5.1. Knowledge Representation and Acquisition

All infectious disease knowledge in MYCIN is contained in packets of inferential knowledge represented as production rules [Davis 1976]. These rules were acquired from collaborating clinical experts during detailed discussions of specific complex cases on the wards at Stanford Hospital. More recently the system has been given the capability to acquire such rules directly through interaction with the clinical expert[1].

MYCIN currently contains some 600 rules that deal with the diagnosis and treatment of bacteremia (bacteria in the blood) and meningitis (bacteria in the cerebrospinal fluid). These rules are coded in INTERLISP [Teitelman 1978], but routines have been written to translate them into simple English so that they can be displayed and understood by the user. For example, one simple rule which relates a patient's clinical situation with the likely bacteria causing the illness is shown in Fig. 1. The strengths with which the specified inferences can be drawn are indicated by numerical weights, or certainty factors, that are described further below.

```
RULE300
   [This rule applies to all cultures
and suspected infections, and is tried in
order to find out about the organisms
(other than those seen on cultures or
smears) which might be causing the
infection]

  If:  1) The infection which requires
          therapy is meningitis, and
       2) The patient does have evidence
          of serious skin or soft tissue
          infection, and
       3) Organisms were not seen on the
          stain of the culture, and
       4) The type of the infection is
          bacterial
  Then:  There is evidence that the
         organism (other than those seen
         on cultures or smears) which
         might be causing the infection
         is staphylococcus-coag-pos (.75)
         streptococcus-group-a (.5)
```

Figure 1 - A Sample MYCIN Rule

5.2. Inference Methods

5.2.1. Reasoning Model

Production rules provide powerful mechanisms for selecting those that apply to a given consultation. In MYCIN's case the rules are only loosely related to one another before a consultation begins; the program selects the relevant rules and chains them together as it considers a particular patient. Two rules chain together if the action portion of one helps determine the truth value of a condition in the premise of the other. The resulting reasoning network, then, is created dynamically and can be seen as a model of one approach to the patient's problem.

MYCIN's strategy in rule selection is goal-oriented. The program "reasons backwards" from its recognized goal of determining therapy for a patient. It therefore starts by considering rules for therapy selection, but the premise portion of each of those rules in turn sets up new questions or subgoals. These new goals then cause new rules to be invoked and a reasoning network is thereby developed. When the truth of a premise condition is best determined by asking the physician rather than by applying rules (e.g., to determine the value of a laboratory test), a question is displayed. The physician enters the appropriate response and the program continues to select additional rules. Once information on the patient is obtained, some rules will fail to be applicable; in this way the invoked applicable rules will provide a customized patient-specific reasoning network for the case under consideration.

Portions of a sample consultation session are shown in Fig. 2. The physician's responses are in capital letters and follow a double asterisk.

---

[1]This capability was implemented in rudimentary form in early versions of the system [Shortliffe 1976] but was substantially broadened and strengthened by Davis in his Teiresias program [Davis 1979].

Note that the physician can indicate that he does not know the answer to a question (Question 13) and MYCIN will proceed to do the best it can on the basis of the data available. Typographical or spelling errors are automatically corrected (Question 12) with the presumed meaning printed at the left margin so that the physician knows what assumption has been made. Note also that the physician's responses are generally single words.

In Fig. 3 the form of MYCIN's final conclusions and therapy recommendations are demonstrated. Note that the program specifies what organisms are likely to be causing the patient's infection and then suggests a therapeutic regimen appropriate for them. There are also specialized routines to calculate recommended drug doses depending upon the patient's size and kidney function.

```
    --------PATIENT-538--------
    1) Patient's name:
    ** PT538
    2) Age:
    ** 45
    3) Sex:
    ** MALE
    4) Race:
    ** CAUC
=CAUCASIAN
    5) Are there any cultures for Pt538
       which may be related to the present
       illness, and from which organisms
       have been grown successfully in
       the microbiology laboratory?
    ** NO
    6) Are there any pending cultures?
    ** YES

       --------CULTURE-1--------
       7) From what site was the specimen
          for CULTURE-1 taken?
       ** CSF
       8) Please give the date and time
          when the pending csf culture
          (CULTURE-1) was obtained.
       ** 11/28/78
       9) Was a smear examined at the time
          that the specimen for the pending
          csf culture (CULTURE-1) was
          obtained?
       ** YES
       10) Were any organisms seen on the
          smear of the specimen submitted
          as the pending csf culture
          (CULTURE-1)?
       ** YES

          --------ORGANISM-1--------
          11) The stain (Gram, India ink,
             or Ziehl-Neelsen acid-fast) of
             ORGANISM-1:
          ** GRAMNEG
          12) Is ORGANISM-1 a rod or coccus
             (etc.):
          ** ROOD
=ROD
          13) What is the form of the
             individual organisms (e.g.
             Lancet-shaped for cocci,
             fusiform for rods, etc.)?
          ** ??
          Expected responses are: FUSIFORM,
          PLUMP or DUMBBELL-SHAPED
          Enter HELP for list of options.
          ** UNKNOWN
       14) Any other organisms noted on the
          smear?
       ** NO
```

Figure 2
The Beginning Of A Sample Consultation

```
   My therapy recommendations will be
designed to treat for organisms that
either are very likely or, although less
likely, would have a significant effect
on therapy selection if they were
present. It is important to cover for
the following probable infection(s) and
associated organism(s):

INFECTION-1 is MENINGITIS

   + <ITEM-1>  E.COLI
   + <ITEM-2>  PSEUDOMONAS-AERUGINOSA
   + <ITEM-3>  KLEBSIELLA-PNEUMONIAE
   + <ITEM-4>  DIPLOCOCCUS-PNEUMONIAE

    41) Does Pt538 have a clinically
       significant allergic reaction to
       any antimicrobial agent?
    ** NO
    42) Do you have reason to suspect that
       Pt538 may have impaired renal
       function?
    ** NO
    43) Pt538's weight in kilograms:
    ** 70

[REC-1] My preferred therapy
        recommendation is:
   In order to cover for Items <1 2 3 4>:
        Give the following in combination:
        1) AMPICILLIN
        Dose: 3.5g (28.0 ml) q4h IV
              [calculated on basis of 50
              mg/kg]
        2) GENTAMICIN
        Dose: 119 mg (3.0 ml, 80mg/2ml
              ampule) q8h IV [calculated
              on basis of 1.7 mg/kg] plus
              consider giving 5 mg q24h
              intrathecally

Since high concentrations of penicillins
can inactivate aminoglycosides, do not mix
these two antibiotics in the same bottle.
```

Figure 3
Example of MYCIN's Recommendations

5.2.2. Management of Uncertainty

The knowledge expressed in a MYCIN rule is seldom definite but tends to include "suggestive" or "strongly suggestive" evidence in favor of a given conclusion. In order to combine evidence regarding a single hypothesis but derived from a number of different rules, it has been necessary to devise a numeric system for capturing and representing an expert's measure of belief regarding the inference stated in a rule. Although this problem may at first seem amenable to the use of conditional probabilities and Bayes' Theorem, a probabilistic model fails to be adequate for a number of reasons we have detailed elsewhere [Shortliffe 1975]. Instead we use a model that has been influenced by the theory of confirmation, and have devised a system of belief measures known as certainty factors. These numbers lie on a -1 to +1 scale with -1 indicating absolute disproof of an hypothesis, +1 indicating its proof, and 0 indicating the absence of evidence for or against the hypothesis (or equally weighted evidence in both directions). The relationship of the model to formal probability theory and the methods for combining evidence from diverse sources (rules and user estimates) have been described [Shortliffe 1975]. Although the system has served us well to date, it does have several recognized inadequacies [Adams 1976] and can only be seen as a first step towards the development of a coherent theory for the management of uncertainty in complex reasoning domains.

5.3. Explanation

Perhaps the greatest advantage of the rules used in MYCIN is the way in which they facilitate the development of mechanisms for explaining and justifying system performance. These capabilities also contribute greatly to MYCIN's educational role [Clancey 1979]. The explanation program has two options, one of which is limited in scope but fast and easy to use. This option is used during a consultation session and allows examination of the reasoning steps that have generated a specific question. The second option is more powerful but it involves language processing and is therefore less reliable and more time consuming. This question-answering capability is designed to allow full access to all system knowledge, both static (i.e., facts and rules obtained from experts) and dynamic (i.e., conclusions reached by the program for the specific consultation session). Each system is briefly described below.

5.3.1. Answers To Questions About The Reasoning Status

During the consultation, MYCIN may ask questions which the physician finds puzzling, either because they appear irrelevant or because the details of their applicability are unclear. When this occurs, the physician can digress momentarily and seek an explanation of the question[1]. Instead of answering the question MYCIN has asked, the user enters the word "WHY" and the program responds with a translation of the rule which

```
   35) Is this a hospital-acquired
       infection?
   ** WHY

[i.e. WHY is it important to determine
whether the infection was acquired while
the patient was hospitalized?]

[2.0] This will aid in determining the
organisms (other than those seen on
cultures or smears) which might be causing
the infection.
   It has already been established that
        [2.1] the infection which requires
              therapy is meningitis, and
        [2.2] organisms were not seen on
              the stain of this pending
              csf culture, and
        [2.3] the type of the infection is
              bacterial, and
        [2.4] neurosurgery is not one of
              the relevant items from the
              history of pt538, and
        [2.5] the age of pt538 is greater
              than 10 days
     Therefore, if
        [2.6] the infection was acquired
              while the patient was
              hospitalized
     then:
           there is evidence that the
           organism (other than those
           seen on cultures or smears)
           which might be causing the
           infection is e.coli (.75)
           staphylococcus-coag-pos (.3)
           pseudomonas-aeruginosa (.3)
           klebsiella-pneumoniae (.5)

   [back to question 35...]
   **
```

Figure 4
Example of the WHY Command

[1]The mechanisms for examining the reasoning status using "WHY" and "HOW" commands were largely the work of Davis in his Teiresias program [Davis 1979]. The techniques he developed are general in their applicability and have been implemented in nonmedical domains as well.

generated the question. An example of this feature is shown in Fig. 4. Note that MYCIN begins its response by phrasing in English its understanding of the "WHY question" asked by the physician. It then displays the relevant rule, specifying which conditions in the premise are already known to be true and which conditions remain to be investigated. In many cases this single rule displayed is an adequate explanation of the current line of reasoning and the physician can then proceed with the consultation by answering the question.

The user can alternatively continue to investigate the current reasoning by repeating the "WHY" command several times. Each additional "WHY" is interpreted by MYCIN as a request for display of the next rule in the current reasoning chain. For example, in Fig. 4 another "WHY" would be equivalent to asking "Well then, why are you trying to determine the organisms which might be causing the infection?" After responding to each "WHY", MYCIN returns to the current question and awaits the physician's response.

The "HOW" command is used in conjunction with the "WHY" command. Note that MYCIN places a unique number before each condition in the premise of a displayed rule (Fig. 4). If the user does not understand <u>how</u> a specified condition was met (or <u>will</u> be met), the "HOW" command, followed by the identifying number of the premise clause in question, can be used. Hence "HOW" requires an argument as shown in Fig. 5. This example follows immediately on the "WHY" example in Fig. 4; thus the argument "2.3" refers to the third condition in the premise of RULE545 in Fig. 4 and "2.5" refers to the fifth. MYCIN's responses to "HOW" commands are self-explanatory. Note also that components of all explanations are given identifiers so that, in Fig. 5, the command "HOW 3.1" would have resulted in a display of RULE526.

Other features of the capability to assess the reasoning status during a consultation are described elsewhere [Davis 1977, Scott 1977, Davis 1979]. Two points require emphasis here, however. First, these features require no English language processing since the user gives simple single-word commands; hence the capability is fast, efficient, and reliable. On the other hand, questions are limited to the immediate reasoning environment surrounding the question asked by MYCIN. Hence general access to other system knowledge or other conclusions previously made by the program is not possible. To provide a more general explanation capability, MYCIN must be able to respond to questions expressed in simple English as described below.

```
   [back to question 35...]
   ** HOW 2.3

[i.e. HOW was it established that the type
      of the infection is bacterial?]

The following rules concluded about the
      likelihood that the type of the
      infection is bacterial:

                 Rule          Certainty
       [2.7]     RULE148         (.19)
       [2.8]     RULE500         (.51)
       [2.9]     RULE501         (.95)
       [3.0]     RULE502         (.97)
       [3.1]     RULE526         (.98)
       [3.2]     RULE504         (.97)
       [3.3]     RULE524         (.98)

Thus, it has been established that the
     type of the infection is bacterial.

   [back to question 35...]
   ** HOW 2.5

[i.e. HOW was it established that the age
     of pt538 is greater than 10 days?]

In answer to question 2 you said that the
     age of pt538 is 34 years. Thus, it
     has been established that the age of
     pt538 is greater than 10 days.

   [back to question 35...]
   **
```

<u>Figure 5</u>
<u>Examples of the HOW Command</u>

5.3.2. <u>Answers To Questions About The Consultation And Knowledge Base</u>

MYCIN's general question-answering feature is automatically made available to the user at the end of each consultation session. After a therapy recommendation has been given, the physician may still have several questions regarding details of the consultation and the basis for diagnostic and therapeutic advice. Fig. 6 shows an example of a simple English question which MYCIN understands and attempts to answer. Our approach to natural language understanding is very simplistic [Scott 1977] but often adequate if questions are kept short without subordinate clauses. In many cases questions are best answered simply by retrieving relevant rules and displaying them in a format suitable for the question being asked. In Fig. 6 MYCIN offers a detailed explanation of how it had decided that the patient might be infected with an <u>E. coli</u>. All rules are specified by number so that the physician can have them displayed if full details are desired.

** WHY DID YOU TREAT E.COLI?

ITEM-1 is E.coli associated with meningitis. I decided to treat ITEM-1 on the basis of clinical evidence only.

The following rules concluded about:
the likelihood that e.coli is one of the organisms (other than those seen on cultures or smears) which might be causing the infection

RULE545 was used to conclude that e.coli is one of the organisms (other than those seen on cultures or smears) which might be causing the infection (.74).
RULE557 was used to conclude that e.coli is one of the organisms (other than those seen on cultures or smears) which might be causing the infection (.93).

The parameters most frequently used in these rules:
 the infection which requires therapy
 whether a smear of the culture was examined
 whether organisms were seen
 the type of the infection

Figure 6
Response To A Free Text Question

Fig. 7 demonstrates the way in which a physician can ask general questions of the knowledge base (static knowledge); neither the questions nor MYCIN's responses refer to any specific consultation session. Note that MYCIN has a sense of its limitations and makes it clear that it has no knowledge specific to pneumonia. Also, a large dictionary includes synonyms and hence permits MYCIN to recognize that "pneumococcal" refers to "pneumococcus", which in turn is the same as "diplococcus-pneumoniae".

Despite their utility, the language understanding features demonstrated in Figs. 6 and 7 have significant limitations. The associated computer code is large, complex, and very slow. Furthermore, we have found that it takes novice users several sessions before they learn the best ways to phra .e questions so that MYCIN will interpret them properly.

5.4. Evaluating MYCIN

As work on MYCIN has progressed, we have begun to devise techniques for analyzing formally the system's

---

** HOW DO YOU DECIDE THAT AN ORGANISM IS AN E.COLI?
The following rules conclude about:
 the likelihood that the identity of the organism is E.coli
555, 304, 243, 3
Which of the rules do you wish to see?
 ** NONE

** WHAT IS THE TREATMENT FOR PNEUMOCOCCAL PNEUMONIA?
The MYCIN system currently can treat the following infections: BACTEREMIA MENINGITIS CYSTITIS. Because the knowledge base is incomplete, all other infections will be treated like bacteremia.

For treatment of DIPLOCOCCUS-PNEUMONIAE in BACTEREMIA, the drugs of choice (ranked by relative efficacy, assuming sensitivity) are:
 1st choice:
  PENICILLIN or AMPICILLIN
 2nd choice:
  NAFCILLIN, CEPHALOTHIN, ERYTHROMYCIN, or CLINDAMYCIN
 3rd choice:
  TETRACYCLINE, CHLORAMPHENICOL, or VANCOMYCIN

** WHAT DOSE OF ERYTHROMYCIN DO YOU RECOMMEND?
When ERYTHROMYCIN is given for BACTEREMIA, the recommended dosage is:
if age is > 12 yrs then 12 mg/kg q6h IV
else if age > .08 yrs then 5 mg/kg q6h IV
else if age > .02 yrs then 5 mg/kg q12h IV

Figure 7
Requests For MYCIN's General Knowledge

---

performance. It must be emphasized, however, that the decision making performance is only one aspect of overall system acceptability; as I have discussed, many of the most significant problems occur when attempts are made to encourage physicians to use a program, even after it has been shown to reach good decisions.

The details of the evaluation studies will not be presented here[1], but a number of specific points are of interest. First any evaluation is difficult because there is so much difference of opinion in this domain, even among experts. Hence, it is unclear how to select a "gold standard" by which to measure the system's performance.

---

[1]See [Yu 1979a] for the details of the bacteremia evaluation, and [Yu 1979b] for the data on MYCIN's performance selecting therapy for patients with meningitis.

Actual clinical outcome cannot be used because each patient of course is treated in only one way and because a poor outcome in a gravely ill patient cannot necessarily be blamed on the therapy that had been selected.

Second, although MYCIN performed at or near expert level in almost all cases, the evaluating experts in one study [Yu 1979a] had serious reservations about the clinical utility of the program. It is difficult to assess how much of this opinion is due to actual inadequacies in system knowledge or design and how much is related to inherent bias against any computer-based consultation aid. In a subsequent study we attempted to eliminate this bias from the study by having the evaluators unaware of which recommendations were MYCIN's and which came from actual physicians [Yu 1979b]. In that setting MYCIN's recommendations were uniformly judged preferable to, or equivalent to, those of five infectious disease experts who recommended therapy for the same patients.

Finally, those cases in which MYCIN has tended to do least well are those in which serious infections have been simultaneously present at sites in the body about which the program has been given no rules. It is reasonable, of course, that the program should fail in areas where it has no knowledge. However, a useful antimicrobial consultation system must know about a broad range of infectious diseases, just as its human counterpart does. Even with excellent performance managing isolated bacteremias and meningitis, the program is therefore not ready for clinical implementation.

There will eventually be several important questions regarding the clinical impact of MYCIN and systems like it. Are they used? If so, do the physicians follow the program's advice? If so, does patient welfare improve? Is the system cost effective when no longer in an experimental form? What are the legal implications in the use of, or failure to use, such systems? The answers to all these questions are years away for most consultation systems, but it must be recognized that all these issues are ultimately just as important as whether the decision making methodology manages to lead the computer to accurate and reliable advice.

6. CONCLUSION

Although I have asserted that AI research potentially offers solutions to many of the important problems confronting researchers in computer-based clinical decision making, the field is not without its serious limitations. However, AI has reached a level of development where it is both appropriate and productive to begin applying the techniques to important real world problems rather than purely theoretical issues. The difficulty lies in the fact that such efforts must still dwell largely in research environments where short term development of systems for service use is not likely to occur.

It is also important to recognize that other computational techniques may meld very naturally with AI approaches as the fields mature. Thus we may see, for example, direct links between AI methods and statistical procedures, decision analysis, pattern recognition techniques, and large databanks. As researchers in other areas become more familiar with AI, it may gradually be brought into fruitful combination with these alternate methodologies. The need for physician acceptance of medical consultation programs is likely to make AI approaches particularly attractive, at least in those settings where hands-on computer use by physicians is desired or necessary. This paper has attempted to explain why the wedding of AI and medical consultation systems is a natural one and to show, in the setting of the MYCIN system, how one early application has responded to design criteria identified for a user community of physicians.

REFERENCES

Adams, J. B. "A probability model of medical reasoning and the MYCIN model." Math. Biosci. 32,177-186 (1976).

Clancey, W.J. Transfer of Rule-Based Expertise Through a Tutorial Dialogue. Dictoral dissertation, Stanford University, September 1979. Technical memo STAN-CS-79-769.

Croft, D. J. "Is computerized diagnosis possible?" Comp. Biomed. Res. 5,351-367 (1972).

Davis, R. and King, J. "An overview of production systems." In Machine Representation of Knowledge (E. W. Elcock and D. Michie, eds.), New York: Wiley, 1976.

Davis, R., Buchanan, B. G., and Shortliffe, E. H. "Production rules as a representation for a knowledge-based consultation system." Artificial Intelligence 8,15-45 (1977).

Davis, R., "Interactive transfer of expertise: acquisition of new inference rules." Artificial Intelligence, 12,121-157 (1979).

deDombal, F. T., Leaper, D. J., Staniland, J. R., et al. "Computer-aided diagnosis of acute abdominal pain." Brit. Med. J. 2,9-13 (1972).

Elstein, A. S., Shulman, L. S., and Sprafka, S. A. Medical Problem Solving: An Analysis of Clinical Reasoning. Cambridge, Mass.: Harvard Univ. Press, 1978.

Gorry, G. A., Kassirer, J. P., Essig, A., and Schwartz, W. B. "Decision analysis as the basis for computer-aided management of acute renal failure." Amer. J. Med 55,473-484 (1973).

Kassirer, J. P. and Gorry, G. A. "Clinical problem solving: a behavioral analysis." Anns. Int. Med. 89,245-255 (1978).

Mesel, E., Wirtschafter, D. D., Carpenter, J. T., et al. "Clinical algorithms for cancer chemotherapy - systems for community-based consultant-extenders and oncology centers." Meth. Inform. Med. 15,168-173 (1976).

Michie, D. "Knowledge engineering." Cybernetics 2,197-200 (1973).

Pauker, S. G., Gorry, G. A., Kassirer, J. P., and Schwartz, W. B. "Towards the simulation of clinical cognition: taking a present illness by computer." Amer. J. Med. 60:981-996 (1976).

Scott, A. C., Clancey, W., Davis, R., and Shortliffe, E. H. "Explanation capabilities of knowledge-based production systems." Amer. J. Computational Linguistics, Microfiche 62, 1977.

Shortliffe, E. H. and Buchanan, B. G. "A model of inexact reasoning in medicine." Math. Biosci. 23,351-379 (1975).

Shortliffe, E. H. Computer-Based Medical Consultations: MYCIN, New York: Elsevier/North Holland, 1976.

Shortliffe, E. H. Buchanan, B. G. and Feigenbaum, E. A. "Knowledge engineering for medical decision making: a review of computer-based clinical decision aids." PROCEEDINGS of the IEEE, 67,1207-1224 (1979).

Shortliffe, E.H. "Medical consultation systems: designing for doctors." In Communication With Computers (M. Sime and M. Fitter, eds.), Academic Press, London, 1980 (in press).

Startsman, T. S. and Robinson, R. E. "The attitudes of medical and paramedical personnel towards computers." Comp. Biomed. Res. 5,218-227 (1972).

Teitelman, W. INTERLISP Reference Manual, XEROX Corporation, Palo Alto, Calif. and Bolt Beranek and Newman, Cambridge, Mass., October 1978.

Watson, R. J. "Medical staff response to a medical information system with direct physician-computer interface." MEDINFO 74, pp. 299-302, Amsterdam: North-Holland Publishing Company, 1974.

Wortman, P. M. "Medical diagnosis: an information processing approach." Comput. Biomed. Res. 5,315-328 (1972).

Yu, V. L. Buchanan, B. G. Shortliffe, E. H. et al. "Evaluating the performance of a computer-based consultant." Comput. Prog. Biomed. 9,95-102 (1979a).

Yu, V. L. Fagan, L. M. Wraith, S. M. et al. "Computerized consultation in antimicrobial selection - a blinded evaluation by experts." J. Amer. Med. Assoc. 242,1279-1282 (1979b).