RUTGERS
THE STATE UNIVERSITY
OF NEW JERSEY

PROCEEDINGS OF THE FIRST ANNUAL

A.I.M.

ARTIFICIAL INTELLIGENCE IN MEDICINE

WORKSHOP
JUNE 14-17, 1975

Theme: KNOWLEDGE BASED A.I. SYSTEMS

Sponsored by:
The Rutgers Research Resource on Computers in Biomedicine
Department of Computer Science
Rutgers University, New Brunswick, N.J. 08903

SAUL AMAREL, *Principal Investigator*
C. A. KULIKOWSKI, *Organizer*
N. S. SRIDHARAN, *Technical Director*
DEIRDRE SRIDHARAN, *Proceedings Editor*

CONTENTS

I. INTRODUCTION

AIM (Artificial Intelligence in Medicine) is a NIH supported national project devoted to the development and dissemination of AI applications in Biomedicine. The SUMEX-AIM computer facility at Stanford University is the major shared resource of the project. This facility is accessed by several research groups in the national AIM community via TYMNET and ARPANET. The Rutgers Research Resource on Computers in Biomedicine is one of the major projects in the AIM community. As part of its responsibilities the Rutgers Research Resource, directed by Dr. Saul Amarel, is sponsoring a series of annual AIM Workshops. The first Workshop was held at Rutgers University on June 14-17, 1975. Dr. C. Kulikowski was the Workshop Organizer; Dr. N.S. Sridharan was the Technical Director; Ms. P. Moore and Mr. K. Brown were administrative coordinators. The Rutgers Research Resource is supported by the Biotechnology Resources branch of the NIH, grant number RR-643. A description of the Resource appears in SIGART Newsletter, An ACM Publication, No.54 (Oct. 1975).

The Stanford University SUMEX/AIM Project is directed by Dr. Joshua Lederberg. Users of the SUMEX facility are divided for administrative purposes into two groups: 1) those at Stanford University School of Medicine, and 2) those elsewhere in the United States. The facility resources (computing capacity and consulting support) are allocated in equal portions to the two groups. As Principal Investigator for the SUMEX grant, Dr. Lederberg reviews Stanford medical school projects with the assistance of a local advisory committee. The governance of AIM includes the AIM Executive Committee and the AIM Advisory Group. The membership of these committees is given in Section VII. National users may gain access to the facility resources with the approval of the national Advisory and Executive groups.

The Workshop was designed to provide insight into existing and potential systems that apply methods of Artificial Intelligence to problems of biomedical research and health care. The attendees were selected from a broad range of investigators specializing in Chemistry, Psychology, Medicine and Computer Science. They were chosen in consultation with an advisory group of AIM investigators and with the approval of the AIM Executive Committee. The 1975 theme of "Knowledge-based Systems in Biomedicine" centered around discussions, demonstrations, and hands-on systems experience in

- medical modeling and decision making for diagnostic/therapeutic consultation;

- psychiatric simulation, psychological modeling, language analysis and common sense reasoning;

- biomolecular characterization of organic molecules on the basis of chemical analysis, protein structure determination and chemical synthesis planning.

No formal papers were prepared for the Workshop. Emphasis was placed on brief presentations of current AIM projects, followed by in-depth discussions of basic issues which underlie AIM activities. Most of the discussions took place in panels which were recorded. Section V of the Proceedings contain summaries of transcripts of five panels. Many of the key issues and concerns that came up in the Workshop are captured in these panel discussions.

Section IV of the Proceedings provides brief descriptions of the systems presented at the Workshop. The list of panel participants and their affiliations is given in Section III.

The Workshop participants were provided continuous access to several working application systems running both at the Rutgers PDP-10 and at the SUMEX PDP-10-TENEX. System access and hands-on experience proved valuable in the dissemination of AI applications in Biomedicine and will be a recurring feature of future Workshops in the series.

II. SCHEDULE OF THE FIRST ANNUAL AIM WORKSHOP

Held at Rutgers University, June 14-17, 1975

GENERAL SESSION (Saturday, June 14)

Morning Session:

8:30 - 9:00 Registration
9:00 - 9:15 Introduction to the Workshop
              (S. Amarel, Rutgers University)

        I. KNOWLEDGE-BASED SYSTEMS IN MEDICINE

9:15 - 9:30   MYCIN: Antimicrobial Therapy Consultation System
               (E. Shortliffe, Stanford University).
9:40 - 10:05 DIALOG: Diagnostic Logic System in Internal
               Medicine
               (H. Pople, University of Pittsburgh).
10:05 - 10:30 Model-based Systems for Consultation:
               CASNET (Causal-Association Network Systems)
               and other approaches (C. Kulikowski, Rutgers
               University).
10:30 - 10:50 Break
10:50 - 11:15 Analyzing and Simulating the Present Illness
               (S. Pauker, Tufts-New England Med Center & MIT).
11:15 - 12:10 Panel Discussion:
               Medical Perspectives of AIM Systems.
               Moderator: A. Safir, Mt. Sinai School of Medicine
               Panelists: R. Engle, Cornell Medical School &
               N. Y. Hospital,
               D. Lindberg, University of Missouri,
               J. Meyers, University of Pittsburgh,
               S. Pauker, Tufts-New England Medical Center,
               W. Yamamoto, George Washington University.

Afternoon Session:

II. KNOWLEDGE-BASED SYSTEMS IN PSYCHOLOGY AND PSYCHIATRY

1:15 - 2:15   PARRY: Improving a Simulation of Paranoid Thought
                   Processes (K. Colby, UCLA).
2:15 - 2:40   BELIEVER: Belief Systems Interpretation
                   (C. Schmidt, Rutgers University).

III. KNOWLEDGE-BASED SYSTEMS IN PSYCHOLOGY AND PSYCHIATRY

2:40 - 3:05   CONGEN: Constrained Generation of Chemical Structures
                   (B. Buchanan, Stanford University).

3:05 - 3:30   SECS: Organic Synthesis System
                   (T. Wipke, Princeton University).

3:30 - 3:55   Protein Crystallography System
                   (R. Engelmore, Stanford University).

3:55 - 4:15   Break

        IV. OVERVIEW OF SYSTEMS AND METHODOLOGY

4:15 - 5:00   Panel Discussion on Artificial Intelligence
                   Methodology in Medicine, Psychology, and
                   Biochemistry. Comparative review of systems
                   and future problems and perspectives.
                   (E. Feigenbaum Stanford University - Moderator)
                   Panelists: S. Amarel, Rutgers University
                   J. Feldman, University of Rochester
                   B. McCormick, University of Illinois at Chicago Circle
                   R. Schank, Yale University

5:00 - 5:40   Panel Discussion on Shared Resources and Computer
                   Networking

Schedule of Technical Sessions of First Annual AIM Workshop

Sunday, June 15, 1975

Morning Session:

8:30 - 9:45	A. Seminar on the DIALOG System (Pople and Meyers).
	B. Seminar on the BELIEVER System (Schmidt).
9:45 - 10:15	Break.
10:15 - 11:30	A. Seminar on Analysis and Simulation of the Illness (Pauker).
	B. Seminar on the CONGEN System (Smith/Carhart).

Afternoon Session:

1:00 - 2:15	A. Seminar on CASNET and related systems (Kulikowski and Safir).
	B. Seminar on Protein Crystallography (Engelmore).
2:15 - 2:45	Break.
2:45 - 3:40	A. Seminar on the MYCIN System (Shortliffe).
	B. Seminar on the SECS System for Organic Synthesis (Wipke).
4:00 - 5:30	Panel Discussion: Analysis and Comparison of Medical Systems.

Dinner:

6:30	Keynote Speech. (Dr. Edward Bloustein, President, Rutgers University) Guest Speaker. (Dr. William Raub, Associate Director, Extramural and Collaborative Programs, National Eye Institute, NIH)

Evening:

8:30 - 10:00	Special Interest Group Meetings and Hands-on Experience with the Systems.

Monday, June 16, 1975

Morning Session:

8:30 - 9:45   A.   Seminar on the PARRY System (Colby).
               B.   Seminar on METADENDRAL (Buchanan).
9:45 - 10:15   Break.
10:15 - 11:30 Special Interest Group Meetings;
               Hands-on Systems Experience.

Afternoon Session:

1:00 - 3:15   Seminar on Artificial Intelligence Systems
               (FUZZY, PEDAGLOT, MDS and other Knowledge-Based
               Systems). (B. Bruce - Moderator)
3:15 - 3:45   Break.
3:45 - 5:15   Panel Discussions on Analysis and Comparison
               of Systems:
                         A.   Biochemistry (Smith - Moderator).
                         B.   Psychology (Colby - Moderator).

Evening:

7:30 - 9:00   Seminar on Medical Systems (MISL Project)
               (McCormick UICC) Digitalis Therapy Advisory
               Program (Silverman MIT)
9:00 - 10:00   Special Interest Group Meetings and Hands-on
               Systems Experience.

Tuesday, June 17, 1975

Morning Session:

8:30 - 9:45   Panel Discussion:
               Methods of Inference: Formal and Clinical
               Problems (T. Shortliffe - Moderator)
9:45 - 10:15   Break.
10:15 - 11:30 Panel Discussion:
               Knowledge Acquisition and Representation
               (B. Buchanan - Moderator)

Afternoon Session:

1:15 - 3:15   Panel Discussion:
               Problems of Systems Development; Issues of
               Collaboration across Disciplines - Shared
               Resources and Computer Networking,
               Methodological Conclusions. (S. Amarel -
               Moderator)
3:30   Break.

Departure:

4:00

III. List of Panel Participants and their Affiliations

AMAREL, Saul   Principal Investigator,
                                Rutgers Research Resource
AXLINE, Stanton   MYCIN
                                Stanford Medical Center
BAKER, William   Biotechnology Resources
                                NIH
BUCHANAN, Bruce   HEURISTIC DENDRAL, Meta-DENDRAL
                                Stanford Computer Science Department
CARHART, Ray   HEURISTIC DENDRAL
                                Stanford Computer Science Department
DAVIS, Randy   MYCIN
                                Stanford Computer Science Department
ENGLE, Ralph   New York Hospital
                                Cornell University Medical School
FEIGENBAUM, Edward   HEURISTIC DENDRAL
                                Stanford Computer Science Department
KULIKOWSKI, Casimir   CASNET
                                Rutgers Computer Science Department
LINDBERG, Don   Chairman,
                                SUMEX/AIM Advisory Committee
MCCORMICK, Bruce   MISL Project
                                University of Illinois at Chicago Circle
MILLER, Randy   DIALOG
                                University of Pittsburgh
MEYERS, Jack   DIALOG
                                University of Pittsburgh
PARKINSON, Roger   PARRY
                                Stanford AI Lab
PAUKER, Stephen   PRESENT ILLNESS
                                New England Medical Center
POPLE, Harry   DIALOG
                                University of Pittsburgh
RINDFLEISCH, Thomas   SUMEX System
                                Stanford Medical Center
SAFIR, Aran   CASNET
                                Department of Ophthalmology
                                Mount Sinai School of Medicine
SAFRAN, Charles   PRESENT ILLNESS
                                Project MAC, MIT
SCHMIDT, Charles   BELIEVER
                                Rutgers Psychology Department
SCHWARTZ, William   PRESENT ILLNESS
                                Tufts-New England Medical Center
SHORTLIFFE, Ted   MYCIN
                                Stanford Medical Center
SILVERMAN, Howard   Digitalis Therapy Advisory Program
                                Project MAC, MIT
SMITH, Dennis   HEURISTIC DENDRAL
                                Stanford Chemistry Department
SRIDHARAN, N.S.   BELIEVER
                                Rutgers Computer Science Department

SRINIVASAN, C.V.   MDS
                          Rutgers Computer Science Department
SZOLOVITZ, Peter   PRESENT ILLNESS
                          Project MAC, MIT
YAMAMOTO, William   Department of Clinical Engineering
                          George Washington University

IV. Brief Description of Systems Presented at the Workshop

BELIEVER [Rutgers]: This system models how a person in the role of an observer, perceives and explains observed or reported actions to others. The goal of the system acting as observer is to answer the question: "Why did person P perform act A at time T?". The question is to be answered by attributing to person P a plan and motives which caused that person to decide to perform action A. Thus the problem is to move from observations to inferences about the internal states (Believes, Expects, Wants etc.) of person P. This type of causal explanation of observation is similar to reasoning in other knowledge based problems such as medical or psychiatric diagnosis. The AI framework adopted for this work called MDS is being developed at Rutgers and provides a formalism for describing the theory.

CASNET [Rutgers]: This system embodies a causal representation of the processes of dysfunction incorporating four main structural elements: the patient findings (signs, symptoms and test results); the patho-physiological states that summarize and explain the findings; the disease hypotheses expressed by their component states; the therapeutic actions which attempt to counteract various aspects of the disease. Such a model has been applied to several dysfunctions, but principally to the glaucomas. Reasoning schemes have been developed for the interpretation of findings, diagnostic decision making, prognosis, therapy selection, and explanation of reasoning in terms of the model and supportive research references.

DIALOG [University of Pittsburgh]: A computer based system for general medical consultation that incorporates a hypothesis formation system using a medical knowledge base now encompassing a substantial portion of the major diseases of internal medicine. The system thereby exhibits diagnostic behavior and competence comparable to that of the skilled clinician, and handles systematically, cases where two or more distinct clinico-pathological entities are present.

HEURISTIC DENDRAL [Stanford]: The objectives of the Heuristic DENDRAL research program are the development of innovative computer and biomedical analysis techniques for application in medical research and related aspects of investigative patient care. The global aim is to apply the unique analytical capabilities of gas chromatography/mass spectrometry (GC/MS) with the assistance of data interpreting computer programs utilizing artificial intelligence techniques, to investigate the chemical constituents of human body fluids in a variety of clinical contexts. A set of artificial intelligence programs interpret data and generate plausible molecular structures. The most important program is the constrained structure generator CONGEN, which generates molecular structures within structural limits. These limits (for example, ring size) are either specified by a chemist directly or inferred from mass spectrometry data by another program called the DENDRAL PLANNER. The problems of organizing and developing this complex system are common to many knowledge based problem solving programs.

META-DENDRAL [Stanford]: Meta-DENDRAL is an induction program for finding rules that characterize the processes that are of interest to

the chemist (for example, rules of fragmentation in mass spectrometry). The name Meta-DENDRAL suggests an effort beyond, but not entirely separate from that of Heuristic DENDRAL and is a response to the immense task of extracting inferential knowledge from experts and making that knowledge accessible to the Heuristic DENDRAL engine. The number of rules is potentially very large and experts have yet to investigate most of them. Therefore, automating the rule formation process seems essential.

As in Heuristic DENDRAL, the heart of the program is a generator of legal solutions, in this case a rule generator called RULEGEN. The generator needs prospective constraints in order to generate plausible rules rather than all possible rules. The planning program for doing this is called INTSUM. The test phase of Meta-DENDRAL under the PLAN-GENERATE-TEST paradigm is a program called RULEMOD which evaluates and modifies rules in the context of other rules.

MISL [University of Illinois]: The Medical Information Systems Laboratory (MISL) is set up to explore the use of artificial intelligence techniques in clinical decision making and pursues three major activities: clinical research and decision support; construction and modeling of a data base in ophthalmology; and network-compatible data base design. The project explores the inferential relationships between analytic data and the natural history of selected eye diseases both in treated and untreated forms. SUMEX/AIM is utilized to build a data base to be used as a test bed for the development of clinical decision support algorithms.

MYCIN [Stanford]: A computer program that uses expert clinical knowledge to advise physicians on the diagnosis of bacterial infections and the selection of appropriate therapy. The distinguishing characteristics of this system are: it acquires information through human engineered interaction; it permits extension of its rule structured knowledge base; it explains its reasoning process in response to simple questions posed in English.

PARRY [UCLA]: An interactive program that simulates the behavior of a paranoid patient during a diagnostic interview in a hospital setting. The conversation is carried out in English. The model consists of a delusional network which operates by detecting flare concepts in the doctor's statements, and thereby modifying its own affect states such as Fear, Anger, Shame, Mistrust in response. The affect states guide the nature of the resonses given by the program. The degree of paranoia can be set at the start of the interview. The model has undergone elaborate validation and sensitivity tests.

THE PRESENT ILLNESS PROGRAM [MIT]: A system which analyzes the history of the present illness for a patient starting with a certain complaint. The knowledge base was developed by analysis of the behavior and declared reasoning of clinicians and by introspection. The knowledge base is organized into Frames as defined by Marvin Minsky, that are linked into an associative memory. The memory is partitioned into long term and short term types which permits likely hypotheses to be arrived at rapidly and considers frames that are closely linked to the hypotheses.

PROTEIN CRYSTALLOGRAPHY [Stanford]: This system has as its goal the application of AI techniques to the Phase Problem of X-ray crystallography in order to determine the three dimensional phase structure of proteins. The system obtains from experts the knowledge and heuristics needed to infer the structure of proteins and to represent them as a cooperative set of processes that can successfully arrive at plausible structure descriptions in a reasonable amount of time. The goals of this project are clearly long term but are organized in such a way that significant intermediate goals can be realized before the project is completed.

SECS [Princeton]: This is an interactive program for computer assisted planning of organic chemical syntheses. It is human engineered and makes extensive use of graphics whenever possible to display chemical structures, synthesis sequences and the solution search graph. SECS has extensive knowledge of chemical transforms and chemical principles and is designed to let the chemist expert do the major portion of the search guidance interactivity. SECS uses an English-like chemical language for describing transforms that the chemist uses to extend the knowledge base. Current work is centered on developing advanced strategies that exploit three dimensional models and an electron structure model that SECS currently knows how to build.

****

V. PANELS

A. MEDICAL PERSPECTIVES OF AIM SYSTEMS

B. ANALYSIS AND COMPARISON OF MEDICAL SYSTEMS

C. KNOWLEDGE ACQUISITION AND REPRESENTATION

D. METHODS OF INFERENCE – FORMAL AND CLINICAL PROBLEMS

E. PROBLEMS OF SYSTEMS DEVELOPMENT

A. MEDICAL PERSPECTIVES OF AIM SYSTEMS

MODERATOR - ARAN SAFIR

ENGLE: In his stimulating book PHILOSOPHY OF AS IF, H. Vaihinger (Routledge and Kegan 1935) presents a thesis which relates directly to the application of so called Artificial Intelligence to the field of medicine. He postulates that we often accept as true the fiction of approximations because of some useful benefits which result. In a sense all of science and mathematics is an approximation of the real world, and there are benefits to be gained if we act as if science were the real world. Similarly, benefits can result from acting as if artificial intelligence were the same as human intelligence though the term Artificial Intelligence seems a bit presumptuous to some individuals. The full benefit of the use of computers as tools of thought can come only when we learn to dissect intelligence into a portion best suited to the human being and a portion best suited to the computer, and then find a way to mesh the two processes. The science of Artificial Intelligence is concerned with that very important task.

YAMAMOTO: Artificial intelligence as it appears to me is attempting to emulate or imitate the performance of the academic physician working generally with the most severe disease patterns. And when you mention artificial intelligence to a number of physicians you arouse a basic hostility because you are threatening them in the area they have reserved for themselves. They are willing to give the IV's to the nurse and the drugs to the pharmacologist and the surgical preparations to the OR nurse. But what they reserve for themselves is what they consider the intelligence.

One can attend conferences devoted to defining the phrase Artificial Intelligence. I have found that you can reach an innocent ground by calling it Artificial Behavior because in identifying what intelligence is we use certain phrases which generally define subtypes of behavior. I would like to list those types of behavior one might refer to in determining whether or not someone or something is behaving intelligently, and more specifically those types of behavior that I think physicians would include if they attempted to assess AI.

The first intelligent component is the choice between alternatives where the alternatives are not necessarily mutually conflicting. I think the AI community has done a fair job of answering that.

Second is execution of pre-determined processes. That is, physicians learn as do others, things which are pretty well defined algorithmically or procedurally which are stored away and then invoked at a select time. The ability to do this very often appears to be intelligent and I think the AI comminity has made substantial inroads here.

Third is learning facts or knowledge by inductive inference and learning by rote. Learning by rote is what we do in medical school, learning by inductive inference is what we hope the doctor will do when

he gets out. There is a questionable level of success here as far as AI is concerned.

Fourth is initiative and invention. These two words we associate with intelligence although there are other components mainly emotional that determine the manifestation of it. I think there has been no contribution from AI in this area.

Fifth is operating under conflicting policy where policy covers a broad range like "don't do harm". As far as I know there has been very little activity along this line in AI although it seems to be an attackable problem.

Sixth, self awareness has to be a component of intelligence and this of course is a basic philosophical, perhaps epistemological problem which AI probably has not attempted to answer.

Seventh is to assign value judgements or assign value to judgements that the performer executes, or values in the context of a society, that is, in the context not just of a patient but also that patient's family. This type of extraneous but nevertheless relevant intelligent activity expands the scope of your problem. This is another area in which AI has not made any substantial inroads.

Eighth is solving problems. This can include playing games to more complicated diagnostic games. I think here AI has contributed a number of very interesting and powerful paradigms.

Ninth is recognition of logical consistency which is something that AI people try to pull into their systems. We cannot say at the present time that AI has a method by which logical consistency of new systems can be determined, but this is a problem which is not unique to AI.

Tenth is operating under tentative decision. Most of the front line physicians operate under tentative decision circumstances. I think MYCIN is an example of an attempt to go in that direction and necessary to emulate if you are going to imitate the intelligent behavior of the clinician.

Eleventh is operating toward an indeterminate or the "qualitative end point". That is, intelligence often allows you to say you don't know what the end point is but you will know when you get there. The ability to operate under that scheme is a manifestation of intelligence.

I am sure all of you can think of other forms of behavior which contribute to the definition of intelligence and until we get a substantial number of these under control we probably will not be able to convince the street physician that AI has a great deal to contribute. Let me say that as far as disseminating AI in the medical community is concerned, I'm greatly heartened by the interest of major medical physicians in the country like Dr. Meyers and Dr. Schwartz because the only way there will be a more congenial reception of AI in medicine in the profession is through clinical leaders becoming interested, and training their students to be aware that thought processes have structure and that structure can be experimented with by using machines.

PAUKER: In the past the clinical importance of computer science in medicine involved both data handling and the dissemination of medical knowledge. Now an additional capability has developed, the ability of the computer to serve as a laboratory to model decision making and to test theories. Our group has explored as have others, the impact of decision analysis on the decision processes in medicine both in diagnosis and in treatments. It has made me far more aware of the necessity for being explicit in my decision making processes after seeking firm and relevant data upon which to base any deduction. Probability theory and especially Bayes rule now form a central part of my diagnostic approach in terms of computer programs. However, our recent studies have emphasized the importance of a richly cross-linked data base of guessing and heuristic approaches. These ideas fit more closely the romantic notion of what clinical expertise is and to some extent have underlined the need for complex learning and indexing processes. With this new kind of laboratory and approach we are beginning to understand better how to teach students what clinical expertise really is. Having more patterns with which to match and explore the expert can plunge in and guess and if he makes a mistake he has rules by which he can back up. And having seen that this is also the procedure of some programs, as a clinician I am pleased to know that there is nothing wrong with exploring in this manner. It works and because it works perhaps AI has something to learn from medicine in the same way medicine has something to learn from AI.

LINDBERG: First I want to say why I consider the SUMEX/AIM project to be of great significance. The first reason is that reliable high performance computing which is required for reasonable AI development is now available at a reasonable cost and hence the experiments may succeed or fail on their own merit without the added complexity of inadequate computer resources. There are still some inadequacies in the system, especially in the area of large files. But these aside, it now seems quite possible to test if AI in fact has anything to offer medicine which I think is the fundamental raison d'etre of the experiment. The SUMEX/AIM is significant because it's mode of providing computer services to medicine is an attractive alternative to the traditional single, large institutional computer center. Personally, I would like to see it succeed. The SUMEX experiment provides that the cost of maintaining an advanced system be borne by a single group, with other institutions using the facilities. In addition one might say that the approach allows networking to reduce the programming/hardware compatibility problem.

For what purpose then should one attempt to employ AI techniques in medicine? For me there is absolutely no doubt on this point. I think AI should be used to do in medicine what cannot be done without a computer. Now that would mean that the universe of choices be divided not between forbidden patient care applications and permissible basic research applications but rather between those things which cannot be done and those things hopefully which can. And I have three examples that I would like to mention.

First, we do not have presently a uniform terminology for medicine let alone a vocabulary, nor do we have a means to create either. It goes without saying there is no meaningful national accumulative data

base effort. Therefore there really is no systematic way for clinical records to become the basis for research. It is likely that AI could create a means to build a vocabulary and I point that out as a problem of major importance.

Secondly, we do not have a general means to test potential causal or non-causal medical associations, a consequence of that being the thalidomide/pregnancy association for example. If there are such assocations to be made today we are no better prepared to recognize them or be alerted to them by a computer than we were ten years ago. When we speak of early warning systems for drug side-effect or drug interaction, we are hypothesizing the particular effects for that special problem. The more general problem would be to prescribe the way in which such an association is actually recognized. If we could do this we probably would not have to plead so hard every year for data collection systems.

Lastly, I would like to suggest that we cannot as yet manage very large files or large and complex data bases. You may say that this is being done already but I am suggesting that we really only think we are doing it. Let me give you the file problems I have in mind.

First, geographical data systems. There are practically no usable systems which allow medical data observations to retain their geographical structures along with their other attributes. The Lighthill Report for the National Research Council in England singled out this application area, geographical systems, as the most promising AI application and I think it is not being followed up. To illustrate the value of such application I need only to remind you of the well known but little understood geographical distirbution of multiple sclerosis in the USA. It is sixteenfold more common in New Orleans than it is in Seattle. Or the varying attack rates of coronary artery insufficiency which is threefold higher in Georgia than in Lincoln Nebraska. We do not have any means to recognize these associations. Those particular ones have been made and validated but how many others are there?

The second data base problem I want to mention has to do with the medical record data file. We are doing the computations but not really managing the information in the files. I think a reasonable solution to that would be to design a system in which the file knew more about what it contained than the inquirer. And that is a problem which I believe is suited to AI methodology.

I want to make a statement about Dr. Meyers system because what I've said may seem in conflict with the fact that I very much admire what Meyers and Pople have done. I think it is very sophisticated work aimed at a very important problem. But I do not feel it is important because they are automating the good consultant. We cannot make another Jack Meyers but american medicine does turn out very good internal medical consultants nonetheless who may grow to be as good as he. For me the importance lies in the fact that they are accomplishing in this program something which cannot be done without the computer by providing a facility whereby diagnostic rules are made accessible and can be applied to a particular case without the presence of the consulting physician.

MEYERS: In spite of Dr. Lindberg's point of view I still believe that the kinds of programs we are developing using the techniques of AI, will continue to have diagnostic application even in the tertiary care institution. Now the number of applications is obviously going to be limited, I thoroughly agree with that. It is probably not so important to develop these AI techniques for routine tasks. No physician by and large needs a program to help diagnose common symptoms. But the paramedical personnel who are taking on care responsibilities may well need this kind of support.

My last comment has to do with the educational application of AI techniques. I have mentioned already the use of our data base for educational purposes, but I hope you can see that these kinds of techniques can be used for standard self-education as well. For example, in our program if a medical student just wants to add "stortness of breath" and stop there, the computer can provide quite a thorough and differential diagnosis of shortness of breath. In addition these systems could be utilized for measuring clinical competence not only in students but also in graduate physicians. And this is becoming an increasingly important aspect of medical practice.

SAFIR: I believe that developing computer methods for intelligent problem solving in medicine can be accomplished only by close collaboration between the computer scientist and physician. And a true understanding of the nature of the data and the problem can be achieved only if the computer scientist is exposed to the very long and difficult process of education in medical problems. He has to serve a clinical clerkship as we call it in medical school because what one gets out of text books and the literature is really just enough to get started. One has to develop a feeling for the complexity and unreliability of the data. Dr. Kulikowski and colleagues have been very involved in observing glaucoma surgery and seeing patients undergoing the measuring process. As a result their understanding and interpretation of the literature has changed tremendously. Likewise, the physician who gets involved with the computer scientist cannot just preach medicine. He must learn how the computer scientist imbeds these clinical lessons in some logical structure and manipulates them. These may sound like relatively easy goals but they require the selection of personalities that are not at all typical of the professions involved. Computer scientists are selected mainly from among those who have a talent for mathematical disciplines and who are encouraged to develop orderly systems of thought that function with predictability and precision. Physicians on the other hand have entered by choice a profession in which disorder and unpredictability are nearly the rule. If someone comes to a mathematical scientist with a problem for which there is yet no solution there is rarely any pressure placed on him to supply one immediately. Clinical physicians obey a very different mandate. They must solve the problem at the time it is brought to them no matter how imperfectly and they are compelled regularly to make crucial decisions in situations that are characterized by inadequate theory and grossly imperfect data. I've often thought that the entire system of medical education is a means of teaching an intelligent and sensitive person to live happily with the intolerable. So computer scientists who can thrive within the disorder of medicine and physicians who can work happily within the logical and mathematical world of computer science

are, to use doctors' terminology not rare but destinctly uncommon. And I believe that good work in computing and medicine will result only from such collaborative teams.

SCHWARTZ: The process of developing large systems that are reliable enough to make an impact on clinical research will require inevitably a large investment of resources over the next few decades. And I wonder if society and the funding agencies are willing to wait that long. Quality care is one of the key issues around the country today. And I feel we ought to be able to convince those who are making the financial decisions that this work really has nothing to do with computer programs but has to do with the development of insights into high quality clinical care and clinical judgement which will allow an enormous up-grading of medical education and medical curriculum.

Most physicians including fourth year medical students are already so professionalized and acculturated in the traditional way of learning medicine that their minds are not open to analyze the structure of their decision making and cognitive processes. I am convinced that we should be teaching problem solving and the nature of the cognitive process in second year of medical school before students are so professionalized. We now know enough to be able to do that. As a community we comprise an important resource which can be a force for encouraging the development of medical curriculum that will emphasize processing of information more than simply acquisition of information. And I believe that is a societal good which a great many people will be able to appreciate and accept on its own merit.

********END OF PANEL DISCUSSION********

B. ANALYSIS AND COMPARISON OF MEDICAL SYSTEMS

MODERATOR - HARRY POPLE

POPLE: I have been asked to summarize a paper I submitted to the IEEE in which I compared three of the four systems represented here, MYCIN, CASNET and DIALOG.

All three systems deal with the problem of hypothesis formation but the hypothesis formation imbedded in MYCIN as I see it, is a special case of deductive reasoning. The organization of rules takes the form of a tree structure and the analysis to derive hypotheses is deductive inference. One begins with the goal which in this case is to prove the occurrence of a disease, and each candidate disease is considered in turn in an attempt to prove the occurrence of that disease by working back to antecedent structures until it is possible to establish a confidence level. The other systems use the alternative reasoning tactic of inductive inference, or reasoning from consequence back to hypothesis and from hypothesis to consequence. DIALOG for example, has pointers running in two directions from manifestations to disease entities and disease entities to manifestations. Going   from   a manifestation to hypothesis is what I call the hypothesis formation step or the abductive step.   Working from hypotheses back to resulting predictions is the deductive step which corresponds exactly to what goes on in MYCIN as I see it. So we are employing two distinctly different forms of logic to achieve the same kinds of activity.

SHORTLIFFE: I interpret the underlying logic of MYCIN differently. MYCIN was conceived originally as a consequence theorem system. We work backwards from a goal and we invoke pieces of knowledge on the basis of what hypotheses we are trying to reach. The introduction of certainty factors into the scheme makes it difficult for me to interpret that as deduction because we are dealing with antecedent rules. We recently introduced antecedent theorems into the system. As soon as we know the identity of an organism we immediately determine the gram stain morphology. And that is a forward looking mechanism that we did not have before.   In the past when we needed a gram stain we had to find rules that would allow us to deduce them in a very round about way.

I agree that there are differences among the systems but your description of those differences, namely deduction vs. induction or abduction is not an accurate interpretation in my opinion.

We have felt from the outset that the perfect system would be one in which the clinician who needs advice could sit down at the terminal and set the scene with information that we in turn would use to ask the appropriate questions.   That is the way patients are presented and discussed in the clinical setting.   That of course would require adequate natural language understanding in the system. So we look for ways of avoiding natural language within the context of the consultation itself.   We needed some natural language processing in order to answer questions and to do some of the explaining, but we at least wanted to let people get out advice without having to deal with the frustrations

of what still is an unfinished piece of AI research, natural language comprehension. The work we did on natural language understanding cannot be defended in any theoretical sense. It was a stop gap measure to get something that would work well enough for our purposes. We recognize the need for it and I believe it is the way these systems should go.

AXLINE: I believe clinicians are more comfortable if they can use the standard format for entering data about a case. But we were interested in simulating the logic process the clinician uses, not necessarily the standard format he uses to gather clinical data which I consider stilted. Our approach is to collect only that information which is going to be used at that time, rather than to accumulate large amounts of data. So in terms of understanding the logic process our approach has been particularly productive. The approach that Ted is describing of setting the scene is of equal merit.

SZOLOVITZ: The clinician combines a highly specialized vocabulary with a set format to enter clinical information. So that in this instance designing a natural language system would be much easier than it is when you must contend with totally context free input. Anything that gives you a structure provides a handle on the problem. And there is a reasonable amount known about parsing so that this is not entirely an impossible problem.

I want to comment on the way we model how our expert clinician deals with data that is presented to him. We are very strongly influenced by Bill Schwartz's absolute refusal to listen to a case in any but the standard order of presentation. And there is a methodological point here. If the program is not able to make use of information as it comes in, then what does it mean to say that you are accurately simulating the deductive or logical process of the clinician?

AXLINE: There are several ways of looking at the information gathering procedure of the clinician. The general internist for example looks at the whole patient and all the problems he presents. This is different from the procedure followed by the consultant who is the person we are talking about here. The consultant plays a much different role in that it is not his function to reproduce all the information related to a case, which in part means that he can collect information for processing in whatever way he wants.

SHORTLIFFE: I'd like to describe the way in which MYCIN's rules have been acquired to make it clear that we are not necessarily trying to make the program perform the way a clinician does. What we are tryng to do is understand well enough the way the consultant analyzes the problem so that we can come up with representation that works. All the rules we have in our system have been acquired at weekly meetings in which Dr. Axline and Dr. Cohen, the two clinicians most closely associated with our project took patient charts and with the end of those charts still unknown to them, began to review them. Those of us unfamiliar with the clinical aspects of what was being discussed would listen and try to pick out the underlying threads of reasoning. We would then code these into rules and use them to run patients' charts. We would then bring back the results to show the expert how the system actually used the rules in order to come up with recommendations. So

our concerns were whether or not the rule we used represented a fact that the expert could agree with, whether or not he had ever used it before in that way, and whether or not the results of the program in terms of recommendations agreed with what he would have recommended for that patient. We want the program to derive the right advice and whatever way we can come up with to do that is all right.

So we are looking at something really very different from what Dr. Pauker and Dr. Schwartz have been doing in trying to understand the actual reasoning process that takes place.

KULIKOWSKI: Our system is a vivid example of how, if you want to give advice in a given area, often imitating the doctor is not necessarily the way to go . What you want to have is a number of alternative models, with the simulation of a particular doctor being just one of them. It clearly depends on the scope of your problem and on the knowledge structure of a particular domain.

SZOLOVITZ: All of us are trying to provide people with expert clinical advice and the methods for doing that can range from simulating the clinician's logic to using a mathematical model. Howie Silverman for example, started out with what looked like a very large AI project, namely to derive a method for prescribing digitalis therapy. It turns out that the major part of the program is a very nice algorithm that does quite well and it uses the AI technology when interacting with the real world. So if we could do that for all internal medicine perhaps that would be the ultimate solution.

SRIDHARAN: I see a tremendous richness of concepts going into the building of these systems, especially those of the MIT group. And I wonder how you go about deciding whether or not you need to do all this processing? Howie Silverman's project is a clear case. If he had wanted to make it look like a flashy AI program he could have done it. But actually the idea would be not to do it. If you can reduce the processing structure and encode your information in a clean form that will do the job, that is the desirable way to go.

SZOLOVITZ: An example of a very rich and complex theory is Andee Rubin's master's thesis which is available at the AI lab at MIT. It deals with medical diagnosis. She observed one of the doctors in our team diagnosing Steve Pauker who pretended to be a patient. The exercise was to go through the resulting transcript and establish the kinds of processes and knowledge involved. And that protocol became the basis for the system. Now unless you have very good models for the underlying medicine it is very difficult to do much better in terms of dissolving the AI part and being left with the concrete model.

AMAREL: It seems to me in most instances the doctor is the decision maker who draws from certain bodies of knowledge that are for the most part systematic and ever expanding. And I see two components in the projects we are discussing. The first is the richness of the hypothesis space which varies between systems and in the way each system keeps track of possible hypotheses, evaluates them, partitions them and uses them. MYCIN for example has practically no hypothesis formation process. On the other hand, DIALOG is very concerned with the taxonomy

of a large number of diseases and syndromes and searching that space entails deliberate processing of hypotheses. And this is where I think AI comes in much more than in some of the other systems. So the size of hypothesis space and the kind of tools you bring into searching the hypothesis space are the determining factors.

The other component is the extent to which a project is interested in simulating the doctor's decision making process in the clinical setting. Some systems are geared toward doing precisely that, while others draw from specific bodies of knowledge in a particular domain and a variety of strategies for using that knowledge.

POPLE: Our system is an example of a simulation. We did it not because I had any specific interest in trying to simulate Jack Meyers but because I had no other way to get at the problem, and he seemed to be a good model for going about it. The heuristic I hit upon was the only one I could find that resembled the behavior I saw. So I think you are right in saying there are different motives represented in these systems and therefore differences in terms of the way one should look at results and evaluate them.

FEIGENBAUM: The problems being discussed here in the context of AI in medicine are almost identical to those issues and problems that arise in other areas of complex interpretation. This is a group of people who share the same sets of concepts, who read each others papers as ARPANET messages the day after they've been generated and so naturally we all share the same sets of concepts. I think everything that people have been talking about has had to do with expectation driven or model based systems for analysis, that these are model based hypothesis formation systems specifically, that the models come in a variety of types, associational, causal and sometimes even statistical, that the knowledge is inconsistent, typically in great quantities, that the knowledge is represented in a rich repertoire of representations we all know and massage each day. We may not use them all the time, but they represent the common tools and techniques for dealing with this knowledge in a highly flexible way. So everyone has come to realize that inserting the knowledge, deleting it, modifying it are the critical problems and we've all invented roughly similar ways of doing it. And coupling all this with these rich inferential processes, we essentially have a kit of techniques that we all can appreciate and explore.

I admire Harry Pople's courage in writing an article comparing these systems. I would say that the easier article to write would be one comparing what we might have heard, say six years ago at a conference on medical diagnosis with what we are hearing today. There is an incredible difference. For instance, compare the current work with that of Ledley/Lusted of more than a decade ago, with Signs and Symptoms Matrix and application of Bayes theorem comprising the rich inferential rule of that system. Or compare the current work with the techniques on which millions of dollars have been spent in statistical pattern classification or clustering techniques for diagnosis. Or compare the current work with what was supposed to be the solution to all this, the so called logic tree which is very static. So the techniques that are being discussed at this conference are light years away from what was being discussed only a few years ago. There is an

enormous gap between what we knew then and what we know now.

SAFIR: I am concerned that computer scientists think they are modeling or simulating a process that they view as static. But it may very well be that the process of medical decision making is undergoing changes almost as rapidly as computer science so that what AI is using as a model today could be the product of medical schools thirty years ago.

SHORTLIFFE: In Dr. Engle's description of the past twenty years of medicine it struck me that a tremendous amount of work and man hours have been poured into the problem of medical decision making during this period. And now it can be automated and analyzed. And I wonder if Dr. Engle gives a talk ten years from now about AI in 1975 whether he will be able to say that AI had the key that had been overlooked for those past twenty years. And I think the challenge we should recognize in all this and take up at this point is to keep from becoming obsolete in the near future.

LINDBERG: I haven't heard anyone attempt to measure the magnitude or quality of our accomplistments. Ted asked what will be said of the work in ten years but I think in much less time we will look back and realize that some of these diagnoses showed great achievement and the programs really did well, some were very simple and the whole thing was over-instrumented, and in some cases the decisions were wrong. And I think we have to make a serious effort to separate out which of our accomplishments are major and which are minor. They cannot all be of the same quality.

SRINIVASAN: There has been a lot of discussion about the usefulness of various techniques for producing advice in medicine, but I wonder what is going to be next. Is it going to be more of the same, more specialized model building? I tend to think that direction is static. Is this for the doctors the general paradigm or is there also some concern for planning functions?

MEYERS: I would say good doctors in most circumstances must have a definite therapeutic plan which may be modified with experience of course. We well recognize in DIALOG that treatment plans are extremely important in the overall scope but to deal with therapy is as big a problem if not bigger than the problem of diagnosis. This is taking two worlds at once which is just too much. Fortunately, smaller programs like MYCIN or CASNET can deal with this but we had to put it in second place. And I believe Dr. Pauker is also in the same situation for the most part.

PAUKER: We are to some extent but I think I have to disagree with your statement. I don't think that the world of diagnosis and the world of therapy are all that separate. I think it is a world of patients and therefore we never really know until the autopsy that we arrived at the right diagnosis. We are always undertaking a therapeutic plan trying to make the patient better, not being certain of what the diagnosis is. And clearly knowing what to model in terms of therapy initiation is very dependent upon and strongly influenced by what we mean by arriving at diagnosis. Often it is very difficult to know when we have reached that

stopping point. What that arbitrary stopping point is depends on what we're going to do next, the seriousness of the situation, the amount of time we have to provide treatment. So we cannot finesse one or the other.

MEYERS: I agree with you. Perhaps I can make my point clearer. Once you get therapy into a system you then perturb the whole system and the data base becomes radically changed by the very presence of the treatment. And that is a very complex change which I am sure causes as big a problem as the original data base.

PAUKER: I would have to agree with that. If you are studying a case in which you find treatment was already prescribed, it changes the whole issue of consistency. You needed a certain finding which is no longer there because a doctor took it away. But the problem of therapy does not go away just because we ignore it. In dealing with two diseases, one can mask the manifestations of the other, in which case we are back in the same ball park. As opposed to being cured we might say that the therapeutic intervention of a physician at some level represents another disease.

SZOLOVITZ: There is also the problem of history. When we study a case history what does it mean to say that a person has had a certain disease for three months? He did not all of a sudden have it. He had a lot of different symptoms which in retrospect amount to this particular disease. Now if he still has this disease in addition to some new disease, we are in exactly the same situation as we are when we initiate treatment. Because in order to understand the historical information we have to cope with this question of how diseases behave in time and with other diseases, what our expectations were as opposed to what actually happened, and how we form hypotheses to account for them.

MCCORMICK: One of the great thrusts of decision making was cost effectiveness, developed by people in the Department of Defense which as far as I can see practically has strangled the community for the past ten years. The problems we are focusing on in the medical area are not that different from what is required for good decision making in other areas including planning in government or business. From among the various techniques we have developed to solve our problems could we find a more flexible mechanism to replace cost effectiveness as the standard criterion for judging the progress and development in a field? The closest any group has come to dealing with that in the context of management is Bill Martin at MIT who is building systems for management decision making.

FEIGENBAUM: I would like to discuss another potential application in that area. When you try to do hypothesis formation you often reach critical points in the analysis where you need some sophisticated piece of data that is extremely costly to obtain and you must decide whether or not it is worth the investment of time and effort to get that information. Right now we give over these decisions to human analysts. One of the things we know about these knowledge based systems is that they are extremely systematic in their application of a body of knowledge and often much better at it than the human experts who build up the rule base in the first place. Could we use these systems for

making those decisions as opposed to trusting the opinion of the physician who may not be as systematic? There have been other types of model builders which have considered this problem but those discussed here are much richer in terms of knowledge employed and I wonder if it ought to be pursued.

********END OF PANEL DISCUSSION********

C. KNOWLEDGE ACQUISITION AND REPRESENTATION

MODERATOR - BRUCE BUCHANAN

BUCHANAN: This panel will discuss the acquisition and representation of knowledge in computer programs. The critical issue is how to transfer knowledge into the program. And as that depends partly on what representation one chooses, both issues are closely related.

With DENDRAL we tried to custom craft the system. We worked with chemists many hours putting their knowledge into LISP code. In the long run it somehow begins to work, but the stability of the project is crucial in this method of collaboration because it is slow and tedious.

Another approach is to move knowledge from the heads of experts into a program by an interactive dialogue system. We tried it with DENDRAL and we are pursuing it more with MYCIN. My own bias is that both methods are inefficient. We are therefore pushing the META-DENDRAL effort which tries to take the knowledge directly and infer the rules that are needed for the program, thereby removing the expert from the picture.

DAVIS: With regard to acquisition, one thing we've found very useful in MYCIN is <u>acquisition in context</u>. That is, not only the knowledge but the reason for entering that knowledge is put into the program, for instance, entering a rule in response to a bug. With this approach you get a step up on the problem of assessing the impact of a particular rule on the knowledge base of the system. One of the constraints on the premise of a rule that has been given in the context of a bug is that it is going to have to evaluate to true in the context of the current consultation. Otherwise the rule simply is not going to fix the bug. This kind of knowledge is in the system I'm developing. It will accept any rule you give it but if it is in the process of trying to fix something and the rule will not be useful in that context, it will say so and request another. This is all clearly predicated on the assumption that working with an expert and putting his knowledge in the form you find suitable is the right way to go. In our system the form happens to be a rule, and we draw the knowledge out of the expert in this way.

This method presents the problem of how to deal with the ramifications of a new data structure on the system. Giving the system some understanding of its own representation seems to help. That is, give the system some capability of dealing with its own data types, and of being able to follow along some of these implications just by the structure of the types. We've done this and it helps. There are semantic implications that I don't know how to handle automatically. At the moment the user has to guide the system.

SCHMIDT: The methodology in our BELIEVER system involves developing a model of the thought process that an expert or anyone for that matter uses, to solve a problem. With that model we try to generate the response we think the expert will come up with in a given

context. We then compare our model's response to the expert's actual response in that situation. We have found that unless we decompose or categorize the information in the same way the expert or subject has, it is difficult to extend the system further.

SRIDHARAN: I would like to show that three of the issues being discussed have a common solution. The first is that of designing the knowledge based system and putting formal knowledge into a predesigned, simple and uniform knowledge structure. This could be greatly facilitated if a natural form of representation were used. There is no bug free system and there is no knowledge base that doesn't have problems. So it's not enough just to design a representation. It has to be designed with the idea in mind that you're going to be changing it constantly. It is not enough for the system to produce right answers. It has to be able to give reasons for those answers, in some sense explain its own processes. It has to be credible. Again, my contention is that a solution to these problems can be found in representing the knowledge in a natural form. Roger Schank's group is doing work in this area using the notion of computable semantics. Srinivasan's Meta-Description System which we are implementing partially in our BELIEVER system is also founded on this idea of a natural representation. The problem we are all experiencing in trying to explain our own systems and understanding others could be alleviated also if the knowledge were represented in a more readable, natural form. It would make it much easier to get down to the concrete stuff of the system and follow its reasoning processes.

PAUKER: These issues of knowledge acquisition and representation depend heavily on how much knowledge you are talking about. The domain of Internal Medicine is representative of the real world in its complexity and number of facts one has to know and work with. The process of acquiring all those facts in a data base is one problem, maintaining consistency in that data base is horrendous and finding the errors in that data base is impossible. I don't know how you are going to go about doing it. Finding it by instance in any reasonable period of time is not possible. What we do with doctors is to produce what we think may be a reasonable approximation, send him out and when he kills a patient we do a CPC, an autopsy, find out what went wrong and correct it. Just acquiring let us say, all 210,000 facts contained in the text on Internal Medicine by Harrison, is not expertise. The medical student who memorizes it all is not a doctor yet. He has to be able to apply that knowledge in the right circumstances, to organize that knowledge at run time, not just at the time of system formation. We can each chop out our own neatly constrained problem where each of our own approaches works. But applying these in real world situations is another problem.

DAVIS: There are two points here. One is, the text does not contain 210,000 unrelated facts. So we are dealing with an order of n not n-factorial when we talk about facts. I don't think we ought to be intimidated by raw numbers and facts. Clearly, there are levels of organization one can work with.

PAUKER: Our experience in developing the Digitalis program has been that we cannot compute all possible implications of a fact we put into the system in a reasonable period of time because of the number of

chains it produces. As your system grows, an added fact becomes harder to deal with.

COMMENT: Perhaps Samuels' checkers playing program offers a useful approach for handling a large medical data base. He found that the best way to debug the checkers data base was to have it run through masters' games and any time the program generated a move that wasn't the next move in the master's game, it adjusted its heuristics so it would generate that move. So a possible way to debug a medical data base might be to have the program run through CPC's and see if it is generating the same decision at each point and if not, adjust itself.

PAUKER: As an approximation it might be interesting to try that approach. But the problem with the CPC is the input and conclusions are in arbitrary order. Some of the conclusions are even wrong and there are no intermediate markers. So finding out where in the chain you went wrong is a problem. In addition the uncertainty remains that perhaps the CPC came up with the wrong diagnosis. The characteristic of medicine is that the data input is incomplete and part of the game with the CPC is the doctor is led down the wrong path because all the data is not given to him.

FEIGENBAUM: I'd like to throw out some numbers also. Simon estimated from some experiments in chess perception that a chess master holds between 50,000 to 100,000 facts about chess. The estimated number of words in a typical adult vocabulary is somewhere between 10,000 for the average person and 100,000 for the extremely intellectual person. Newell estimated that if he were to put together a model of the whole man, he would have about a million production rules. Now the question, is a million a hard number to manage? I think we would all agree that it is, given the kinds of mechanisms for representation we have been using. One thing to consider is, could we cause the necessary evocation to happen in one machine cycle by using active electronics in nets of demons instead of search electronics?. Each demon would be realized in an integrated circuit that would poke its head up when something came by. Now the cost of such a thing if you consider something like a ten property demon, might be about a dollar in the current state of electronics. So for a million dollars you have a million dollar memory which would evoke what is necessary to evoke in one machine cycle. And that's not absurd.

SZOLOVITZ: But that doesn't solve the problem of what you are going to put into the representation and how you are going to debug not the methodology, but the actual content. For instance, it is nearly an impossible task for a panel of clinicians to revise Harrison's Text so there are no errors in it. How can we overcome this problem of working in a domain in which we cannot certify that some new fact we add to the system is in fact correct?

DAVIS: I think we are in trouble if we reach that stage of simply putting things into the system without having any idea of whether or not it's correct. Steve said earlier that it is a near impossible task to follow down all the implications of a newly added fact. The alternative is not to follow it down, but to put it in anyway and wait for something to break.

PAUKER: Let me say something about the nature of the medical data base. It is not factual, it has errors in it. It evolves, it is self-contradictory. When students enter medical school they are told that half of what they will learn is wrong. The problem is that no one knows which half. So given that real world constraint, the data base must be inconsistent. Unless we can deal with that we're in trouble.

FEIGENBAUM: Who cares if there are inconsistencies? The processer can be set up to take care of it. Take Pople's scheme for example. It could be that some critical observation is an outlier and extremely important to the hypothesis. But because it is an oulier the inference scheme doesn't deal with it and that's a mistake. Fortunately, there is enough evidence redundantly available so that the inference of the correct hypothesis doesn't get demoted too much. So the inference scheme can be very tolerant of failures, of bugs, in spite of the fact that you don't check it all the way through in the data base. You sound like a bunch of mathematicians when you say if the system breaks that's it, you can't prove the theorem.

DAVIS: But that has been our experience in programming. A subtle mistake in one place leads to very strange things further down the line.

SRIDHARAN: I would like to suggest that the solution to handling multiple facts and finding contradictions in new occurrences lies in developing multiple representation. We should be able to put abstract concepts into the machine along with specific instances of those concepts so it can relate to them. For example, there is the kind of representation coming up in natural language work called scripts or scenarios which are really concrete instances of those schemes of inference which one immediately invokes in order to assimilate a new fact. These are all heuristic vehicles for handling this complex issue of representation. So the solution is not to design the best representation but to have at your disposal a variety of methods for looking at the various aspects of the same knowledge base.

SAFRAN: We've talked a bit about representation, acquisition and numbers of facts but very little about the eventual use of these systems. How many representations do we need to effect any kind of use?

BUCHANAN: There are many uses of knowledge. Each task domain has its own specific uses and if the representation depends critically on the use and there are no general principles to work with, then we are going to remain in this custom crafting mode, building separate systems for each task.

There is a MYCIN experiment in which we tried using its framework in other task domains diagnosing and recommending therapy for bugs in a Pontiac horn. People at SRI used the MYCIN structure to build a consultant for helping novice mechanics put together an air compressor and fix bugs in it. So we are finding the structure of the system useful in other domains. This was our first venture into a totally different domain and there is no claim that it was a grand success. We did it just to see what kind of things we had hidden away in the program that were purely medical that we wanted to clean out.

POPLE: I'd like to point out that the process we are talking about is something that in most professional education is considered to be if not unteachable, then at least the most difficult thing to try to convey to the student. The process of course is using the knowledge of a given discipline to solve real world problems. I think we have given the various professions some good insight into this process that they may use effectively. There is now a transfer from the computer programs back into the classroom that can take place. And it is not at the level of facts but rather at this process level, something that has been very difficult to articulate to students in the past.

********END OF PANEL DISCUSSION********

D.   METHODS OF INFERENCE - FORMAL AND CLINICAL PROBLEMS

TED SHORTLIFFE - MODERATOR

SHORTLIFFE: The topic for this panel is methods of inference.   I have a list of issues we could address in this session that come under the general heading of hypothesis generation and testing.

The first issue is how to quantify inferences. They may be causal or associational but we've all found a need to put a number on them. This includes knowledge that has been given to the system rather than what is actually derived during the process of reaching inferences.

A related problem is the accumulation of quantification numbers for the hypothesis. We've all had to handle the problem of relating positive and negative evidence as well as the clanker in diagnosis, that is, the one thing that seems to be against everything else. We have all had to design functions or algorithms for combining the numbers that have been accumulating in order to reach decisions.

Another issue relates to validating our models. If we start perturbing the numbers that we have from the outset does this really affect performance? If the numbers were available could we use statistical theory or are we dealing with issues that seem to go beyond statistics?

Can we define testing procedures that will convince ourselves and the observer that the kinds of techniques we are using for measuring inference are reasonable and justifiable at least in a theoretical sense?

To what extent are we trying to avoid issues of independence of evidence in favor of the hypothesis? We try to keep our rules separate and individually executable to avoid having to relate them explicitly to one another. And I think many of us have come up with schemes that allow us to skirt this issue mainly because we just don't know how to handle it.

KULIKOWSKI: There is a certain amount of uniformity among the clinical projects in dealing with quantification. Obviously, we have relied heavily on the clincians' judgement in acquiring these weights. One important issue in work of this kind is to relate these weights of evidence to some of the more objective statistical measures that one could obtain say, from a data base. Part of the problem in all of our systems is that they are over-determined in some sense. We have a lot of redundancy in them quite deliberately because we attempt to explain the structure of hypotheses in alternative ways. As a result if you want to validate or test one of our systems or acquire new knowledge, in some sense what one has to do is to freeze the part of the system that is under examination. And that is a very difficult job because we have often skirted the issue of interdependency as Ted has suggested. In our project we are very interested in seeing how far we can get with the independence assumption and where it breaks down. We haven't yet done

any formal study of this. On the other hand we are reaching a point where we often do need rather complex combinations of events to give us a useful clinical state to reason with. As we learn more about the necessary description of diseases and ways to reason about them, we will be able to extract those parts of the description that need strong interdependencies from those which do not. We've found in glaucoma as long as you stick to a relatively vague description you can maintain a very simple causal flow. The moment you want to characterize more precisely some of the interactions, you find that many things are not just a simple sequence of cause and effect but rather a set of interdeterminers in some form.

As for accumulation of quantification, all of us fall back on the notion of independence. But I think there are significant differences between composing things along a causal chain and composing on a purely taxonomic basis. On the whole, I would say there is more arbitrariness in a taxonomy than in a causal scheme, although we must be certain that the causal scheme is really causal and not just something we imagine to be causal which is part of our problem in the medcial domain.

BUCHANAN: The DENDRAL program does not present many of these problems of uncertainty. In the chemistry domain the inference mechanisms are largely stochastic processes. Essentially, we are able to get from the chemist predictive rules. These are probabilistic so there is some weight associated with them. The chemical structure is described and you expect to see evidence for certain actions. As a result of the action, new situations are produced for which there is some evidence and data. Now all of that can be run in a straight forward predictive way and there is really no inference problem there. The problem comes when you try to read those rules backwards. That is, from the evidence derive the processes and the fundamental situations from which those processes arose. In the Meta-DENDRAL program we are working with the same packets of information but they are arranged differently. Given some collection of evidence and a global structural description namely the whole molecule, infer the rules that one needs to use or test the program in either a predictive or inferential way. We tried discontinuous scales for our inference rules and found that they didn't work. The problem was that in different contexts "strong" or "weak" weights meant different things. We found we could do better on a more or less continuous scale.

MILLER: I would contrast DENDRAL with the medical systems because to replace the clinical experts with a meta rule forming system would be giving machines more responsibility than they can handle at least in the medical domain. I think that in this area the computer is not out on its own to derive clinical expertise because humans already have that. The problem is to apply that human expertise which is already in the system and I think most of the medical systems have done this.

PARKINSON: We have had a problem which is common to belief systems and that is knowing what to do with negation and reciprocal belief. What we've done is to use both. For instance, PARRY has the following beliefs:

The doctor desires to harm PARRY
The doctor desires to help PARRY

At first one might approach this by assuming that if he's not harming than he wants to help, and if he doesn't want to help than he wants to harm. But that is not the case. We have to add to each of these, the negation of it since the belief that the doctor doesn't want to harm PARRY still says nothing about his desire to help. Likewise, the belief that the doctor desires to harm PARRY still says nothing about his desire to help.

Another problem is that on a scale of 0 to 10 we start out with a belief system that contains zero information except for the initial assumption that he's probably the doctor and that he probably does want to help. We have found in our model and we believe this happens with humans in real world situations that as one gathers evidence to affirm a certain belief, in this case it would be that the doctor does want to help, it tends to get believed strongly enough so that any further evidence that might challenge that belief gets explained away. So a belief can start from 0 and rise up to 10 and if counter evidence comes into play it may have little affect on it. We believe for humans if there isn't too much counter evidence to challenge a belief it probably does not change unless it is really important. Likewise in our system we look at the importance to the model of inferring that belief. For example, it is fairly important to find out if the doctor is trying to help us. It is very important to find out if he is trying to harm us. So if we decide that he really doesn't want to harm us and then some counter evidence appears and indeed the doctor starts attacking us then certainly this is important enough to alter the initial belief. There is also a problem when both the positive and negative evidence say, for the doctor's desire to help is of equal weight so that neither one is believed.

I have one last comment about the strategy in the system itself. All these mechanisms are related to the original reason for proving or disproving the belief and that is self-interest. In order to make certain actions possible we have to find out if the environment allows it. And at that point we try to infer belief. It is not as if the program tries to prove everything it can. It wants to do something and it makes these inferences to find out if indeed it can.

PAUKER: Something intersting happened with a program we developed. A wrong number was accidently inserted and about a month later I discovered it but the program had worked anyway. I changed the number put in a different one and it still worked! And that really raises the issue of whether a specific number really does make a difference. Perhaps there is a simpler mechanism. To some extent the method of inference is embodied more in the links than it is in the measures you put on the links. I think if you have the appropriate links, the apporpriate structure of the data, the exact quantification that goes on there probably is insensitive within a reasonable range. It is strange that we could put relatively arbitrary untested numbers in it and still have it work.

As a physician my view is that the key to the program's performance is experts. It is not more facts or numbers it is the doctor using more interconnections and heuristic rules. And I don't think these kinds of heuristics can be built into numbers. So that right numbers or

algorithms really don't make any difference.

SAFRAN: I think this gets us into the issue of the credibility of any medical system. Given a data base and a set of numbers it is very important to be able to explain to a physician who is using them how these numbers were arrived at and their relative importance and how the system goes about reaching a decision or a hypothesis. The arbitrary assigning of numbers leads you away from credibility.

SCHMIDT: I would like to reinforce that statement. There is very little you can say to the expert when he wants to know how you came up with your answer. And I think that is something worth considering if we hope to attract other experts to the system. They are the responsible persons in this case and they must have all the necessary information with which to evaluate the system.

As a psychologist I use numbers all the time but I've avoided using them in common-sense reasoning because I find I need something more symbolic. Typically, I'm working in a world of partial matches, the evidence only partially matches the entire rule I'm looking for. To substitute these symbolic rules means I do have a residual after that partial match whereas with a number I just have a difference. There is no further computation I can do in my system with that difference between the number I would have liked, say probability 1, and the number I got of probability 8. So I think if you want to organize very complex evidence you probably will do well to stay away from numbers.

SHORTLIFFE: Certainly your first comment is a potential problem in our system and probably one for all these systems. There are numbers that guide our rules and we've gone to great length to implement some capability to explain the reasoning. The expert may ask us how the program reached the conclusion and we can list for him the six rules. And each one of those rules may look just great to him but he simply cannot accept the conclusion.

MILLER: In our system we have found the actual number in any particular instance plus or minus 1 doesn't make a lot of difference. But I think doing away with numbers or saying numbers aren't important is something we really can't do. I ran an experiment whereby I wiped out DIALOG's evoking strength and gave it equal weights in terms of confirming a diagnosis. I then used this altered version of DIALOG in a case it had solved previously. Its behavior was very different and it didn't perform the way you would expect a physician to perform in terms of coming to a diagnostic conclusion. So these experiments showed us that the numbers do matter quite a bit.

SILVERMAN: As computer scientists we learn to deal with numbers quite a bit and I think there is an over- propensity toward looking at numbers for answers. Once we began to define our model and associate more limits between items that were coming in, the actual numbers that we were using became unimportant. What happened was we got it to a tertiary system and that seems to work just fine because we have a thorough enough model. So instead of having a range of seven or eight possible values, we have three along with a great deal of information as to how to choose which is the appropriate one.

SHORTLIFFE: You are saying that a discontinuous three valued scale seems to do very well. If proper associational links between evidence exists, do you think you can simplify the numbers more and more?

PAUKER: Let me add one point to that. When we talk about a discontinuous three valued scale I think we mean using that to measure strength and belief.

SHORTLIFFE: Yes.

PAUKER: The three by three matrix that Howie talked about, that is, toxic, a little toxic, not at all toxic, is a statement matrix. It is not a level of belief matrix.

SHORTLIFFE: It is to the extent that a set of observations about a patient has got to be mapped in one of those states. So there is some element of belief about which state the person is in which is reflected in the three values.

POPLE: I had a strong aversion to the use of number at first but it became clear that in going through cases, Dr. Meyers did use terms that definitely suggested the strength of associations. So we found in the language of the clinician relationships which we eventually had to incorporate into the data structure of DIALOG. And I don't think it is all that difficult to take these numbers and translate them back into the kind of terms or ideas that they were intended to convey in the first place.

KULIKOWSKI: Our approach was slightly different. I started out being quite a lover of numbers having worked in a number of pattern recognition applications. My motivation in moving away from them was because I found them unsatisfactory for explaining the structure of our reasoning to a clinician and more significantly because if a numerical method alone doesn't work you are not able to trace back symbolically that residue that Chuck had mentioned.

In the early stages of our system we removed the numbers from the causal links but kept the numbers between the evidence and each state. The system worked comparably well in doing that. So I would say if you go from a structure of a subgraph of the causal net to a higher level hypothesis the numbers can be important if you are dealing with a large number of hypotheses the way Harry is doing. When you are dealing with only a few hypotheses the mapping can be deterministic. We've defined the problem so well by the causal subgraph that it is one to one mapping.

COMMENT: I don't see how the use or absence of numbers has anything to do with the difference between the CASNET model and the DIALOG model. I feel it is solely due to the degree to which you have been getting close to the metabolism. If the ophthalmologist has a very good understanding of the metabolism then you have relatively firm linkages that can be described in a binary way. Now we are very far away in the general case of internal medicine from having such detailed understanding so that we approximate a more complex situation statistically by linkages to which we assign values.

MEYERS: I don't think it is a difference in level of understanding but as Cas said, the complexity of the problem. I think we can take any subset of internal medicine and follow exactly the same rules that we are talking about. It is the complexity of the problem that requires numbers so that you can keep your facts straight.

KULIKOWSKI: I would tend to agree with Dr. Meyers. To go between levels one needs numbers. But once you are at some level of understanding you can operate symbolically.

COMMENT: I think everyone would agree that the purpose of AI is to produce machines that will do intelligent things at some level. And if they do things intelligently the way people do them they inherently run into the same kinds of errors that even experts can produce. So the point of view could be taken that by using a quantification scheme with a consistent numerical process, even though the machine has been up for 48 hours, it is more likely to give a consistent answer than the physician who has been up for the same amount of time and not at peak of performance level. So I think a good argument can be made for a quantification scheme that it does at least have the virtue of being consistent if nothing else.

PAUKER: At the current level of technology, do machines stay awake that long?

SAFIR: I think we are probably at a stage now of complicating the assumption and getting further from the forest for the trees. We go through stages like this where things begin to look more and more complicated and after a while somebody backs up and looks at it critically and offers another simplified hypothesis. We are right now at a phase of complicating the science and waiting for that next step. Somebody once said that all diseases come down to the simple phenomenon of a tube getting plugged up somewhere and it is true. You get very involved in the clinical richness until someone comes along and finds out by electron microscopy that it is a tube getting blocked. Things work in amazingly simple ways but we organize them in our thinking in ways that are complicated and have nothing to do with what is actually happening. And the numbers don't really exist, they happen to be a good cerebral mechanism for dealing with ideas that we cannot handle otherwise. I think at this stage our representations are quite imperfect and it would be nice to have been back when we thought we knew what we were doing. That must have been a comfortable time before we had machines to test these hypotheses.

*******END OF PANEL DISCUSSION*******

E. PROBLEMS OF SYSTEMS DEVELOPMENT

SAUL AMAREL - MODERATOR

AMAREL: This panel will discuss the management of systems development. We will try to get a feeling for the more practical aspects of managing projects, and share problems, advice and experiences we have had in collaborating across disciplines. It is the counterpart but no less important part of the project oriented, scientific activities we have discussed so far.

I want to start by asking Ed Feigenbaum who has had experience in large projects for almost ten years that involved the application of AI in scientific and medical programs, to tell us of his own experiences.

FEIGENBAUM: Let me say something about experts because they represent the kernel of what it's all about in the knowledge based system design area.

In discussing how one picks an applications area in AI heuristic programming in particular, aimed specifically at Medicine, I listed as one of the criteria under knowledge base: "Is there in your environment at least one highly knowledgeable, highly motivated, computer oriented and computationally sensitive expert who can serve as an informant, through whom the knowledge base can be acquired?". One can partition the classes of experts into the computer oriented and computationally sensitive experts, and those who are not. The only place I've ever gotten into trouble in work in a knowledge based system is the one place I had an expert who didn't know the first thing about computers. The kind of mental model a person has about what a computer can and cannot do is extremely valuable. Without that it's hard to make any progress at all, as the person has no scale of measurement against which to suggest an idea.

The computer oriented and computationally sensitive experts break down into three classifications. There are the area experts who are not computer science oriented, but who understand scientific research. There are the quasi-computer scientists who know a great deal about computer science and technology and could be computer scientists or not depending on in what university they sat. Examples of such people are Lederberg and Ray Carhart. Those are people about whom one often has guilt feelings. That is, they are so good at what they do, in let's say chemistry, one feels very guilty for having yanked them that far over into computer science.

Then there is the very special brand of experts represented by Ted Shortliffe, who could call himself a professor of computer science or professor of medicine or a practicing doctor. The equally rare complement to that is the computer scientist who will make the trip more than half way into somebody else's discipline. It is becoming increasingly difficult to find applications oriented computer scientists who are willing to become minor experts in somebody else's domain in order to translate the conceptual terms. They just don't see the

payoff.

Let me talk about payoff. You have to arrange that the expert you find sees the payoff in what it is you want to do. You may not be able to demonstrate that on day one, but you have to have some way of getting to a point where that person gets to a terminal or at least a seminar in which the payoff is made clear to him. So there are two problems. First, get right into the heartland of that expert's domain. Then plan for incremental payoff so that you can get over the first threshold and sustain his interest. He has to see that those first five facts he put in made a difference, or he won't put in the next five after that.

After you manage to bootleg some resources to get to stage one of having a credible running program that can serve as the platform for an NIH proposal, plan that the very first renewal application for that proposal involves a study section of at least half that expert's peers. You want a discipline oriented evaluation as well as a computer science evaluation at the very first stage which is, say three years down the pipe. Then if you plan to carry this on for more than the period of the first renewal, plan for an almost totally disciplinary evaluation at the end of it for the second renewal. Nothing is guaranteed to make the expert more attentive to getting knowledge into the program than knowing he is going to display it to his own peers.

Finally, the computer scientist and expert must be assured of a very good computing resource with adequate amounts of computing right at the beginning, and a level of sustained funding that is reasonable. And by that I mean nothing less than three years is reasonable for an effort in this area. If you can't get a three year go right away at it, don't bother. It's just not worth getting into if you have to struggle with these problems of resources.

SAFIR: I'm sensitive to the peculiar role of the medical practitioner doctor who gets involved with computer science because it's happening to me and it's an involuntary act. The doctor who gets involved is likely to be a full time academic doctor who works in a medical college or teaching hospital. He is different from the practitioner who delivers health care in that he is more likely to be involved in trying to improve the appications process and not someone who practices application all the time. He is an investigator. And that's a different sort of person from the practitioner. There is a big spectrum of doctors, and you are being exposed to a very biased sample here. I would in no way hold myself up as a typical ophthalmologist. I think freak is probably the best word for people who are interested in something this far from the practice of medicine.

The academic investigator doctor finds that if he's going to be a good investigator, he has to learn how the tools work in order to use them intelligently. He finds himself involved in computer science without being able to help it. And if he has the particular cast of mind, it becomes an exciting new discipline and he finds himself putting a foot into that camp. I think medical doctors who start out as practitioners as I did are going to be replaced by people who went about it in an orderly and disciplined way, like Ted Shortliffe who decided to learn both disciplines from the ground up and then put them together in

one head.

I think that the problems of collaboration between doctors and computer scientists are far more complex than they seem to be. It's not just a question of getting a good doctor interested in computer science and learning the technology. I think you ought to try to capture medical graduate students right at the outset with displays, devices and services that they can understand, that are non-threatening. They may seem terribly mundane to you, but they are things that doctors want and understand. Once a doctor learns that he can get a useful service from the computer through having fun at a terminal, then you've got him. Then you can entice a larger and larger percentage of them into doing something more scientific.

SMITH: There are several questions of collaboration across disciplines, in our case computer science and chemistry. I would use a broad definition of interdisciplinary collaboration and include the various subdivisions of chemistry. There is much collaboration that can go on, and SUMEX of course, provides one mechanism for doing this. But it hasn't removed all the difficulties. Some people for example, wouldn't mind so much being users of our programs, but the mention of collaboration conjures up perhaps some interference in their own particular research projects. I think the way to extend the kinds of things we're doing to an outside community is to demonstrate utility, and to provide them with information that is difficult to get in any other way. The traditional method of demonstrating utility has been to publish papers in the literature. But that method breaks down when you are talking about computer programs which are applied to chemistry. There is no way you can describe a computer program of any complexity to enable another chemist to replicate it. Again, we have the hope that SUMEX will provide a mechanism for removing some of these difficulties. We hope it will allow chemists to get their hands on a program, try it out, and see what it can do in a problem of their own definition.

AMAREL: I would say that a primary challenge for the AIM community and especially the AIM workshop in terms of goals, is how to transmit current stages of development of a very complex program, not only to collaborators but to other interested people.

YAMAMOTO: Perhaps we need some different technique for scientific communication than the traditional journal article in this area. These articles are based on the fact that the science performs essentially a demonstration on an existing natural object. Whereas, in Computer science in many cases the science is creating the object on which it is also demonstrating. The problem of publication therefore is simultaneously to report the demonstration and to make available to another interested scientist the opportunity to either verify, perform or alter. And this cannot be done by writing a journal article. The root problem that this community faces therefore, is the problem of creating a new science, where science is a social activity. And it is in this sense that AIM's sharing facility is a very important component of the future of those who want to continue working in applied areas of AI.

RINDFLEISCH: Through the networking and the direct contact with these programs, people can start to share code, and take these concepts more directly than from published journal articles which give only conceptual descriptions of what is going on.

AMAREL: Don Lindberg, as chairman of the AIM advisory committee, has raised some very fundamental issues about what this community should be doing and how it should be interacting.

LINDBERG: I want to comment on the two matters that Ed Feigenbaum and Aran Safir raised, as I can agree with everything both of them said.

Ed suggested that a permanent alliance must be made between the computer scientist and medical man. I would extend that by including an alliance between the medical man and bioengineering. And if you can't anticipate working together for five years, it's probably not going to be profitable. It doesn't mean that nothing will come out before that, but it takes at least that much time before an easy working relationship matures.

With regard to what Aran said about ophthalmology, if you look at the historical sequencing of medical specialists becoming involved with computing, I think that right now ophthalmology is new to computing, and some of the pizzazz elements that you are after stem from that specialty having just now gotten ready to be interested. Probably the first group of clinical specialists to be users of the computer were the pathologists because they produce lots of numbers in the laboratories and it made sense. Next was radiology because they were quantitative people too. They had recording problems and image problems. These two groups are settled down and they are essentially, as specialties, committed to computing. They're locked in. Their views toward the kind of applications you give them are quite different now then they were fifteen years ago. Just a week or so ago, my colleagues were dealing with a problem of a wholly automated magnificent AVL blood gas machine which comes with a little micro processor on a card. The manual for a technician is such now that all you have to do is hold the tube in your hand and find the hole in the machine. It turns out that they had enough operational problems with it, even though it's a glorious methodology, that they're going back to an IL which is essentially a semi automatic machine and they are making the decision without a backward glance. They are fully able to evaluate the nice technology and trade it off against better performance. They are really launched. The ophthalmologist will be stuck with computers too and their desires will shift as the association matures.

That leads me to the the point I wanted to make about collaboration. I would urge you all developing this collaboration, particularly the computer side, to be very very slow about promising working systems and urging your colleagues to use them for others until you are really ready. The difference between coming up with a program that can be demonstrated on SUMEX at a meeting like this is an order of magnitude away from making that a working system. There are all sorts of people arrangements that are necessary to make it work. If your chemist collaborator is using the system and only he observes that it doesn't work, that's one thing. But if you are serving the

ophthalmologist who is going out and making promises to his colleagues and his patients, or the pathologist who is running a service for the whole institution and those go sour, well, that's why people move from one school to another. The partnership is no good unless both parties feel they are winning. Assuming that can be accomplished and patients and colleagues don't get hurt and feelings are not bruised, then the obvious question is what is the computer scientist going to get out of it? Well, he wants to get some papers which are of importance to the field of computer science. And often the first impulse is to generalize everything, which is a good scientific approach I guess. But if it takes the form of looking at a simple problem and making it hopelessly complex merely so that it can be described in a jazzy new terminology, my advice is don't do it. When something can be done simply, do it simply and be proud of having made life simpler and not more complex.

So, if you are going to commit yourself, make something that your collaborator is seriously concerned about and is going to use. Use the best methods and don't make it complicated if you don't have to.

AMAREL: Bill Baker of NIH is involved in the management and administration of all these enterprises.

BAKER: I was a biomedical engineer before I ever went to NIH, and chairman of a biomedical engineering department that was multidisciplinary. So I'm not part of this community, I'm an outsider looking in.

Our annual budget at NIH is around 12 million a year. It's very very small. We also have in other agencies within HEW, programs that have direct impact on your activities. These have a problem that I'd be concerned about if I were on your side of the fence in that some of them are sheer impulse functions. They come and they go. Unfortunately, a great many people here are being supported today by these impulse type programs. Other programs at HEW have an instability in size of the activity. The technology supported by the National Center of Health Services Research seems to have a very indefinite size. It almost changes week by week. They have a lot of money one week and none the next. These other programs differ from ours in that NIH has a very carefully prescribed area of responsibility in the health enterprises, and that is basic and clinical research or health knowledge. These others deal more with the patient care and health services system. And NIH is being pressured by congress right now to move over into what NIH calls disease control and demonstration. It's a very big issue and when we meet with Dr. Yamamoto, it is very frequently the most important thing that we discuss. It has to be worked out so that the nation's needs can be met without diluting or sacrificing any of the effort that is going on in the basic and clinical research activities. We are supporting research that is also supported by sister agencies and we will continue to do this. The rationalization that I see for doing this is that all these projects have the subset of AI dealing with organizational disease built into them. SUMEX/AIM is not the only nationally shared resource that we deal with. Our concerns in such activities lie in that we have developed a certain capacity of high technology, of complex methodology that can be shared across the country. But the mechanisms that are most appropriate to marry the

collaborators to the system, the financial support for these activities, we are very concerned about and are working on. In fact, I've come up with a new resource called an interface resource which Saul Amarel really represents. He started out with no equipment and interfaced his collaborators to the hardware that was not under his resource. He's quickly adding to the resource capabilities and will go into the regular category out of the interface resource. Hopefully, the decision makers of NIH and OMB and congress will think this is a good idea and additional funds will be put into our program to bring more attention to this method of work.

FEIGENBAUM: I'd like to make a point that relates Don Lindberg's comment about not letting systems out too soon, to some of Bill Baker's comments. It has to do with the nature of the enterprise and our ability to sustain it over a period of time.

I think we have failed to persuade society's decision makers that the pursuit of intelligence in machines is of great value to society and ought to be pursued as a long range endeavor though it's difficult and expensive. There was a time when we had almost no one to make happy as a result of our research because there was plenty of money to go around for all at NIH and other funding agencies. Now the emerging science is being asked what it has done for society lately, and being given eighteen months in which to answer. Whereas, in the sixties we had ten years to answer. In most of what we do, we haven't yet reached the first order consumer and unless we conduct some kind of field testing for our programs and come up with numbers that justify our work and show that an intellectual task is being performed better or cheaper, with some greater social utility than it was before we did it, the interest of society in sustaining this research is not going to last beyond the first SUMEX/AIM grant. Academic collaborators are accustomed to getting about 70% of the way in developing usable programs. We are not good at engineering products. We need to construct a mechanism whereby someone else who likes the task of going the last 30% of the way and is good at that task, takes it over for us. As it stands now, when it comes time to make a DENDRAL work for the consumer, we have to do it. We'd like to move on to other things, but to sustain ourselves we have to do it. So we need to invent this other kind of institution that will close that last gap, that will make our systems usable at least up to the stage of some kind of user evaluation.

BAKER: Ed, you should read our guidelines in biomedical engineering resources because we have that built in. Now when DENDRAL is ready to go from computer science into this phase, you just switch modes with a whole new resource approach. The mechanism really sits in our program to make this transition.

FEIGENBAUM: But who is going to do it?

BAKER: It's a different kind of mode. We have one biomedical engineering resource we support now under this set of guidelines in micro-electronics at Case Western. Its advisory committee has membership from industry to give advice on this. Its mission is to carry our enough collaboration with the prototypes that it develops to take the risks that the interested stock holders of the company would

normally take. NIH is willing to go that far.

LINDBERG: I feel totally resonant with Ed's remarks. I don't think that preparing a guideline creates the people out there to do the job. University people are creative, that's their strong suit and nailing together a finished production system practically never happens.

BAKER: It lets them know where they can get support to do it.

YAMAMOTO: I'd like to make a comment that deals with marketing and the issue of creating a science.

Science is sustained by a cloud of people in society who are sympathetic to it, many of whom practice a sub version of it or identify themselves with it. To build a science you must build a broad pyramid in the society that recognizes that what you are doing is akin to what they are using, and it is the small actions that you take that create a science. If you market an idea using an agency of the government and the good will of a portion of the academic community, you can then begin to create that cadre of sympathetic individuals in the society who use computers for intelligent tasks. There are now at least ten companies that say they are selling AI kinds of things, and there are many hospitals that are using intelligent activities performed by machines. And the next thing we succeed in selling in the medical care area will sell that much more easily. So you really should pay some attention to the marketing of ideas because it is a very significant portion of scientific endeavor.

FEIGENBAUM: I think you have just highlighted again the need for an institution that likes to do that.

YAMAMOTO: If you should find someone that wants to sell your ideas or promote them you should not regard them as either inferior or in a different area. You should always be willing to embrace that cloud of other activity around you or else you will not be embraced by society. I don't like the market. My background is just as esoteric in a different science as yours is. But I've come to realize recently the importance of the market in my field.

COMMENT: I wanted to point out that there is some pure basic scientific research that has to be done in AI, and we don't want to alienate those researchers from doing their work in the problem domain of medicine. Even though their product may not be sellable, their work is just as important to this community as producing DENDRALS that are going to be sold.

SAFIR: Marketing implies that you first find out what is needed and that's much more complex than it seems. If there is a great need for a product in the community of users of medicine, it is not for a diagnostic aid in serious metabolic or internal medical diseases. It is for something so terribly mundane that the academicians are not interested in it. I think that the problem may be almost insolvable in that the community providing solutions is not interested in studying the problems people want solved.

AMAREL: Before we conclude, I would like to have a more general discussion about the workshop itself, its content and suggestions for preparing next year's workshop.

RINDFLEISCH: We will put a list of the participants and their addresses and telephone numbers in a file on SUMEX so people can get it right away.

COMMENT: I suggest that for future workshops we have proceedings or some mechanism of publication of people's ideas. It would be helpful to have something that would inform us of what will be discussed beforehand and then take with us to peruse after having seen the systems and having run them.

AMAREL: We are going to make an effort to put together something which can approximate proceedings and will give a fairly concise account of what went on at this first workshop, and have it on file in SUMEX/AIM. This can be done and we will try to have this accessible to at least those people who can access SUMEX/AIM which is a fairly large population. As far as other publications are concerned, I am not certain.

YAMAMOTO: Let me give the problem a challenge. You are AI people with an AI facility. Your AI report of the proceedings of a workshop like this ought to be some type of mnemonic recollection in your AI systems. It should record in your machine base systems, what as a consequence of this workshop has happened in your respective AI prototypes. It seems to me that if you have a true AI philosophy, your AI system ought to grow as a result of encountering. I think that would be the ultimate form of publication for this community rather than the bound proceedings, even though the distribution base in this would be very narrow, perhaps only to those people who have attachments to SUMEX. I don't know how to do it, I can only imagine what it might be like. But it ought to be something uniquely different, and that's a challenge.

I think the workshop was too long. I heard on the first day some comments about who is allowed to stay and who is not and so on. I think you ought to blur that edge next time so it isn't so sharp.

AMAREL: I think the publication is a marvelous challenge, and a formidable one. Your challenge is in five years to get to that point. What do you think Bill, we can do it?

BUCHANAN: With respect to that issue there are some things that are going on at a very low level in MYCIN. We try to keep track of the author of a rule, and that is some sort of an acknowledgement of how the system if growing. In SUMEX itself we are trying to work on a bulletin board facility for dissemination of informal ideas, where you can be notified when something of interest to you is posted by someone else.

SRIDHARAN: Bruce mentioned that there is a long lead time between the actual birth of an idea and the time you put it dowm on paper. And one of the things I was looking forward to at the workshop was hearing nascent ideas that I have not already read. But I think it has worked out quite the contrary. Therefore, for the next workshop, I really hope

that we get the fresher ideas from everyone and not a rehash of what we have already written up or thought about.

SAFIR: Writing up your ideas in an organized form is a difficult task, and busy researchers are not likely to do it unless they get a significant reward which is generally a recognized publication. If the National Library of Medicine would recognize a file that could be accessed by anyone who had a printout device so that one could get this electronically stored manuscript that would have to be suitably refereed, I think you might start an electronic journal that would have value and would be rapidly responsive to people's thoughts.

FEIGENBAUM: To follow up on what Sridharan had to say, one of the consequences of the workshop could have been an open forum in which people suggested to each other the next set of problems to work on and not just absorb what has been done already in various projects. There are many problems yet to be tackled in the computer science area and also problems from the medical domain.

AMAREL: With these suggestions with us, let us bring this first AIM workshop to a close. In behalf of all my colleagues I wish to thank first and foremost, Cas Kulikowski who is the organizer of the workshop, N.S. Sridharan who worked with him throughout on many different problems and Saul Levy who organized the computing activities with respect to the workshop. I also want to thank Pat Moore and Ken Brown who were vital to its success, and our graduate students here at Rutgers for their valuable help. Lastly, many thanks to the AIM advisory committee, the AIM executive committee, and of course, SUMEX/AIM itself which provided a very useful way of working and planning for this workshop. Thank you very much.

********END OF PANEL DISCUSSION********

VI. REFERENCES

1. Smith, D.H., Masinter, L.M., and Sridharan, N.S. (1974) "Heuristic DENDRAL: Analysis of Molecular Structure". In Computer Representation and Manipulation of Chemical Information (W.T. Wipke, Editor) John Wiley.

2. Buchanan, B.G. (1974) "Scientific Theory Formation by Computer", in Computer-Oriented Learning Processes, (Proceedings of NATO ASI, Bonas, France).

3. Wipke, W.T. (1974) "Computer Assisted Three-Dimensional Synthetic Analysis" in Computer Representation and Manipulation of Chemical Information . (W.T. Wipke, Editor).

4. Colby, K.M. (1973) "Simulations of Belief Systems" in Computer Models of Thought and Language. W.H. Freeman.

5. Schmidt, C.F. (1975) "Understanding Human Action: Recognizing the Motives". Computers in Biomedicine, Technical Report 45, June 1975, Department of Computer Science, Rutgers University. Also to appear in Cognition and Social Behavior, J.S. Carroll and J. Payne (Eds.), New York: Lawrence Earlbaum Associates, in press.

6. Sridharan, N.S. (1975) "The Architecture of the BELIEVER System, Part I: General Description and Illustration of the Inference Mechanism". Computers in Biomedicine, Technical Report RUCBM-TR-46, June 1975, Department of Computer Science, Rutgers University.

7a. Weiss, S.M. (1974) "A System for Model-Based Computer-Aided Diagnosis and Therapy", Computers in Biomedicine, RUCBM-TR-27-Thesis, Department of Computer Science, Rutgers University, June 1974.

7b. Kulikowski, C., Safir, A. and Weiss, S.M. (1973), "A Representation of Medical Knowledge for Problem Solving: Application to a Model of Glaucoma", Computers in Biomedicine, RUCBM-TR-21, Department of Computer Science, July 1973.

8. Pople, H.E. (1975) "Artificial Intelligence Approaches to Computer-Based Medical Consultation", 1975 IEEE Intercon Conference Record, Session 31, April 1975.

9. Shortliffe, E. (1974) "MYCIN: A Rule-Based Computer Program for Advising Physicians Regarding Antimicrobial Therapy Selection", Computer Science Department, Stanford University, STAN-CS-74-465, October 1974.

10. Shortliffe, E.H., Davis, R., Axline, S., Buchanan, B., and Cohen, S. (1975) "Computer-Based Consultations in Clinical Therapeutics: Explanation and Rule Acquisition Capabilities of the MYCIN System", Computers and Biomedical Research, August 1975.

11. Miller, P.B., (1975) "Strategy Selection in Medical Diagnosis", Report MAC-TR-153 (M.S.Thesis from MIT), September 1975.

12. Pauker, S., (1975) "Towards the Simulation of Clinical Cognition", submitted to American Journal of Medicine, June 1975.

VII. AIM Organization

AIM EXECUTIVE COMMITTEE 1975:

LEDERBERG, Dr. Joshua
Principal Investigator of the SUMEX/AIM Project
        Department of Genetics, S331
        Stanford University Medical Center
        Stanford, California 94305

AMAREL, Dr. Saul
Principal Investigator of the Rutgers Research Resource
        Department of Computer Science
        Rutgers University
        New Brunswick, New Jersey 08903

BREWER, Dr. Carl R.
        Biotechnology Resources Branch
        National Institutes of Health
        Building 31, Room 5B25
        9000 Rockville Pike
        Bethesda, Maryland 20014

LINDBERG, Dr. Donald
        605 Lewis Hall
        University of Missouri
        Columbia, Missouri 65201

AIM ADVISORY GROUP 1975:

LINDBERG, Dr. Donald   [Chairman]
        605 Lewis Hall
        University of Missouri
        Columbia, Missouri 65201

AMAREL, Dr. Saul
        Department of Computer Science
        Rutgers University
        New Brunswick, New Jersey 08903

BREWER; Dr. Carl R.   [Executive Secretary]
        Biotechnology Resources Branch
        National Institutes of Health
        Building 31, Room 5B25
        9000 Rockville Pike
        Bethesda, Maryland 20014

BOBROW, Dr. Daniel G.
        Xerox Palo Alto Research Center
        3333 Coyote Hill Road
        Palo Alto, California 94304

FEIGENBAUM, Dr. Edward
        Serra House
        Department of Computer Science
        Stanford University
        Stanford, California 94305

FELDMAN, Dr. Jerome
        Department of Computer Science
        University of Rochester
        Rochester, New York

LEDERBERG, Dr. Joshua   [Ex-officio]
        Department of Genetics, S331
        Stanford University Medical School
        Stanford, California 94305

MILLER, Dr. George
        The Rockefeller University
        1230 York Avenue
        New York, New York 10021

REDDY, DR. D.R.
        Department of Computer Science
        Carnegie-Mellon University
        Pittsburgh, Pennsylvania

SAFIR, Dr. Aran
        Department of Ophthalmology
        Mount Sinai School of Medicine
        City University of New York
        Fifth Avenue and 100th Street
        New York, New York 10029