September 1983

Report No. STAN-CS-82-998
*Also numbered HPP-83-40*

# Knowledge Engineering: A Daily Activity on a Hospital Ward

by

Benoit Mulsant and David Servan-Schreiber

**Department of Computer Science**

Stanford University
Stanford, CA 94305

HPP-83-40

# Knowledge Engineering: A Daily Activity On A Hospital Ward


Benoit Mulsant and David Servan-Schreiber

Faculte de Medecine
Universite Laval
QUEBEC

*and*

Faculte de Medecine Necker-Enfants Malades
Universite Rene Descartes
PARIS

Benoit Mulsant and David Set-van-Schreiber
31 rue Des Remparts
Quebec PQ
G1 R 3R4
CANADA



*This paper, minus the Appendix, has been accepted for publication in Computers and Bromedrcal Research. It describes work undertaken as a Stanford class project in Computer Science 221B (Computer-Based Medical Decision Aids) in Spring 1983.*

## Abstract

Two common barriers against the development and diffusion of Expert Systems in Medicine are the difficulty of design and the low level of acceptance. This paper reports on an original experience which entails potential solutions of these issues : the task of Knowledge Engineering is performed by medical students and residents on a hospital ward using a sophisticated Knowledge Acquisition System, EMYCIN. The Knowledge Engineering sessions are analysed in detail and a structured method is proposed. A transcript of a sample run of the resulting program is presented along with an evaluation of its performance, acceptance, educational potential and amount of endeavour required. The impact of the Knowledge Engineering process itself is then assessed both from the residents and the medical students standpoint. Finally, the possibility of generalizing the experiment is examined.

# 1. INTRODUCTION : THE BLUE-BOX PROJECT

During the past decade AIM (Artificial Intelligence in Medicine) researchers have developed new tools to help physicians in their daily practice: the expert *systems* (ES) **(1).** Some of these ES have undergone extensive tests and their ability to perform has been shown to be at the level of the best expert physicians in their domain. Nevertheless most of these programs have never left AI laboratories. The systems actually in clinical use are the exceptions **(2).** This issue is not new and many explanations have been proposed. Some of the most eminent figures in the field argued that the problem was a technical one and that the solution would be to build better performing systems (performance being taken in a broad sense) implemented on powerful interactive computers providing instantaneous answers and allowing good man/machine interaction in natural language. Other important capabilities include explanation and justification of the line of reasoning, broad knowledge about real clinical problems, easy access to the knowledge and of course a good level of expertise.

Now a new generation of ES displays these features, still they are no more used on the wards than their predecessors. Clearly technical performance is not enough. We have come to believe that part of the problem resides in the way the existing systems were conceived and built. The Blue-Box Project was developed to define a new approach for the construction of medical ES and to test if it will help them to penetrate the real (clinical) world.

In the first part we briefly discuss the concepts of this new approach. Then we describe how we tested it in developing the Blue-Box program -- a psychopharmalogical ES specialized for the treatment of depression -- using the EMYCIN program in the environment of the Palo Alto VA Mental Health Clinical Research Center. We present a sample of a Blue-Box consultation, and address some control and presentation issues. We then discuss the results of our knowledge engineering experience and finally we conclude by evaluating how we have reached our goal: to show that a different development strategy for ES is possible and could improve their acceptability on the clinical ground.

# 2. MOTIVATION

Most medical ES have, so far, been developed by AI specialists -- the knowledge *engineers* -- working with one or two senior physicians specialized in a subfield of medicine -- the *domain experts* -- in AI research laboratories far from the clinical wards where these programs were supposed to be used. Very few of the knowledge engineers were physicians and sometimes they were mostly interested in developing new AI techniques using the medical knowledge as a sample knowledge to test their ideas. No educational efforts were directed towards the potential medical users. “Perfect” unique and definitive products were built, theoretically ready to be used in every possible medical environment. The reasons behind this one-sided AI emphasis were:

1. These programs being considered as research prototypes whose usefulness in the real world was not the major concern of their authors.

2. The human and financial cost of development (many man-months of highly specialized physicians and computer scientists using expensive machine).

3. Most of all: the absence of alternatives i.e the lack of tools to build ES implying long and difficult program development by sophisticated AI programmers starting from scratch.

We can understand why, in the tough and self-centered medical world, these programs did not meet the success expected by their authors, even though they represented, from an AI point of view, real . breakthroughs.

Today the situation is different. Computer time is cheaper. Tools to build ES, with no or little experience in computer science are available. These programs -- KAS (3), EMYCIN (4) -- are in fact ES stripped of their specific domain knowledge and expanded with knowledge base building facilities. This made possible the Blue-Box project.

Our goal was to test whether medical students without AI experience could use one of these -programs to develop a useful medical ES in a clerkship time (one or two months). To do so, they would not work with a senior expert but would take advantage of their basic medical knowledge to use (1) the specific domain knowledge available in the literature and (2) the available “on ward” help from the residents. If it was possible, it would allow any MD or medical student to develop the ES he needs and could provide a medical community with a bank of ES dealing with the clinical problems that were locally felt challenging. The problem of the human cost could be solved and we believe that the routine use of computer systems in the clinical decision making process could be greatly improved : (1) it would introduce these new tools in a non-threatening, gentle way to the medical community because computer literate physicians would not feel that they lose control of their task (2) the performance issue would be partially solved since one tends to be far more satisfied with his own program than with “ready to use” programs.

# 3. METHOD

We tested the feasibility of this approach during a two month clerkship in psychiatry at the Stanford University Medical School. During this regular clerkship done at the Palo Alto VA Mental Health Clinical Research Center (MHCRC) we had the possibility aside from our clinical work to spend some time to develop a research project. Thanks to Professor Bruce Buchanan and Professor Edward Shortliffe, we also had the opportunity to use the EMYCIN program available on the SUMEX-AIM DEC 20. We decided to use this tool to develop a computer advisor on the psychopharmalogical antidepressant therapy.

Before analyzing the development process and its concrete results (not only the resulting ES but also the impact on the ward of this process) we shall describe briefly EMYCIN, the MHCRC ward and the medical problem of antidepressant therapy.

### 3.1. The Knowledge Acquisition Program : EMYCIN

Among the research done in AI a focus of interest has been to develop programs that aid 'scientists with complex reasoning tasks. One key to the creation of such intelligent systems is the incorporation of large amounts of task-specific knowledge. This knowledge is used to "understand" the domain (descriptive knowledge) and to reason about it (inference rules). But building such "expert systems" from scratch can be very time consuming. Thus some general, domain independent tools to aid the construction of these programs have been designed. EMYCIN is one of them. It is based on the knowledge-independent core of the MYCIN program developed by Dr. E. Shortliffe at Stanford University (5) -- MYCIN was constructed to advise a physician on the selection of anti-microbial therapy. Results of formal evaluations of MYCIN's competence in the domains of bacteremia and meningitis indicate that in this area it has begun to approach that of an expert. Quoting from the EMYCIN manual (4):

> EMYCIN is used to construct and run a consultation program, a program that offers advice on problems within its domain of expertise. The consultation program elicits information about a particular problem (a "case") by asking questions to a user. It then applies its knowledge to the specific facts about the case and informs the user of its conclusions. The user is free to ask the program questions about its reasoning in order to better understand or validate the advice given.

For example, the Blue-Box program was constructed to advise a physician on the selection of an appropriate therapy for a patient complaining of depression. As the reader will see, the system reasons about the type of depression, evaluates the suicidal risk of the patient and proposes a management plan. Blue-Box bases its reasoning on raw data the physician supplies regarding the clinical presentation, the medical history, the psychiatric history, the drug history and the family history.

So, there are really two users of EMYCIN. The know/edge engineer enters descriptive knowledge of the domain and rules about how to carry out the consultation. He uses EMYCIN to produce the knowledge-base. EMYCIN then interprets this knowledge to provide advice to the *consultation user* on a specific domain problem.

### 3.1.1. Rules

The central concept behind the control structure of EMYCIN is that of a decision rule. An example of such a rule, taken from Blue-Box, is :

```
RULE 027
--------
If :  1) The patient has faced severe psycho-social
         stressors (cumulative loss history...) over
         the past years,
      2) The duration of this depressive episode is
         less than or equal to 6 months,
      3) The sleep-pattern of the patient is one of :
         late-evening-insomnia middle-night-insomnia, and"
      4) The highest level of functioning prior to
         this episode is good
 Then :  There is suggestive evidence (.7) that
         grief-reaction is one of the type of depression
```

All the medical decision making knowledge of the system is stored in this rule form. Rules have the general form : Premise -> Conclusion. Premises consist of boolean predicates on the value of parameters. In the example given above the *duration* of *the depressive episode* is a parameter and a Boolean predicate is the “less than or equal to” relationship to the value 6 months.

### 3.1.2. Parameters

EMYCIN uses three types of parameters : Yes/No parameters (Boolean), Single-value and **Multi**-value parameters. *Intoxicated* -- whether the patient is currently intoxicated -- is a Yes/No parameter, its value is simply yes or no. *Age* or *severity* are Single-value parameters, they can take one of **several** distinct values. Finally, the *mood of the patient* is a Multi-value parameter **as** it can be simultaneously sadness, anxiety and hopelessness.

Parameters contain all the "static" knowledge of the system. Some are used to describe the patient *(name, age, sex,mood,* libido...), others describe treatments *(types of treatment recommended, best tricyclic anti-depressants for this case...).* They store all the data used to build a representation both of the general knowledge used and of the case at hand.

### 3.1.3. Backward Chaining

The basic control structure operates by using the rules to deduce the value of parameters through *backward-chaining. The* system operates backward, starting from the desired conclusion to find the information necessary to reach that conclusion. For example, when the system tries to determine the type of depressive episode, it will try to apply Rule 27 displayed above. For this rule to apply, the four premises must be true. Thus the system sets new goals, among which one is "Has the patient faced severe psycho-social stressors ?". It then looks for rules concluding about the psycho-social stressors, which yield yet other premises to be tested and so on. The process terminates when, eventually, values will be needed for parameters for which there are no concluding rules. When this happens, the user is asked to supply the values, and the succession of rules can then be applied to reach the original conclusion.

### 3.1.4. Certainty Factors

Often, when describing their clinical decision making knowledge, physicians use expressions like : "If... Then I am quite sure that..." **or"If...** Then I am less likely to...". EMYCIN provides a way to weigh the evidence in favor of a conclusion in order to represent this type of reasoning: the *Certainty Factors* (CF). A certainty factor is a number between - 1 and + 1 that indicate the strength of evidence in favor or against a conclusion (assigning a particular value to a parameter). In the example given above the **(.7)** shown in the conclusion is an indication of the strength of evidence in favor of assigning *grief reaction* as a value for *the type of the depression.*

Certainty factors can also be associated to responses given by the user during the consultation to represent the "uncertainty" about the presence or the degree of severity of a finding (if no CF is specified, the default value is 1) :

4) Please, characterize the depressed feelings of Jaime with any of followings : sadness, hopeless, helpless, anhedonia, anergia, anxiety -guilt-feelings, worthless, other or none:

SADNESS ANHEDONIA (.7) ANXIETY (.4)

The system provides the following **translation** of CF entered by the user : 2 < CF < = 4 : there is *weakly suggestive* evidence that...; 4 < CF < = 8 : there is *suggestive* evidence that...; 8 < CF <1 : there is *strongly suggestive* evidence that...; CF = *1 :* it is *definite* that...

**EMYCIN's** use of the rules for encoding knowledge, together with the certainty factor formalism for combining evidence, makes it relatively easy to "translate" medical knowledge in a machine understandable format. It also allows an incremental and modular development of the program. Not all rules have to be defined from the onset. The scope of the program can be progressively broadened by adding new rules, independently of the knowledge already present.

**3.1.5.** Explanation module and natural language interface

EMYCIN also tries to be user-friendly, both to the Knowledge Engineers and to the consultation user. Rules entered in the system are automatically translated in English ; this translation provides a transparent and comfortable material to present to non-computer-literate physicians associated to the development of the Knowledge base. Another feature of EMYCIN allows the consultation user to ask why a request is made by the program or how a conclusion has been drawn. Finally, the user can call a sub-system (the Question Answering Module) which understands questions in a sub-set of natural english and gives information either about the present consultation (e.g "Why did you choose TRAZODONE for this patient ?") or information about the general knowledge of the system (e.g "How do you conclude that ECT is a possible treatment ?")

Thus EMYCIN was a sophisticated tool to start with. Its very sophistication made our experiment possible. Moreover, numerous systems had already been developed with EMYCIN before we started this study, and it was generally acknowledge that the task was feasible.

3.2. The Medical Subject : Management of Depression

### 3.2.1. Initial Considerations

In order to give a fair trial to our initial assumption - that a reasonable expert system can be developed by medical students with no experience in Knowledge-engineering on a ward where the project is not a main focus of attention - our medical subject had to exhibit the following characteristics :

- Motivating strength : sufficiently close to the staff's everyday practice, and yet sufficiently complicated at the resident level to require clarification and in-depth study.

- Extensibility to non-expert practice : causing frequent problems to physicians outside the field of Psychiatry, to demonstrate the possible impact of an ideal, final system as a medical teaching and consulting tool.

- Compatibility with the rule-based approach of **EMYCIN** : deductive in nature rather than constraint-satisfaction or iterative.

- Accessibility of the knowledge : core structure accessible in the medical literature. Moreover, it was important that at least part of the knowledge is shared among the staff in order to divide the burden of knowledge provider. A potential problem was that Psychiatry has the undeserved reputation of using an unstructured terminology, not agreed upon by most practitioners. In fact, this has not been an issue as we decided to use the largely shared terminology of the DSM III (6).

- Modularity : allowing an incremental growth of the system by **stepwise** refinement.

Among the subjects initially envisaged, consultation for the management of a depressed patient seemed to best comply with these constraints.

### 3.2.2. Justification

Management of a depressed patient involves two separate tasks : prescription of a drug therapy (Does the patient require a pharmacological treatment ? And what drug ?) and evaluation of the suicidal risk (should the patient be hospitalized and special suicidal precautions taken ?) These two aspects are examined here.

Before prescribing a drug treatment, the physician should first assess if it is warranted. This is typically done by ruling out certain organic diseases (like hypothyroidism for example) and drugs which might induce depression. Some psychiatric diseases can also simulate a depressive episode (Dementia, Schizophrenia,...) and should be screened for.

The second step in the evaluation is to characterize the type of depression. Some patient might principally suffer from a grief reaction when other became depressed for no apparent reason. Overall, there are approximately nine different "prototypes" of depressions.

Once the physician has a general idea about the type, several cues help him decide what type of treatment is recommanded. In addition to drugs, treatment types include Electra-Convulsive Therapy (ECT) and psychotherapy. A specific depression type does not, in most instances, directly require a particular treatment and even less so a particular drug. Multiple cues, which can be present in any of the different types, are relevant to the choice of a treatment. Moreover among the eight most commonly used "tricyclic" and related anti-depressants very little difference could be claimed in control studies. Thus the choice of a drug in that class is based mainly on empirical knowledge. Factors like the degree of anxiety, a drug history, successful treatments in a family member, unwanted side-effects and organic diseases of the patient become very relevant to the decision. In addition, each treatment has its own set of adverse reactions and drug-drug, drug-age, drug-gender, drug-state-of- health interactions and are often complicated by non-prescription remedies : alcoholic beverages and illicit drugs.

## 3.3. The Knowledge Engineering Process

In this section we discuss the basis of our knowledge engineering method, describe how we used it to develop our system and assess its "time efficiency".

Considering the availability of our experts, their level of expertise, and the short amount of time we had to build a reasonably performing program, we tried to devise a new knowledge engineering method :

(1) We used our basic medical knowledge to structure the problems, reviewing textbooks, **major** psychiatric publications and discharge summaries. It allowed us to use the residents' time most efficiently asking them precise questions in order to clarify obscure or controversial issues or to check the validity and usefulness in **real** clinical situations of the knowledge extracted from the literature. "Directly" working on well structured problems **(i.e** checking a proposed structure) with our experts -- instead of classically using them to structure the problems -- saved a lot of time and made their task far easier.

(2) We took advantage of the availability of several "experts", working with each of them **on** different portions of the knowledge base. It was then our responsibility to assure its general coherence. In any case the residents being trained in the same hospital, by the same senior psychiatrists, usually shared common views. So the coherence of their knowledge had never been **a** real issue. Of course this also implied that our ES would reflect the particular subjective opinion **of** the local psychiatrists (indeed very "drug oriented"), as a medical textbook reflects the opinion of its authors. Let us describe the building of the Blue-Box program to illustrate how we applied these knowledge engineering principles.

After spending ten days on the ward, getting used to the place, the people and the local procedures, we started our work by a general presentation to introduce AIM and our particular subject (the development of an ES) to the ward staff. Three residents decided to work with us. We divided the task in three major phases : (1) Assessment of the clinical problem (2) Eliciting the decision making knowledge (3) Encoding the knowledge into rules.

### 3.3.1. Assessment of the Clinical Problem

The initial questions we needed to answer in order to define the EMYCIN structures **were : "What is** the clinical strategy to follow in front of a depressed patient ? How can we divide this process in clear different steps ? What data do we need to evaluate and treat a depression ? In what order are these data elicited ?". We sketched tentative answers to the questions studying the discharge summaries **of** depressed patients treated on the ward during the past months. Then we checked our hypotheses asking the residents to describe "a telephone consultation with a general practitioner consulting them about a patient presenting with a chief complaint of depressed mood". The description fitted into the structures we had defined : we "translated" them into the EMYCIN format, determining the parameters - independently of any rule -- and the goal rule of our system.

At this point, we realized that to be considered intelligent on the ward, our program should both respect the medical format (eliciting the data in a certain order) and provide the physician with the ability to report the facts he considers important (even if they are not used internally). These considerations determined the clinical strategy of our program, encoded in the goal-rule and the parameters. The information is requested in the fashion a GP would present it to a consultant : presentation of the patient, chief complaint and history of the present illness, past history of similar episodes and their treatment, personal and family psychiatric history, concomitant medical problems and their treatment. The program then uses these data (I) to rule out the possibility that the depressive mood of the patient is due to a medical cause or to another possible psychiatric disease (2) to determine the type of depression and the class of drugs it should respond to (3) to assess the suicidal risk of the patient (4) to select and prescribe a therapeutic regimen (e.g, a specific drug among a class, a combination of drugs, an ECT...).

### 3.3.2. Eliciting the Decision Making Knowledge

Once we had defined the data we needed (the EMYCIN parameters), our strategy of questioning, the sub-tasks to perform (our goal rule) and the initial data to request, we had to write the rules in which the medical decision making knowledge of the program would be stored. To do so in the shortest amount of time, we selected the drugs our program should know and the special cases it should be able to handle and asked the residents to "reason backward". We gave them a list of the 'available clinical information (the parameters defined during the previous phase) and we asked them to tell us which findings a patient should present (i.e which values of the parameters they expect) in order to consider a particular class of drug ; and the, which findings they expect to prescribe a specific drug in this class. In other words we asked them to build the ideal candidates for the different possible treatments. The residents reported that they found it far easier to reason this way than to have to structure the domain for us. We had provided them with building blocks and the goals to reach. All they had do was to tell us how to arrange the blocks. On the other hand, to reduce the possible sterilizing effects of this constraining method, while listening to them "thinking aloud", we were very attentive to the context of their explanations and their terminology. Even when we knew (from the review of the literature) tentative responses to the questions we were asking, we avoided any interference with their reasoning and encouraged them to depart from the original structure we had defined. At the end of the session, we tried to see if the new rules could be expressed within the frame work of the existing parameters. If not we created new parameters and added them to the pool of "building blocks" for the next session.

### 3.3.3. Encoding the Knowledge into Rules

Since the residents directly reasoned in the rule format using EMYCIN parameters, the translation of the elicited medical knowledge into EMYCIN rules was usually straightforward. We just had to type these parameters and their desired values in the EMYCIN pseudo-Lisp. For instance, the third premise of Rule 027 displayed above was coded as: (SAME CNTXT SLEEP-PATTERN (ONEOF LATE-EVENING-INSOMNIA MIDDLE-NIGHT-INSOMNIA)); the conclusion: (CONCLUDE CNTXT TYPE GRIEF-REACTION TALLY 700). The certainty factors were first checked and refined on paper. We expected this task to be very difficult. The pleasant surprise was to find that the residents did not have much difficulty in determining and adjusting the certainty factors. In some rare cases we needed to hand-simulate the certainty factor of a conclusion reached by the program yield by a cluster of findings -- triggering a cluster of rules. It generated this kind of question : "A man in his sixties who has made a past suicidal attempt using a lethal method and who has a poor family support would be considered at high suicidal risk but would not be hospitalized against his will, unless he has a clear plan. Does that make sense ?".

Then, we ran the program and made the different corrections suggested by these trials : the residents were particularly sensitive to the way the questions were phrased (especially appreciating "their" questions) and were willing to input accurate clinical information about organic diseases, personality disorder, past history...even when they knew that it would not be used internally. Of course, all their suggestions were not limited to presentation issues; some rules and certainty factors had to be corrected and some new rules had to be added. But, to our surprise, these modifications were not cumbersome and the basic structures we had predefined were adequate and generally we did not need to change them.

During this third phase, we started to train one of the residents. We showed him how to run the program and also introduced him to EMYCIN and the way we use it, so that he would be able to modify and expand the knowledge base after we leave. We considered the education as an integral part of the knowledge engineering process. In the same way, we end of our clerkship by a forrnal presentation of the program to the ward staff. We discuss their reactions in front of the program below, after the display of a sample consultation.

### 3.3.4. Time Efficiency

Before displaying a sample Blue-Box consultation, we would like to provide an approximation of the time spent to develop the program. We spent between one and two hours per week with the residents, during five weeks (total: about 22 hours). The first four hours were spent to define the clinical data and the strategy. Then it took about ten hours to specify the knowledge later encoded in the 200 rules, and eight hours to check rules and to test the program with the residents. Ten hours were also spent to train one resident (five hours explaining to him how EMYCIN works and how we have used its different structures: five remaining hours practicing). Besides, during five weeks (we did not work on the program during the first week and had had to stop working during two weeks) each of us spent an average of four hours a day, encoding in the EMYCIN format the knowledge elicited from the residents, checking the rules and testing the program (total : 280 hours). Altogether it required 300 man-hours to produce a 200 rule system yielding a rate of 1.5 man-hour per rule. This estimation is consistent with the estimation of 1 man-hour per rule done during the development of a previous system called CLOT (7), a computerized consultant for diagnosis of bleeding developed using EMYCIN.> We would like to stress that this rapid development was possible because we used EMYCIN. We were able to focus our efforts on the knowledge acquisition process and not on programing (in fact, building this ES we did not write ONE line of LISP code).

# 4. RESULTS

The most concrete "result" of our work is of course the program. So, we present here a sample consultation followed by a few examples of general questions about depression you can ask to the QA module. All the answers input by the users are in capital letters and follow the prompt: "**". The user can answer U or UNKNOWN to any question. The case presented is a real one, inferred from a comprehensive discharge summary. Comments in italics have been added for a better comprehension.

EMYCIN 16-May-83...

(...)

Do you want to enter Rules, Parms, Go, etc (? for help): Go

----------PATIENT-1----------

1) What is the name of the patient ?

** MR. BLUE

2) What is the age of Mr. Blue ?

** 60 YEARS

3) What is the sex of Mr. Blue ?

** MALE

4) Please, characterize the depressed feelings of Mr. Blue with any of followings : sadness, hopeless, helpless, anhedonia, anergia, anxiety guilt-feelings, worthless, other or none

** ANHEDONIA SADNESS ANXIETY

6) What is the duration of this depressive episode ?

** 2 MONTHS

6) Is the appetite of Mr. Blue normal, increased or decreased ?

** DECREASED

7) Weight variation ( - or + pounds) since the beginning of this depressive episode ?

** - 10

8) Please describe the sleep-pattern of Mr. Blue:

** LATE-EVENING-INSOMNIA MIDDLE-NIGHT-INSOMNIA EARLY-MORNING-INSOMNIA

9) At what time are Mr. BLue's mood and/or activity more depressed ?

** UNKNOWN

10) Please characterize Mr. Blue's psychomotor activity:

** VERY-RETARDED

11) Does Mr.Blue have some somatic complaints, hypochondriasis and/or somatic delusions (other or none) ?

** SOMATIC-DELUSION

12) Please list (if any) the psychotic symptoms of Mr. Blue :

** PARANOID-DELUSION

13) Are there any PAST episodes of affective disorder you would like to mention ?

** NO

14) Do you have any information about the psychiatric history of Mr. Blue?

This includes any known psychiatric disease (Axis 1), personality disorders (Axis 2), or past suicidal attempts.

** NO

16) Is there a history of psychiatric disease, or of suicidal attempts Mr. Blue's family ?

• * YES

----------FAMILY-PSYCHIATRIC-HX-1----------

16) Please list any psychiatric disease (Axis1 diagnoses) known in the family of the patient :

** DEPRESSION

17) Is there an history of suicide in the patient's family ?

** YES

18) Does Mr. Blue have any organic disease you would like to mention ?

** YES

**----------ORGANIC-DISEASE-1----------**

19) Please , name the disease **?**

● * LEFT-BUNDLE-BRANCH-BLOCK

20) **Howlong** had the patient been known to have Left-bundle-branch-block **?**

** UNKNOWN

21) Is there any other organic disease you would like to mention **?**

● ✉ ☠⚐

22) Are there any drugs prescribed for Left-bundle-branch-block, you would like to mention **?**

● ✉ ☠⚐

23) Does Mr. Blue have some genital symptoms (loss of libido...) **?**

** NO

*[The program has asked basic questions about the patient it has an idea about the type and severity of depression (psychotic, biologic, atypical, grief-reaction ...). It will now try to assess the suicidal risk.]*

24) Is Mr. Blue suicidal ? (If you answer U for unknow, I will try to assess the suicidal risk myself).

** u

26) Does Mr. Blue have access to fire arms **?**

** UNKNOWN

26) Please grade the quality of the family support of Mr. Blue

** POOR

27) Please grade the severity of the Left-bundle-branch-block in terms of its emotional impact on the patient's **life**

** NONE

28) Has Mr. Blue faced severe psycho-social stressors (cumulative

loss history...) over the past years ?

** YES

29) What is the cultural background of Mr. Blue

** WASP

30) Please characterize the suicide plan of Mr. Blue

** VAGUE

31) Has Mr. Blue accepted to make a No-Suicide decision ?

● ✉ ✍

*[When the user type a "?" he can get a more explicit question, possibly with some literature reference.]*

The No-Suicide decision (as defined by Robert C. Drye et al. in Am. 3. Psychiatry **130:2,** February 3) consists of having the patient pronounce the sentence : No matter what happens, I will not kill myself accidentally or on purpose, at any time.

Expected responses are : YES or NO

Enter HELP for list of user options

** UNKNOWN

*[The evaluation of the suicidal risk is over. Using its estimation of the depression type and the suicidal risk, the program has chosen one or more general category of treatment (ECT, tricyclic anti-depressants, MAO Inhibitors, Neuroleptics, psychotherapy,...). New questions are asked to discriminate more accurately between these categories and/or between elements of one category (i.e different tric yclic anti-depressants).]*

32) You mentioned that a patient's relative has had a depression. Please single out a drug that has been successful in treating

this depression :

** UNKNOWN

33) You said **earlier** that Mr. Blue presents some anxiety. Please characterize the degree of anxiety ?

** MILD

34) Is the treatment of the patient planned in California **?**

** WHY

*[Instead of answering the question, the user can always ask "WHY" i.e "why is the program asking this particular question at this particular time ?" EMYCIN will explain why it is asking the question by displaying the rule that has generated it.]*

(i.e. WHY is it important to determine whether the treatment of the past is planned in California ?)

(1.0) This will aid in determining the types of treatment recommended.

It has already been established that

(1.1) ECT is one of the types of treatment recommended

Therefore, if

(1.2) there is evidence that the treatment of the patient is not planned in California, and

(1.3) Mr. Blue has given his consent for Electra-Convulsive Therapy

then

it is definite (1.0) that ECT is one of the types of treatment **recommanded**

(RULE087)

(back to question 34...)

** YES

36) Are you over with CHANGEs, Examination of parameters and SAVE ?

** YES

*[The reasoning process is over and the program now dumps its conclusions about (1) whether the patient needs to be hospitalized, (2) whether special suicidal precautions should be taken (if they are not, nothing is specified) (3) the different treatment procedures recommended.]*

*** Considering what you have told me about your patient, my recommendations would be :

Mr. Blue needs to be hospitalized (.48)

Even though I know it is difficult to prescribe an ECT in California, the severity of the depression being what it is. I would keep ECT in the back of my mind, and start the legal procedures. (.4)

You could consider treating Mr. Blue with MAPROTILINE, starting with a total daily dose of 160 mgs given PO, TID. Increase the dose by 60 mgs every other day until you reach the maximum of 300 mgs (.44)

Remember to warn your patient about the possible unwanted effects. If the side effects are intolerable, consider discontinuation of the treatment. If your patient does not show any improvement after 6 weeks of this treatment, consider switching to another antidepressant; before doing so you can check the plasma concentration,

Since Mr. Blue presents some psychotic features, you could add HALOPERIDOL to the treatment, starting with a daily dose of 1 mgs given PO, TID. Increase the dose by .5 mgs every other day until you reach the maximum dose of 2 mgs (.43)

Do you want to enter Rules, Parms, Go, etc (? for help) : Go

*[Now, instead of runing a consultation we are going to use the QA (Question-Answering) module to ask questions in plain English to the program about its knowledge. Of course the set of possible questions is not predetermined.]*

Special options (type ? for help) ;
** QA

You are entering the Q-A module module

** WHAT ARE THE POSSIBLE TYPES OF DEPRESSION ?
The legal values for, the type of depression are : BIOLOGIC, PSYCHOTIC, BIPOLAR, GRIEF-REACTION, CHARACTEROLOGICAL, ATYPICAL, ANXIOUS, PRIMARY-ANXIOUS, DRUGS-INDUCED, DISEASE-INDUCED

*** HOW DO YOU CONCLUDE THAT A DEPRESSION IS BIOLOGIC
The following rules conclude about :
the likelihood that biologic is one of the type of depression

4, 6, 6, 8, 9, 7, 10, 11, 12, 13 ,14, 98

The parameters most frequently used in these rules are :

. the weight variation of the patient

. the duration of this depressive episode

Would you like a summary of ALL the parameters used by these rules ?

• * YES

Parameter (rules with this parameter)

the weight variation of the patient (6, 8, 7)

the duration of this depressive episode (6, 8, 7)

the sex of the patient (12, 13)

the age of the patient (12, 13)

the depressed feelings of the patient (4)

the appetite of the patient (6)

the organic diseases (98)

the sleep-pattern of the patient (90

the patient's psychomotor activity (10)

the onset of this depressive episode (14)

the various somatic symptoms of the patient (11)

. the duration of the organic disease of the patient (98)

the highest level of functioning prior to this episode (14)

whether the patient has some genital symptoms (loss of libido...) (12)

the severity of the organic disease of the patient (98)

whether the patient has some genital symptoms (loss of libido and/or menstrual disturbances) (13)

Which of the rules do you wish to see ?

**

** HOW DO YOU USE THE SLEEP PATTERN OF THE PATIENT ?

The following rules use :

the sleep-pattern of the patient

9, 28, 70, 80, 81, 26, 36

The parameters concluded by these rules are :

the type of depression

the types of treatment recommended

the severity of the depressive episode

Which of the rules do you wish to see ?

** 28

RULE 028

-------

(This rule applies to patients, and is tried in order to find out about the type of depression)

If : 1) A: The appetite of the patient is increased,

B : The patient has gain a significant amount of weight, or

C : Hypersomnia is one of the sleep-pattern of the patient,

2) Anergia is one of the depressed feelings of the patient, and

3) Anhedonia is not one of the depressed feelings of the patient

Then : There is strongly suggestive evidence (.8) that atypical is one the type of depression

Justification :

It has been reported that depression with such clinical features should be treated with **MAOI** and not with TRICYCLICS. That is why we have created a type of depression called ATYPICAL.

Literature :

WHITE and SIMPSONS ; Archives of General Psychiatry ; Sept. 81

*[Rules can be associated with justifications and/or literature references. the followings are other examples of questions the QA could also answer.]*

• * HOW DO YOU CONCLUDE THAT THE PATIENT IS AT HIGH SUICIDAL RISK ?

**(...)**

• * HOW DO YOU DECIDE THAT AMITRIPTYLINE IS A POSSIBLE TRICYCLIC ?

**(...)**

•* HOW DO YOU USE THE ORGANIC DISEASES OF THE PATIENT ?

**(...)**

• * WHAT DO YOU KNOW ABOUT THE SEVERITY OF A DEPRESSION ?

**(...)**

We hope this demonstration has given the reader a general view of how the program works and what it does. What we would like to present here, is not only an evaluation of the program, but an analysis of the implications of the task of Knowledge Engineering itself. Furthermore, it should be understood that it is difficult to present a classical evaluation of a system that has not been designed with the same original goals as others of its kind. Blue-Box is based on the knowledge of a specific ward and design for it. We make no claim about its potential use on some different clinical setting.

**4.1. Evaluation of the Program**

Traditionally, an expert system is evaluated along three criteria: *performance, acceptance and educational potential.* Recently another criterion has been raised (8) : the amount of endeavor required to build the program.

To assess the performance, the subject the system is dealing with should'be kept in mind. The narrower the domain, the easier it is to reach a respectable level of performance. Another important consideration concerns the qualities of the domain. When the clinical problem can be reduced to **a** physiological model easily formalizable, the performance of a system is greatly enhanced -- see for instance CASNET (9) which reasons about Glaucoma. On the other hand, when the task is mainly based on empirical knowledge, it requires a large number of rules and extensive testing with **a** wide variety of clinical situations before a minimal competence can be claimed.

Blue-Box has not been submitted to a formal comparison with experienced clinicians. Still, when we ran cases using discharge summaries of former patients of the MHCRC we found a general consistency between the advice of the program and the discharge plan. At the end of the clerkship, we also made a formal presentation of the program to the ward staff. The reactions were positive even though some residents expressed their concerns about the hypothetical use of such a program by general practitioners or by internists to treat depressed patient without consulting psychiatrists. They deplored that such a practice could possibly deprive patients of the more subtle and human psychiatric help they deserve. But everyone agreed that this' program could at least be an interesting teaching and research tool. It was underlined that the program would be particularly useful to teach to the beginners in the field how to use theoretic knowledge in a structured and rational fashion. The residents who worked with us expressed their desire to evaluate the program in a more systematic way and to try to expand it ; two of the residents who did not work on the program expressed their interest in taking part in this process.

Finally, the endeavor required to implement the project, is really minimal if compared to other systems. It involved only minimally qualified people : two medical students with no formal training in AI or Knowledge Engineering and three psychiatric residents over a two months period. At the beginning, none of the system builders had had any previous acquaintance with EMYCIN, and the medical students had only a basic understanding of the specific medical domain. Moreover, the clinical activity of each of them has been fulfilled normally during the entire period of the experiment.

### 4.2. Evaluation of the Knowledge Engineering Task

The benefits of the experiment have to be assessed differently for the residents who served as pseudo-domain experts and for the medical students, pseudo-knowledge engineers.

#### 4.2.1. The Resident's Standpoint

First, it seems important to stress that the idea of becoming involved in the development of a computer program dealing with an important aspect of their daily activity has been accepted by the residents. Among the three who volunteered for this project only one had a limited previous experience with computers and none with Artificial Intelligence. One of the two remaining had even mentioned jokingly a "computer phobia" but was willing to undergo what he himself called a "behavorial therapy".

Globally, the residents were willing to formalize their knowledge and to gain insight in their clinical problem solving process. This project gave them a rare opportunity to "stand back" and spot their own weaknesses as welt-as powerful heuristics. The review of the literature that was pursued in parallel also raised new issues and data they were not always aware of. Overall, the task they found themselves working on was comparable to that of the initial phase of writing a book on the subject, which is an all too rare opportunity in clinical practice. An issue which became apparent, however, was how much easier it was to develop an expert system than to write a book. None of them dedicated any extra-time to this research, outside of the knowledge engineering sessions. Moreover, all the formalization did not need to be done at once, since the ideas could be tested on the computer. By testing them, all that was not useful became obvious, and what was not precise enough could be refined further depending on the performance of the program. Thus, the residents perceived their task as both easy and rewarding.

By the end of the clerkship, this experiment increased their awareness of the possible impact of computers on their practice. It also aroused their interest. All of the three residents separately decided to buy a home computer and to learn how to program. Two of them were seriously considering investigating further the domain of Artificial Intelligence in Medicine, possibly in association with the researchers of Stanford University. Finally, their encounter with this "new area of Medicine" abolished most of their fear that computers would be designed to replace them.

#### 4.2.2. The Medical Students' Standpoint

For the medical students involved in the project, trying to understand and teach to a computer program a domain of medicine largely new to them, was a great incentive toward learning and discovering. It is one thing to dig into the literature with hope that it will be useful in a remote future (if your memory is good...), it is much more rewarding to formalize it, find where it fits in regard to previous knowledge, transfer it to the computer and look at it "in action". It gives an immediate feed-back about the usefulness of new information. It presents the student with a concrete representation of how he, as a physician, will look (or think) when he will know all of that.

This type of formalization, sometimes abstract, is an aspect of medical thinking all too often neglected in favor of acquiring more and more facts. When practiced on a ward, while learning the concrete task of managing patients, it provides a broader understanding of the problems and situates the clinical case at hand in its proper general category that the student will later be able to recognize. In addition, implementing their knowledge on the computer has also been an opportunity for the students to "debug" their clinical problem solving techniques before inconsistencies and "loose associations" deeply rooted in their memory.

**4.2.3. What we did not (but could have done) and what could be expected from it**

Among the numerous things that we could possibly have developed further, based on the quality of the ward's staff and potential of EMYCIN, we would like to describe a few here and sketch their possible implications.

First the nursing staff has not been involved in the project, which we think deprived our team of a great amount of useful information. This was principally our fault since we did not make any formal offer to participate to the nurses and social workers. Giving it some afterthought, the nursing staff could certainly have provided some relevant comments about the "peripheral knowledge" for example. This involves phrasing of questions, ordering of questions, quality of reprompts, clarity of justification for rules etc... During the final presentation of the project, the nursing staff appeared really enthusiastic about the program and willing to investigate and participate in this type of research. They also considered that the program could be a useful introduction to the ward for new nurses and possibly a "person to turn to" in order to understand better orders given by busy physicians with no time to explain their reasoning process.

Another aspect of the project we have not been focusing on concerns the extensive justification of rules with references to the literature. We could only give "a taste" to residents of what a system with such a documentation would be like. They thought of it as a new way of going through the literature. Their idea was not really to replace the literature review but to add a new dimension to it. Every time a new idea, concept or "rule" is found in a journal it could be integrated in the reasoning chain of the program. Thus the goal of the quest is not only "What's new" but "How and when can I use that". By running old cases with the new, updated, knowledge base, the implication of the new rule can be assessed at every level of the decision making process. This simulation provides an analysis of new material before it might be relevant to solve a clinical case. It actually provides a way of putting new data into practice right away.

At the student level, the main application we could foresee for Blue-Box, but did not investigate, is to provide a new clerk with a rapid acquaintance to the ward. Typically, every clerkship is characterized by a limited number of procedures that will be performed repeatedly during the time spent on the ward. Unfortunately, it often takes the entire length of the clerkship before students are sufficiently familiar with these procedures to become interested in more rewarding aspects of the domain. By interacting with a program like Blue-Box, the student could learn more rapidly what data should be elicited during the interview of a ward's patient, what is the terminology commonly used, what are the most widely prescribed tests, what are the typical management methods, and so on and so forth. Thus the student might learn not only new procedures but how to justify them. When confronted with the program, he will be asked numerous questions relevant for the case he is working on. If some data are missing because his initial work up is not complete, by asking "WHY" when the computer makes a request, the student would be told how the piece of information he did not collect could be used by the program. He could also take notes of the missing data, go back to his patient and find out what they are. Thus the interaction with the computer would resemble to some extent what it currently is with the faculty of the ward, but, without the stress.

## 5. CONCLUSION : ONE MILLION EXPERT SYSTEMS

We have presented how we tried a new approach for the development of medical expert systems and how this method might help solve the traditional problem of acceptance. We would like to conclude by assessing whether our experiment can be generalized.

First, this type of project, like many others, requires a certain amount of motivation. Physicians have always had the reputation of being afraid of computers. One could think that they might be very reluctant to be involved in the task of knowledge engineering. In fact, we think that the incentive toward AIM research among physicians is largely underestimated. They are only afraid of what they don't know, but presented with the concrete task, they find it pleasant and exciting. We also speculate that some of the reluctance toward computers displayed by the medical community might stem from those clinicians who consider themselves too old to get involved with computers and yet too young to escape from their growing wave. Thus, the young residents with whom we could work might well be more prone to this type of reasoning and more willing to be involved than the previous generation of physicians.

Another important issue, correlated to the latter, is the consideration of the medical community for this type of research. If physicians are to be involved in the development of expert systems, there must get some positive feedback from their peers. Their work must be acknowledged as valuable and useful. This is certainly even more true at the level of medical students, when looking for models to follow. Thus, the more developed and recognized the field of AIM will be, the easier it will be to initiate projects in different hospitals.

Finally, we think three important practical conditions should be put together before any experiment of the kind we have described can be launched in a medical center:

- The students in charge of the knowledge engineering should have a minimal background in computer programming. By minimal we mean the ability to interact with a computer, learn his way around the system, and understand how a computer program works. This corresponds approximately to the level reached after an introductory course of computer science in a college, or even in high school.

- The quality of the computer environment is another important aspect to consider. Blue-Box has been developed on a DEC 20 computer, and even though it would probably be technically possible on a smaller machine, it would not have allowed us to work as comfortably or as quickly as we did.

- The software used, or Knowledge *Acquisition System,* in this case EMYCIN, is not commonly available. Without such a tool to start with, the Expert System development is focused on programming, to the prejudice of the medical knowledge acquisition. This burdens and lengthens the task of the knowledge engineers and decreases the quality of communication with the medical "experts" ; they do not work on the same thing.

  Moreover, even though EMYCIN has efficiently supported our efforts, it also entails certain limitations. Newer and more sophisticated programs could further facilitate knowledge engineering quality of the interaction with the final product. Indeed, we believe that further development of medical knowledge acquisition systems should be an area of choice for the AIM research in the coming years.

These are, presently, concrete but not insurmountable limitations to the generalization of the approach. Nevertheless, the trends in the field of computer sciences seems to indicate that we shall be able to overcome them in the near future. We hope that our study will have helped demonstrate that there is no theoretical limitation to the development of one million expert systems.... and more.

## Acknowledgements

The authors wish to express their deepest appreciation to the following people : Dr. Bill Gekeler, Dr. Kerry Kravitz and Dr.Steve James, the residents who helped us to develop Blue-Box with constant attention and patience. Dr Steve Stahl who accepted our last minute request to attend his clerkship and later introduced us to the moving world of Psychopharmacology. The entire ward who warmly received us and helped discover the other aspect of psychiatry. The patients whose kindness and understanding enlightened our stay at the MHCRC.

Dr. Robert Blum who introduced us to the Stanford Computer Science Department. Michael Walker gave us constant encouragement and useful critics during our stay. Michael Rychener and Duvvuru Sriram reviewed the final draft of this paper. And, most of all, Professors Bruce Buchanan and Edward Shortliffe guided the project, and stimulated most concepts developed in this paper. We wish to express our heartfelt thanks to them for without them our exposure to and experience in Knowledge Engineering in Medicine were not conceivable.

# I. Appendix :Technical problems and some solutions

We report here on the techniques we used to support some conceptions of user interaction, representation of time and knowledge that evolved in the course of the knowledge engineering, process within the framework of EMYCIN.

### 1.1. User-interaction

As mention in the previous section, we thought that limiting the number of requests from the user is a major design goal in the construction of an expert system. During the development of Blue-Box, we tried to respect two conditions :

1. The content of the requests must be perceived by the user as important and relevant to the task at hand.

2. The number of requests must be the minimum that is consistent with maximum utilization of medical inferences.

#### I.1.1. Controlling the format of the consultation

To allow the user to enter data in medically relevant format we made an abundant use of contexts. Some context correspond specifically to a section of the "case presentation". Examples of contexts are :

- The patient : includes all information about the patient's identification (name, age, sex,...), the "Chief complaint" and the "History of present illness" as the mood, the weight variation, the sleep-pattern,...

- The "Past History of affective disorders", is the patient chronically depressive or is this the first episode of this type ?

- The "Medical History" which covers any medical disease and associated drugs.

- The "Psychiatric History" in which information about a known psychiatric disease, a personality disorder, the best level of functioning and other data are contained.

- The "Family Psychiatric History", includes diseases and trials of treatments for similar depressive episodes, as well as the quality of psychological support provided to the patient.

To each of these contexts is associated a list of questions to ask as soon as the context is generated ("Initial data"). This representation, along with the explicit ordering of context generation through the use of a “goal rule",allows a very good control of the global format of the consultation. It provide a consultation where the age and sex of the patient are asked right after the name and among the first questions (Mr Q. is a fifty years old male...).

This highly structured interaction has not generally been encouraged by previous designers of consultation programs. They argue that, left with a maximum degree of freedom, the program will tend to ask questions only when relevant or potentially useful (when required by the premise of a rule).

But, even if this is true in a number of cases, it still seems intolerable to a physician to see the program ask what is the age of the patient only when evaluating the kidney function. It also seem “stupid” to the same physician to see the program going back and forth between questions on the family history and questions used to characterize the clinical profile, at least during the initial phase of the consultation.

Another interesting argument concerns the answers to “why questions” during the consultation. If an information is required by the program merely because it is part of the initial data list, a why question will not allow the user to see how this datum is used in the reasoning process.

Even this has not proved to be a-significant tradeoff for two reasons. First, the reason why the question is asked is frequently obvious to the user during this phase of the interview. Even the unsophisticated user understands that data such as the presence of an organic disease are of general usefulness when evaluating a depressed patient. Secondly, why questions provide only a snapshot of the on going evaluation of the patient which bears very little information on how the answer will be used internally elsewhere in the reasoning process. E.g, even if a question about a recent weight loss was asked to evaluate the premise of a specific rule concluding about the likelihood of one type of depression, the answer to a why question would not show how this parameter is used to conclude about other types of depression, later in the reasoning chain. Indeed, the user should rather invoke the QA module to ask “How do you use the weight variation of the patient ?" to gain a better insight in the pertinence of the question.

#### 1.1.2. Accepting multiple answers

A practical way to limit the number of requests is to create “Multi-Value Parameters” asked only once and that can contain a great deal of information. One of our multi-value parameters is the “mood of the patient”. The interaction looks like that :

4) Please, characterize the depressed feelings of **Jaime** with any of followings: sadness, hopeless, helpless, anhedonia, anergia, anxi guilt-feelings, worthless, other or none

** SADNESS ANHEDONIA ANXIETY **(.3)**

Each answer is then used by different rules, independently of the other.

### 1.1.3. Refining answers when appropriate

The program had to be able to ask more questions about a particular finding in some instances and to ask them in the appropriate order. A typical example concerns "anxiety". If the patient is said to be anxious it is pertinent to ask what is the degree of his anxiety (mild, moderate, severe) and to check if he has been anxious prior to this episode of depression.

To perform this task correctly, we often created multiple parameters (one for each of the information) which are asked to the user only when the "parent parameter" is known (in this case "anxiety"). We used the order of premises in rules to control the order of questions.

### 1.1.4. Adapting questions to the level of the user

In the ideal interaction with the user; all questions should be perfectly adapted to his level of sophistication. This involves three aspects of the interaction. First, the terminology employed should match what is commonly understood at his level of medical sophistication. Second, the questions asked should take into account the potential ability of the user to make valid clinical decisions. Third, the program should allow the user to report spontaneously data he knows are relevant, requesting them or double-checking only when they are required for an important decision, and have not been provided.

1. **Sophistication of terminology** : In many instance-we used the REPROMPT feature of **EMYCIN** to provide the user with a more detailed question or to explain the terminology used in the original prompt. Here is an example of a user request for more details:

```
43) Has Jaime accepted to make a No-Suicide decision?
** ?
The No-Suicide decision (as defined by Robert C. Drye et al. i
Psychiatry 130:2, February 73) consists of having the patient pr
the sentence: "No matter what happens, I will not kill myself ac
or on purpose, at any time."
Expected responses are: YES or NO
Enter HELP for list of user options.
** UNKNOWN
```

2. **Let the user make conclusions** : In some instance, the user "knows" the answer to a high level question (e.g "Is the patient suicidal ?") and confronted to a series of question might become irritated. In order to prevent this type of situation from happening too often,

we chose to have the high level question asked first. If the user answers “Unknown”, then the system tries to find out the answer by itself, asking all the questions relevant to this task.

3. Let the user enter data spontaneously : Some data apparently remote to the task of the program are in fact very relevant. The most sophisticated users who know this, will want to enter them at what seems to be the most appropriate point of the interview. For example, before a tricyclic anti- depressant is prescribed, it is important to check that the patient does not currently have a cardiac abnormality (Bundle Branch Block, Coronary Artery Disease,...). A psychiatrist familiar with this fact reports this when describing the medical history of the patient. He would indeed be very frustrated if not allowed to report it early on during a consultation, being irrationally afraid that the program is “missing” a very important piece of information.

   In Blue-Box, the user is offered to report any organic disease he considers relevant to the problem during the “Medical History” phase of the interview. Later, if no cardiac problem has been mentioned, and tricyclic anti-depressants are warranted, the program double-checks by directly asking the user if there is any cardiac abnormality.

```
46) When I-asked you whether Jaime had an organic disease, you d
    mention any cardiac problem. Before prescribing a tricyclic
    like to be sure of that. Does the'patient have a cardiac di
```

### I.1.5. Creating fuzzy thresholds

Looking at user-interaction from the other standpoint we tried to make an optimum use of the information provided. A frequent problem encountered when processing symbolic information is the inability of thresholds to represent decision points. An example of this is the utilization of the age of the patient. Very often, it is not barely the age the clinician is interested in but whether the patient is “an elderly”. Unfortunately this is not always equivalent to "age > 60”. On the other hand it is not conceivable to ask a question like “Is the patient an elderly ?". This would make the consultation much too cumbersome. In cases like this one we used other information already entered to conclude about the value of the parameter. For example, a rule to conclude about whether the patient is an elderly is :

```
RULE094
-------
If: 1) The age of the patient is between 55 years and 60 years,
       and
    2) The highest level of-functioning prior to this episode is
       poor
Then: There is weakly suggestive evidence (.4) that the patient is
       an elderly
```

At the same time this increases the quality of the english translation of rules displayed in the QA module or when answering to a why question? It is indeed often more meaningful to say "the patient has had a significant weight loss recently" than "the weight variation of the patient divided by the duration of the depressive episode is less than -10 pounds per week".

In EMYCIN the formalism of Certainty Factors (see [Shortliffe 76]) also provides a graceful way to overcome thresholds, Depending on the value of a parameter in the premise, the parameter in the action part of the rule can take a given value with different certainty factors. For example, Blue-Box uses the following rules to conclude about the type of onset of the current depressive episode :

RULE01 6
-------
If : The duration of this depressive episode is between 2 weeks and 4 weeks
Then : There is suggestive evidence (.67) that the onset of this depressive
episode is acute ,

Thus the rules which use the type of onset to conclude about the diagnosis or reason about the best therapy are less dependent on an arbitrary value of the duration. In some cases we used both mechanisms (additional information and CFs) to play around thresholds (e.g for "Elderly") and obtained very satisfactory results.

1.2. Management of time

Once it has been decided that a pharmacologic treatment is warranted for the patient, the next problem is to choose a particular drug. Some of the most useful information then lies in the drug history of the patient. What drug has already been tried ? How well did it work ? If it failed, is it because of a low dosage, too short a treatment, intolerable side-effects, or because the patient simply does not respond to it ? Thus, at first, we thought the program should prompt the user to describe all prior psychiatric episodes and give a detailed drug history for all of them. But listening to case presentation and reading discharge summaries we found that physicians much rather talk of "a cyclic depressive state during the last 10 years, episodically treated by Amitriptyline and Imipramine which failed to yield any positive results.". Moreover, if questions were to be asked for each and every prior episode of psychiatric illness, a user might repeatedly be asked about medications given years ago, at other hospitals, concerning which little or nothing is known, and the procedure would quickly become more irritating than informative.

---

[5] The english translations of rules are also very important during the knowledge-engineering sessions with the residents where they symbolize the result of the interaction with the knowledge- engineers.

In Blue-Box we tried to find a optimal balance between no information at all and too much detail by simulating as closely as possible the type of descriptions used by physicians. One context includes information to describe the type of past episodes of affective disorders (PAST-AFFECTIVE-EPISODES). An offspring context of the latter includes information about a drug used at one point in association with past episodes of one type (PAST-AFFECTIVE-EPISODES-RX). Thus numerous episodes can be described by filling only one context (e.g Depressive episodes on a cyclic basis) or by filling more than one context (e.g "Depressive episodes on an acute basis" + "depressive episodes on a chronic basis", describes an outburst on chronic background). As many treatments as the user wants can be described in association with the different types of episodes. A typical interaction looks as follows :

14) Are there any PAST episodes of affective disorder you would like to
mention?
** Y

--------PAST-AFFECTIVE-EPISODES-l --------
15) Were these PAST episodes (or this PAST episode) DEPRESSIVE, MANIC
or MANIC-DEPRESSIVE? (Please, one type of episode only; if more
than one type, you will be offered to report it)
* * DEPRESSIVE
16) Would these past Depressive episodes be best described as CHRONIC,
CYCLIC, or (one or more) ACUTE?
** CYCLIC
17) Are there any other PAST episodes of affective disorder you would
like to mention?
* * *

--------PAST-AFFECTIVE-EPISODES-2--------
18) Were these PAST episodes (or this PAST episode) DEPRESSIVE, MANIC
or MANIC-DEPRESSIVE? (Please, one type of episode only; if more
than one type, you will be offered to report it)
• * MANIC
19) Would these past Manic episodes be best described as CHRONIC,
CYCLIC, or (one or more) ACUTE?
• * ACUTE

20) Are there any other PAST episodes of affective disorder you would like to mention?

** N

21) Did the patient receive any drug to one of the past Depressive episodes?

** Y

--------PAST-AFFECTIVE-EPISODES-RX-l --------

22) Please, generic-name of the drug? (If more than one, you will be offered to report it)

** AMITRIPTYLINE

23) What was the dose of amitriptyline?

** U

24) Howlong had the patient been taking amitriptyline to treat this PAST episode?

** 1 MONTH

25) Please characterize the response to the PAST amitriptyline treatment:

** NEGATIVE

26) Did the patient receive any other drug to treat one of the past Depressive episodes?

** Y

--------PAST-AFFECTIVE-EPISODES-RX-2--------

27) Please, generic-name of the drug? (If more than one, you will be offered to report it)

** DESIPRAMINE

28) What was the dose of desipramine?

** U

29) Howlong had the patient been taking desipramine to treat this PAST episode?

** U

30) Please characterize the response to the PAST desipramine treatment:

** INTOLERABLE-SIDE-EFFECTS

31) Did the patient receive any other drug to treat one of the past Depressive episodes?

** N

32) Did the patient receive any drug treat one of the past Manic episodes?

** N

This technique of representation is relatively handy to describe the drug history associated to past episodes of affective disorders but it it is not well suited to describe treatments the patient is currently on. We have not had the time to design a special structure to handle this case and for the moment the only solution available is to characterize the current treatment along with the previous ones.

### 1.3. Representation of knowledge

During the process of Knowledge Engineering we were confronted with three methods of reasoning, all used by the residents depending on the problem at hand : Reasoning with independent findings, reasoning with small definitional clusters, and reasoning with prototypes and thresholds. Each of them had to be implemented in EMYCIN and raised specific issues.

#### 1.3.1. Reasoning with independent findings

When trying to solve a well structured problem, e.g one with relatively few possible outcomes, the residents often used a combination of very simple rules, each dealing with a single finding and increasing or decreasing the certainty of one possible outcome. This method of reasoning **is well** illustrated by the problem : Is the patient suicidal ?

Typical rules used by the residents and thereafter by the program are :

RULE046
.......
If: The patient has not accepted to make a No-Suicide decision
Then: There is strongly suggestive evidence (.9) that the patient
is at high suicidal risk

RULE054
.......
If: There is a history of suicide in the patient's family
Then: There is weakly suggestive evidence (.4) that the patient is
at high suicidal risk

The formalism of rules associated with Certainty Factors of EMYCIN is particularly well suited to this type of knowledge representation. It has (obviously) been very easy to transfer the clinical rules obtain from the physicians into EMYCIN rules. As mention in section 4 the resident were surprisingly comfortable in assigning CFs to these rules. Moreover, when we started to run the first cases through the program, very little modification were necessary and the program assessed the suicidal risk relatively well from the outset.

### I.3.2. reasoning with small definitional clusters

In other instances, physicians find themselves willing to cluster a few findings in order to make **a** particular conclusion- When they feel that these findings should be strongly bound together and bear on a specific hypothesis, the transformation into a rule is again straight forward. For example :

RULE027

.......

If: 1) The patient has faced severe psycho-social stressors
(cumulative loss history...) over the past **years,**
2) The duration of this depressive episode is less than or
equal to 6 months,
3) The sleep-pattern of the patient is one of: **late-evening-**
insomnia middle-night-insomnia, and
4) The highest level of functioning prior to this episode **is**
one of: fair good
Then: There is suggestive evidence **(.75)** that grief-reaction is
one of the type of depression

### I.3.3. Reasoning with prototypes and thresholds

In some cases, the most natural representation the residents could give of their reasoning process consisted of a set of cues grouped into a "prototype" with which they try to match the findings of the case. If the match "more or less" corresponds with the prototype, they conclude that a hypothesis is confirmed. It is worth noticing that this method of reasoning is officially reinforced by the proeminent publication of the American Psychiatric Association : "Diagnostic and Statistical Manual of Mental Disorders (Third Edition)" (DSM III) [APA 80]. In this manual, which sets the standards for psychiatric terminology and diagnostic criteria in the United States, diagnoses are frequently presented as :

1. A fundamental symptom, characteristic of the disease.

2. A list of symptoms <u>among which</u> a certain number should be present.

3. Other miscellaneous conditions.

'This type of reasoning is not easily implemented in EMYCIN. It has always been very artificial (or cumbersome !) to break these prototypes in smaller clusters. The easiest solution is to individualize each finding in a single rule which increases by a small amount the certainty of the final diagnosis. But doing so the clinicians loose their understanding of how the program reasons and are much less efficient when asked to debug and refine the rules.

In conclusion, most of the clinical knowledge elicited during the knowledge engineering sessions could readily be represented in the formalism of EMYCIN, we felt that the task of the KEs could be greatly facilitated if, in addition to simple rules, EMYCIN could allow a representation for prototypes.

# References


1. Mulsant, B. and Servan-Schreiber,D. “A gentle introduction to Artificial Intelligence in Medicine”. Technical Report, Robotics Institute Carnegie-Mellon University (1982).

2. Shortliffe, E. and **Duda,** R. Expert Systems Research. Science. Vol. 220, pp. **261-267** (1983).

3. Reboh, R. “Knowledge Engineering Techniques and Tools in the PROSPECTOR Environment”. Technical report, SRI International (1981).

4. Van Melle, W. and Carlisle S. “The **EMYCIN** manual”. Technical report, Stanford University (1981).

5. Shortliffe, E. “Computer based medical consultations : MYCIN”. Elsevier (1976).

6. American Psychiatric Association. “Diagnostic and Statistical Manual of Mental Disorders”. Third Edition (1980).

7. **Bennet,** S. and Goldman,D. “CLOT : A Knowledge-Based Consultant for Diagnosis o Bleeding Disorders”. Working paper HPP-80-7, Stanford Heuristic Programming Project (1980).

8. Davis, R. Expert Systems : Where are we ? And where do go from there ? A/ Magazine. Vol. 3, pp. 3-22 (1982).

9. Weiss, S. M., Kulikowski, C. A., and Safir, A. A Model-Based Consultation System for the Long-Term Management of Glaucoma. *IJCAI 5,* pp. 826832 (1977).