fundamental research experiments with limited clinical usefulness other than as demonstration projects. Our past experience has shown, however, that SUMEX provides a superb vehicle for demonstrating systems, even at a distance.

The new book writing effort will in particular be facilitated by SUMEX, much as the *AI Handbook* was in the past. A multi-authored text of this type, particularly one for which the authors are spread at numerous different universities around the country, would be a nightmare to compile if it were not for the SUMEX resource. Many of the contributors to the book have been assigned SUMEX accounts for purposes of manuscript preparation. Online manuscript work through the shared facility, coupled with messaging capabilities, will greatly enhance the efficiency and accuracy of the developing chapters and the editing process.

B. Sharing and Interaction with Other SUMEX-AIM Projects

Although our EXPEX work is young, we are already benefiting from interactions with other researchers who use the SUMEX-AIM resource. The NESTOR work in particular has depended on access to the INTERNIST-1 knowledge base and on frequent exchange of messages with the researchers at the University of Pittsburgh. Similarly, our collaboration with the GUIDON research team for the implementation of an explanation capability would not have been possible without the facilitated communication and shared file access available via SUMEX.

C. Critique of Resource Management

SUMEX continues to provide a superb environment for research of this kind. Not only is the 2060 a well managed resource under Ed Pattermann's leadership, but the hypothesis assessment and graphical query systems are dependent upon access to high performance professional workstations, and we are delighted with the resources that SUMEX has provided us in this regard.

III. RESEARCH PLANS

A. Project Goals and Plans

We anticipate completion of many of these basic research efforts during the coming year. Cooper's NESTOR work is largely complete, and a thesis document is anticipated in June 1984. Similarly, Kunz has completed his work on AI/MM, and his dissertation is approaching completion. Both Cooper and Kunz have completed their oral examinations on this work.

The project of Tsuji is complete and she has now left Stanford. However, the code she developed is being modified for ongoing use in the ONCOCIN environment.

The project of Langlotz continues to be an active research effort within the ONCOCIN project. His plan for the coming year is briefly outlined in the ONCOCIN portion of this annual report.

The work of Rennels, which is just getting underway, will be better formulated by next year at this time. We expect the project to last at least two more years, however.

The textbook preparation is scheduled for completion in approximately one year, with publication anticipated during 1985.

B. Requirements for Continued SUMEX Use

All the work we are doing is largely dependent on the SUMEX resource. The new work of Rennels is using Hewlett-Packard 9836 workstations owned by the Medical Information Sciences training program, but Dr. Rennels continues to be dependent upon SUMEX for communication and collaboration. Of the other projects, only the hypothesis assessment and graphical query projects are sufficiently mature to justify their transfer to one of the SUMEX personal workstations, so the new 2060 continues to be a key element in our research plan.

In addition, we have long appreciated the benefits of GUEST and network access to the programs we are developing. SUMEX greatly enhances our ability to obtain feedback from interested physicians and computer scientists around the country. As our programs continue to mature, it will become increasingly important that we be able to make them available for demonstration and for access by distant collaborators via the SUMEX network.

C. Requirements for Additional Computing Resources

The mainframe machine should continue to provide a suitable environment for most of our work in the months ahead. We have no plans to transfer NESTOR, or AI/MM to other hardware soon.

D. Recommendations for Future Community and Resource Development

We are very satisfied with the facilities SUMEX has provided since the upgrade to the DEC 2060. Other than continued acquisition of professional workstations that can be shared by some of the more mature programs in this set of projects, we have no requests for additional acquisitions or resource development at this time.

II.A.1.2. GUIDON/NEOMYCIN Project

GUIDON/NEOMYCIN Project

William J. Clancey, Ph.D.
Department Computer Science
Stanford University

Bruce G. Buchanan, Ph.D.
Computer Science Department
Stanford University

I. SUMMARY OF RESEARCH PROGRAM

A. Project Rationale

The GUIDON/NEOMYCIN Project is a research program devoted to the development of a knowledge-based tutoring system for application to medicine. This work derived from our first system, the MYCIN program. That research gave way to three sub-projects (EMYCIN, GUIDON, and ONCOCIN) described in previous annual reports. EMYCIN has been completed and its resources reallocated to other projects. GUIDON and ONCOCIN have become projects in their own right.

The key issue for the GUIDON/NEOMYCIN project is to develop a program that can provide advice similar in quality to that given by human experts, modeling how they structure their knowledge as well as their problem solving procedures. The consultation program using this knowledge is called NEOMYCIN. NEOMYCIN's knowledge base, designed for use in a teaching application, will become the subject material used by a family of instructional programs referred to collectively as GUIDON2. The problem-solving procedures are developed by running test cases through NEOMYCIN and comparing them to expert behavior. Also, we are using NEOMYCIN as a test bed for the explanation capabilities that will eventually be part of our instructional programs.

The purpose of the current contract, now in its sixth of six years, is to construct an intelligent tutoring system that teaches diagnostic strategies explicitly. By strategy, we mean plans for establishing a set of possible diagnoses, focusing on and confirming individual diagnoses, gathering data, and processing new data. The tutorial program will have capabilities to recognize these plans, as well as to articulate strategies in explanations about how to do diagnosis. The strategies represented in the program, modeling techniques, and explanation techniques are wholly separate from the knowledge base, so can be used with many medical (and non-medical) domains. That is, the target program will be able to be tested with other knowledge bases, using system-building tools that we provide.

B. Medical Relevance and Collaboration

There is a growing realization that medical knowledge, originally codified for the purpose of computer-based consultations, may be utilized in additional ways that are medically relevant. Using the knowledge to teach medical students is perhaps foremost among these, and NEOMYCIN continues to focus on methods for augmenting clinical knowledge in order to facilitate its use in a tutorial setting. A particularly important aspect of this work is the insight that has been gained regarding the need to structure knowledge differently, and in more detail, when it is being used for different purposes (e.g., teaching as opposed to clinical decision making). It was this aspect of the GUIDON research that led to the development of NEOMYCIN, which is an evolving computational model of medical diagnostic reasoning that we hope will enable us to better understand and teach diagnosis to students. An important additional realization is that these structuring methods are beneficial for improving the problem-solving performance of consultation programs, providing more detailed and abstract explanations to consultation users, and making knowledge bases easier to maintain.

As we move from technological development of explanation and student modeling capabilities, we will in the next year begin to collaborate more closely with the medical community to design an effective, useful tutoring program. Stanford Medical School faculty, such as Dr. Maffly, have shown considerable interest in this project. A research fellow associated with Maffly, Curt Kapsner, MD, joined the project last year to serve as medical expert and liaison with medical students at Stanford.

C. Highlights of Research Progress

C.1 Accomplishments This Past Year

C.1.1 The NEOMYCIN Consultation Program

NEOMYCIN is distinguished from other AI consultation programs by its uses of an explicit set of domain-independent meta-rules for controlling all reasoning. These rules constitute the diagnostic procedure that we want to teach to students: the stages of diagnosis, how to focus on new hypotheses, and how to evaluate hypotheses. It has been a major undertaking, separate from the problem of representing disease knowledge, to design and test this diagnostic procedure. Such modifications require changing our conception of how disease knowledge is organized. For example, this year we partitioned disease findings into "non-specific" and "red flag" (those requiring explanation), augmenting the diagnostic procedure to use this information for focusing on hypotheses. A second change is to have the program reason about the disease process more generally. By associating symptoms by organ system, NEOMYCIN now has primitive means to infer when a disease process began. It also makes more complete use of severity, location, and progression information to discriminate among hypotheses.

During this past year, we completely reworked the program's knowledge of non-meningitis cases. This is important if we wish to teach students to consider the competitors of meningitis and how to discriminate among them. The goal is to prepare the program for presenting these (or similar) cases to students. In order to test the modeling component, it is necessary to ensure that the program has sufficient expertise to recognize good student behavior. All data that might be relevant to solving a given problem must be known to the program. The key problem here is establishing a base of synonyms and knowledge about classes of data. To do this, we have been collecting protocols of students solving problems, requiring them to request all by simple initial case information. Student behavior also suggests disease knowledge that must be added to the knowledge base that an expert might not consider, but which the modeling program must recognize in a student. In general, we find that students carry out a much broader, inefficient search, requesting much more information than an expert and drawing fewer conclusions from the information that they receive.

The Image Student Modeling Program -- Teaching diagnosis involves recognizing the intent behind a student's behavior, so that missing knowledge can be distinguished from inappropriate strategies. The teacher ***interprets*** behavior, ***critiques*** it, and provides *advice* about other approaches. To do this successfully and efficiently in a complex domain, the teacher benefits from multiple, complementary modeling strategies. IMAGE is a student modeling program that uses NEOMYCIN's meta-rules and disease knowledge to understand student diagnostic plans.

A student is presented with a problem to diagnose. As the student requests more problem data (i.e., takes a history and physical of the patient), IMAGE looks for regularities in sequences of his data requests. IMAGE contains a body of knowledge about how to map such sequences of behavior onto a strategical interpretation of what the student is doing. The process is heuristic in nature because the program will sometimes lose track of what the student is doing, because he is being inconsistent or using unexpected strategies.

The IMAGE uses a dual search strategy. The program first produces multiple predictions of student behavior by a model-driven simulation of NEOMYCIN. Focused, data-driven searches then explain incongruities. By supplementing each other, these methods lead to an efficient and robust plan understander.

A model of student strategies in medical diagnosis must disambiguate the possible purposes and knowledge underlying the student's actions. The approaches followed by other plan recognizers and student modelers are not sufficient here because:

1. the complex domain makes thorough searches impractical, whether top-down or bottom-up;
2. we are not modeling only facts and rules used in isolation, but also the procedures for applying them;
3. every one of the student's actions must be monitored in case the teaching module decides to interrupt;
4. his behavior must be evaluated and not just explained; and
5. we might not have any explicit goal statements from the student, so we expect to rely only on his queries for problem data as evidence for his thinking.

The IMAGE program is a prototype system which is now being extended. Specifically, a more useful system would examine its own interpretations and strive for coherence. We are designing such a such a system now, using the "blackboard model" for posting interpretations that may change over time. The levels of this blackboard are: 1) the student's data requests, 2) a classification of question type (e.g., triggered, follow-up, hypothesis-directed, general), 3) a strategic interpretation in terms of NEOMYCIN's diagnostic procedure (tasks and meta-rules). By incorporating a strategic level of interpretation, this program can be expected to make significant contributions to our understanding and use of the blackboard model of interpretation. The first version of this program will seek to explain student behavior in terms of deletion and reordering of procedural knowledge, plus simple variations of disease knowledge (e.g., false data/hypothesis relations). Study of student protocols is now suggesting what kinds of variations are common that we might easily identify automatically.

C.1.2 The NEOMYCIN Explanation System

The initial explanation system of NEOMYCIN, now completed, enables the user to answer WHY and HOW questions during a consultation. That is, when the program prompts the user for new data, the user may ask WHY the data is being requested or HOW some strategic task will be (or was) accomplished. Unlike MYCIN's explanation system, upon which this kind of capability is patterned, explanations in NEOMYCIN are in terms of the diagnostic plan, not just specific associations between data and diagnoses. The program can provide abstract and concrete paraphrases of strategy rules (based on canned text). We have begun the next phase, which is to answer WHY questions by condensing the entire line of reasoning. The program will use models of the user's disease and strategic knowledge, plus general explanation heuristics, to select the task and focus information that is most likely to be of interest. Prototypic user models are now implemented. Heuristics have been designed and include: 1) mentioning the last task whose focus (or argument) changed in kind (e.g., from a disease hypothesis to a finding request); 2) never mentioning tasks that are merely iterating over a list of rules, findings, or hypotheses; and 3) only mentioning tasks with a rule as an argument to programmers.

Related to our explanation condensations is an effort to teach the strategic language of tasks to students. For example, we will have students annotate a NEOMYCIN typescript in terms of tasks and foci, to help them recognize good strategic behavior. This requires a common language of what the tasks are, e.g., "grouping" and "asking general question." Rather than just marking annotating tasks, we seek the ***principles*** by which the tasks could be consistently structured into primitives and auxiliary. These same principles could be used by the explanation system for choosing tasks to mention. Our current theory is that these primitive or "interesting" operations correspond to meta-rules that establish a new focus.

C.1.4 Graphics for Teaching

We are continuing make extensive use of graphics in our programs. For example, we are implementing a program that will mostly automatize the protocol collection process (though we are cautious about how menus will bias student behavior, even when lists are very long and full of irrelevant findings). As part of our series of instructional programs, GUIDON-WATCH is now being implemented as a graphic system for watching NEOMYCIN's reasoning. For example, we can highlight the hypotheses under consideration and show graphically how the program "looks up" its hierarchies before refining hypotheses. The design of GUIDON-ANNOTATE is also mostly complete. It will allow a student to mark up a typescript of NEOMYCIN's behavior using the same language of tasks the program uses when explaining its own behavior; iconic menus are very useful to avoid natural language difficulties (though it is clear that the student will sometimes need to "talk back").

C.2 Research in Progress

The following projects are active as of June 1983 (see also near-term plans listed in Section III.A):

1. augmenting NEOMYCIN's disease knowledge so we can fairly evaluate the program's focussing strategies and evaluate IMAGE;
2. developing capability to automatically produce summary explanations of NEOMYCIN's reasoning.
3. development of GUIDON-WATCH and GUIDON-ANNOTATE for teaching NEOMYCIN's knowledge to students.
4. developing new student modeling program based on the blackboard model.

D. Publications Since January 1983

1. Hasling, D., Clancey, W.J., Rennels, G.: ***Strategic explanations in Consultation.*** Int J Man-Machine Studies, in press.
2. Clancey, W.J.: ***The advantages of abstract control knowledge in expert system design.*** Proceedings of AAAI-83, pages 74-78.
3. Clancey, W.J.: ***Acquiring, representing, and evaluating a competence model of diagnosis.*** In Chi, Glaser, and Farr (Eds.), THE NATURE OF EXPERTISE. In preparation. HPP-84-2.
4. Clancey, W.J. and E. H. Shortliffe.:*READINGS IN MEDICAL ARTIFICIAL INTELLIGENCE: THE FIRST DECADE.* Reading: Addison-Wesley, in press.
5. Clancey, W.J.: ***Classification Problem Solving.*** HPP-84-7. Submitted to AAAI-84.

E. Funding Support

Contract Title: "Exploration of Tutoring and Problem-Solving Strategies"
Principal Investigator: Bruce G. Buchanan, Adjunct Prof. Computer Science
Associate Investigator: William J. Clancey, Research Assoc. Computer Science
Agency: Office of Naval Research and
 Army Research Institute (joint)
ID number: N00014-79-C-0302
Term: March 1979 to March 1985
Total award: $683,892

II. INTERACTIONS WITH THE SUMEX-AIM RESOURCE

A. Medical Collaborations and Program Dissemination via SUMEX

A great deal of interest in GUIDON and NEOMYCIN has been shown by the medical and computer science communities. We are frequently asked to demonstrate these programs to Stanford visitors or at meetings in this country or abroad. GUIDON is available on the SUMEX 2020. Physicians have generally been enthusiastic about these programs' potential and what they reveal about current approaches to computer-based medical decision making.

Perhaps our most significant project to disseminate our research via SUMEX in the past year has been the completion of a book, "Readings in Medical Artificial Intelligence: The First Decade," edited by Dr. Clancey and Dr. Shortliffe. All of the significant SUMEX-AIM products of the past decade are described in this collection. Each chapter is preceded by a one-age historical introduction. In addition, opening and closing chapters by the editors survey issues in the field and the promise of the future. A complete index should make the book of considerable educational value. Preparation of this volume has been greatly aided by use of editing and formatting programs available on SUMEX-AIM. Royalties for the book, beyond production costs, will be used to sponsor an invited lecture at a major AI national conference, such as AAAI.

As mentioned earlier, a physician joined our group this year to help us develop the disease knowledge of the program (our first collaborator, Tim Beckett, MD, died of cancer in July 1983). This physician has found the convenience of accessing SUMEX from his laboratory or at home to be extremely important for finding time to test NEOMYCIN and to communicate with us by electronic mail.

B. Sharing and Interaction with Other SUMEX-AIM Projects

GUIDON/NEOMYCIN retains strong contact with the ONCOCIN project, as both are siblings of the MYCIN parent. These projects regularly share programming expertise and continue to jointly maintain large utility modules developed for MYCIN. In addition, the central SUMEX development group acts as an important clearing house for solving problems and distributing new methods.

C. Critique of Resource Management

In the winter of 1984, the SUMEX staff efficiently and effectively shifted our operation away from the center of campus to a professional office building adjoining the medical center. The placement and installation of LISP workstations proceeded smoothly. After a year with Ed Pattermann as director of SUMEX, we can report that the stability and excellence of the resource we have come to expect has been completely maintained. Very important to us, the RAVEN laser printer installed at our new site not only provides excellent-quality output, but as a machine devoted to the Heuristic Programming Project has eliminated the delays we were experiencing a year ago.

With the shift to personal machines, we are continuing to experience a few difficulties. The greatest problem appears to be inadequately debugged software from XEROX. In particular, Interlisp-D relies heavily on network capabilities and must be compatible with several operating systems. This transition to new kinds of hardware and software can be expected to continue for several years. Therefore, we are extremely reliant upon the availability of experienced systems support. We believe that additional SUMEX staff is necessary to accommodate growing community needs.

III. RESEARCH PLANS

A. Project Goals and Plans

Research over the next year will continue on several fronts, leading to several prototype instructional programs by early 1985.

1. Continue to develop the knowledge base so the program can understand and anticipate any reasonable approach to the cases chosen for teaching.

2. Test student modeling program on these cases, collecting data for further development of the program, as well as exploring about the range of student approaches to diagnosis.

3. Extend the explanation system to do full summaries. Incorporate modeling capabilities that relate inquiries to a user model. Provide explanations tailored to this interpretation of the motivation behind the user's inquiry.

4. Integrate current display capabilities into running NEOMYCIN consultation to show how the space of diagnoses is explored and how diagnostic tasks are generated. Develop these capabilities to explore forms of graphic explanation useful in tutoring. (GUIDON-WATCH)

5. Extend student modeling system to include heuristics for generating tests that will confirm and extend the model. Improve the model to include analysis of patterns in model interpretations, including dependency-directed "backtracking" in the belief system and some capability to critique the modeling rules. Relate this to knowledge acquisition research.

6. Work closely with medical students to package NEOMYCIN capabilities in a "workstation" for learning medical diagnosis, determining what mix of student and program initiative is desirable.

B. Long term plans: the GUIDON2 Family of Instructional Programs

We sketch here our general conception of the research we plan for 1984-87, specifically the construction of instructional systems that use NEOMYCIN. Our ideas are strongly based on recent proposals by JS Brown, particularly his paper "Process versus Product --A perspective on tools for communal and informal electronic learning" and some related papers that he wrote in 1983. The plan is to implement at least three of these programs (here called GUIDON-WATCH, GUIDON-MANAGE, and GUIDON-ANNOTATE).

The key idea is that NEOMYCIN provides a ***language*** by which a program can converse with a student about strategies and knowledge organization for diagnosis. NEOMYCIN's tasks and structural terms provide the vocabulary or parts of speech; the meta-rules are the grammar of the diagnostic process. We will construct different graphic, reactive environments in which the student can observe, describe, compare, and improve diagnostic behavior of himself and others. There are many shared, underlying capabilities that will be constructed in parallel and improved over time.

Our approach is to delineate clearly different kinds of interactions that a student might have with a program concerning diagnostic strategies. Thus, each instructional system (but one) has a name of the form GUIDON-<student activity>, where the name specifies what the student is doing (e.g., watching, telling). The programs can be made arbitrarily complex by integrating coaches, student models, and explanation systems. We try here to separate out these capabilities, trying to get at the minimum interesting activities we might provide for a student.

GUIDON-WATCH The simplest system allows a student to watch NEOMYCIN solve a problem, perhaps one supplied by the student. Graphics display the evolving search space, that is, how tasks, as operators, affect the differential (Differential ---(Question X)---> Differential'). The student can step through slowly and replay the interaction. He can ask for prosaic explanations and summaries of what the program is doing. The program will also indicate its task and focus for each data request. This introduces the student to the idea that the diagnostic process has structure and follows a certain kind of logic.

GUIDON-MANAGE In this system the student solves a problem by telling NEOMYCIN what task to do at each step. Essentially, the student provides the strategy and the program supplies the tactics (meta-rules) and domain knowledge to carry out the strategy. The program will in general carry through tasks in a logical way, for example, proceeding to test a hypothesis completely, and not "breaking" on FINDOUT or APPLYRULES (two low-level tasks that mainly test domain knowledge and not strategy). The program will not pursue new hypotheses automatically. However, the student will always see what questions a task caused the program to request, as well as how the differential changes. This activity leads the student to observe the entailments of strategies, helping him become a better observer of his own behavior. Here he shows that he knows the structural vocabulary that makes a strategy appropriate.

GUIDON-ANNOTATE This system allows the student to annotate a NEOMYCIN typescript, indicating the task and focus associated with each data request. The program will indicate, upon request, where the student is incorrect and which annotations are different from NEOMYCIN's, but still reasonable interpretations. The student will be able to choose these tasks from a menu of icons, either linearly or hierarchically displayed, as he prefers. (Again, NEOMYCIN will annotate its own solutions upon request and allow replaying.) This activity gets the student to think strategically by recognizing a good strategy. In this way, he learns to recognize how strategies affect the problem space.

GUIDON-APPRENTICE This is a variant of NEOMYCIN in which the program stops during a consultation and asks the student to propose the next data request(s). The student is asked to indicate the task and focus he has in mind, plus the differential he is operating upon. The program compares this proposal to what NEOMYCIN would do. In this activity we descend to the domain level and require the student to instantiate a strategy appropriately.

GUIDON-DEBUG Here the student is presented with a buggy version of NEOMYCIN and must debug it. He goes through the steps of annotating the buggy consultation session, indicating what questions are out of order or unnecessary, indicating what tasks are not being invoked properly, and then trying out his hypothesis on a "repaired" system. He is asked to predict what will be different, then allowed to observe what happens. This activity teaches the student to recognize how a diagnostic solution can be non-optimal, further emphasizing the value of good strategy. It also provides him with key meta-cognitive practice for criticizing and debugging problem behavior.

GUIDON-SOLVE This is the complete tutorial system. The student carries through diagnosis completely, while a plan recognizer attempts to track what he is doing and a coach interrupts to offer advice. Here annotation, comparison, debugging, and explanation are all integrated to illustrate to the student how his solution is non-optimal. For example, the student might be asked to annotate his solution after he is done; this will point out strategic gaps in his awareness and provide a basis for critique and improvement. A "curriculum" based on frequent student faults and important things to learn will drive the interaction. In this activity, the student is on his own. Faced with the proverbial "blank screen," he must exercise his diagnostic procedure from start to finish.

GUIDON-GAME Two or more students play this together on a single machine. They are given a case to solve together, and each student requests data in turn. All students receive the requested information. When a student is ready, he makes a diagnosis, indicated secretly to the program while the others are not watching. He then drops out of the questioning sequence. However, he can re-enter later, but of course will be penalized. Afterwards, score is based on the number of questions asked and use of good strategy. The coach will indicate to weak players what they could learn from strong players, encouraging them to discuss certain issues among themselves. Variation: one person solves while one or more competing students annotate the solution and show where it could be improved. Variation: one team introduces a bug into NEOMYCIN (and predicts the effect) and the other team finds it (as in SOPHIE). This activity will encourage students to share their experiences and talk to and learn from each other about the diagnostic process.

C. Requirements for Continued SUMEX Use

Although most of the GUIDON and NEOMYCIN work is shifting to Xerox Dolphins and Dandelions (D-machines), the DEC 2060 and 2020 continue to be key elements in our research plan. Our primary use of the 2060 will be to develop the NEOMYCIN consultation system, possibly by remote ARPANET access. Because of address space limitations, the consultation program can be combined with explanation or student modeling facilities, but not both, as is required for GUIDON2 programs. We continue to use the 2020 for demonstrating the original GUIDON program. As always, the 2060 will be essential for work at home, writing, and electronic mail.

D. Requirements for Additional Computing Resources

The D-machine's large address space is permitting development of the large program that complex computer-aided instruction requires. Graphics will enable us to develop new methods for presenting material to naive users. We also plan to use the D-machine as a reliable, constant "load-average" machine, for running experiments with physicians and students. The development of GUIDON2 on the D-machine will demonstrate the feasibility of running intelligent consultation or tutoring systems on small, affordable machines in physicians' offices, schools and other remote sites.

We currently have access to 1 1/2 DOLPHINs. We expect that 3 full time programmers will need access to two full machines. We are keeping logs so we can begin to understand patterns of activity and how these "personal" machines can be effectively shared.

E. Recommendations for Future Community and Resource Development

As we shift our development of systems to personal LISP machines, such as the DOLPHIN, it becomes more difficult to access these programs remotely for access from our homes (so that we may work conveniently during the evenings and weekends) and from remote sites for collaboration and demonstration. This problem will be partly ameliorated by "dial-up" (modem) access to these machines, but the use of bit-mapped displays requiring a high-bandwidth makes the phone lines inadequate for our purposes. Further technological development of networks, probably involving access over cables, will be necessary.

As computer resources become more distributed, the need for a central machine does not diminish. Programs and knowledge bases continue to be shared, requiring high-speed network connections among computers and file servers. SUMEX-AIM's role will shiftly slightly over the next few years to accommodate these needs, but its identity as a central resource will only change in kind, not importance. Moreover, sophisticated printing devices, such as the Xerox RAVEN, must necessarily be shared, again using a network. Maintenance of this network and its shared devices will become a key activity for the SUMEX staff. Thus, while computing resources will be provided by the "outboard engines" of personal machines, the community will remain intricately linked and dependent on common, but peripheral, resources.

From this perspective, future resource development should focus on improving the capabilities of networks, file servers, and attached devices to respond to individual requests. For example, it is now common for 10% of a user's time at a personal machine to be spent waiting for a file server or printer to process a request. Multi-processing becomes a necessity in such an environment, so a request can be honored, while the user returns to continue his programming or editing.

II.A.1.3. HPP Core AI Research

Heuristic Programming Project
Principal Investigator: Edward A. Feigenbaum
Co-Principal Investigator: Bruce G. Buchanan
Department of Computer Science
Stanford University

I. SUMMARY OF RESEARCH PROGRAM

A./B. Rationale and Medical Relevance

Medicine and the biological sciences are knowledge-intensive with an exponential rate of growth in relevant knowledge. This means that problem solving of all sorts is becoming increasingly complex in these disciplines. Further, most problems are symbolic in nature rather than amenable to mathematical formulation and numerical solution. Artificial Intelligence (AI) methods have been focused on medical and biological problems for over a decade with considerable success. This is because, of all the computing methods known, AI methods are the only ones that deal explicitly with symbolic information and problem solving and with knowledge that is heuristic (experiential) as well as factual.

One particularly fast-moving area of AI is expert systems. An expert system is one whose performance level rivals that of an human expert because it has extensive domain knowledge (currently usually derived from an human expert); it can reason about its knowledge to solve difficult problems in the domain; it can explain its line of reasoning much as an human expert can; and it is flexible enough to incorporate new knowledge without reprogramming. Expert Systems draw on the current stock of ideas in AI, for example, about representing and using knowledge. They are adequate for capturing problem-solving expertise for many bounded problem areas. Numerous high-performance, expert systems have resulted from this work in such diverse fields as analytical chemistry, medical diagnosis, cancer chemotherapy management, VLSI design, machine fault diagnosis, and molecular biology. Some of these programs rival human experts in solving problems in particular domains and some are being adapted for commercial use. Other projects have developed generalized software tools for representing and utilizing knowledge (e.g., EMYCIN, UNITS, AGE, MRS, GLISP) as well as comprehensive publications such as the three-volume ***Handbook of Artificial Intelligence*** and books summarizing lessons learned in the DENDRAL and MYCIN research projects.

But the current ideas fall short in many ways, necessitating extensive further basic research efforts. Our core research goals, as outlined in the next section, are to analyze the limitations of current techniques and to investigate the nature of methods for overcoming them. Long-term success of computer-based aids in medicine and biology depend on improving the programming methods available for representing and using domain knowledge. That knowledge is inherently complex ¨ it contains mixtures of symbolic and numeric facts and relations, many of them uncertain; it contains knowledge at different levels of abstraction and in seemingly inconsistent frameworks; and it links examples and exception clauses with rules of thumb as well as with theoretical principles. Current techniques have been successful only insofar as they severely limit this complexity. As the applications become more far-reaching, computer programs will have to deal more effectively with richer expressions and much more voluminous amounts of knowledge.

This report documents progress on the basic or core research activities within the

Heuristic Programming Project (HPP), funded in part under the SUMEX resource as well as by other federal and industrial sources. This work explores a broad range of basic research ideas in many application settings, all of which contributes in the long term to improved knowledge based systems in biomedicine.

C. Highlights of Research Progress

In the last year, we made progress on several major topics of research. The style of research that we believe is most productive at this stage of development of AI is the experimental style. Thus, within the HPP we build systems that implement our ideas for answering (or shedding some light on) fundamental questions; we experiment with those systems to determine the strengths and limits of the ideas; we redesign and test more; we attempt to generalize the ideas from the domain of implementation to other domains; and we publish details of the experiments. In order to carry out this style of research, then, we select specific problems to help focus the general questions. Many of these specific problem domains are medical or biological. In this way we believe the HPP has made substantial contributions to core research problems of interest not just to the AIM community but to AI in general.

Progress is reported below under each of the major topics of our work. Citations are to HPP technical reports listed in the publications section.

1. *Knowledge representation*: How can the knowledge necessary for complex problem solving be represented for its most effective use in automatic inference processes? Often, the knowledge obtained from experts is heuristic knowledge, gained from many years of experience. How can this knowledge, with its inherent vagueness and uncertainty, be represented and applied?

   Work on the logic-based MRS and the rule-based NEOMYCIN systems continues, attracting wide interest within the AI community. Numerous copies of MRS have been sent to collaborators elsewhere who are experimenting with it on the own machines. The book on rule-based expert systems by Buchanan & Shortliffe was completed in this year.

   [See HPP technical memos HPP-83-26, HPP-83-28, HPP-83-29, HPP-83-34, HPP-84-1]

2. *Advanced architectures and Control*: What kinds of software tools and system architectures can be constructed to make it easier to implement expert programs with increasing complexity and high performance? How can we design flexible control structures for powerful problem solving programs?

   A major effort in exploring and understanding the Blackboard architecture has been undertaken. A new pilot project using this architecture was started in the domain of protein chemistry (see description of Jardetzky & Buchanan pilot project). We have also begun investigating Blackboard systems as a way of organizing expert systems to exploit concurrency. Initial work has begun using the HASP/AGE systems as an application example.

   [See HPP technical memos HPP-83-30, HPP-83-33, HPP-83-38, HPP-83-43, HPP-83-44, HPP-84-4, HPP-84-6]

3. *Knowledge acquisition*: How is knowledge acquired most efficiently from human experts, from observed data, from experience, and from discovery? How can a program discover inconsistencies and incompleteness in its knowledge base? How can the knowledge base be augmented without perturbing the established knowledge base?

We have continued to make progress on two on-going projects for learning by experience and learning by analogy, and have initiated work on three new systems for acquiring knowledge. Those three are learning by watching, learning from text, and learning rules & meta-rules inductively. All three of the new systems use medical problems as their test-domains.

[Preliminary results have been published in HPP-83-27, HPP-83-36, HPP-84-2, HPP-84-8.]

4. *Knowledge utilization*: By what inference methods can many sources of knowledge of diverse types be made to contribute jointly and efficiently toward solutions? How can knowledge be used intelligently, especially in systems with large knowledge bases, so that it is applied in an appropriate manner at the appropriate time?

   These issues are being explored in the development of MRS (Meta-Representation System) where one of the roles of meta-knowledge is to guide the effective use of lower level knowledge. They are also central in the studies of Blackboard control systems and their use in concurrent expert systems.

   [See HPP technical memos HPP-83-26, HPP-83-28, HPP-83-30, HPP-83-33, HPP-83-38, HPP-84-1, HPP-84-2, HPP-84-6]

5. *Software Tools*: How can specific programs that solve specific problems be generalized to more widely useful tools to aid in the development of other programs of the same class?

   We have continued the development of new software tools for expert system construction and the distribution of packages that are reliable enough and documented so that other laboratories can use them. These include the old rule-based EMYCIN system, MRS, and AGE.

   [See HPP technical memos HPP-83-26, HPP-83-28, HPP-83-29, HPP-83-33]

6. *Explanation and Tutoring*: How can the knowledge base and the line of reasoning used in solving a particular problem be explained to users? What constitutes a sufficient or an acceptable explanation for different classes of users? How can knowledge in a system be transferred effectively to students and trainees?

   The NEOMYCIN program has undergone preliminary comparison with medical students' protocols to understand the extent to which its medical concepts match those of the students. Analysis of experts' problem solving has also been done. NEOMYCIN's explanation capabilities have been improved. New work on student modelling has started in order to test NEOMYCIN in the context of tutoring.

   [See HPP technical memos HPP-83-41, HPP-83-42, HPP-84-2, HPP-84-7]

7. *Planning and Design*: What are reasonable and effective methods for planning and design? How can symbolic knowledge be coupled with numerical constraints? How are constraints propagated in design problems?

   The Palladio system for assisting in the design of VLSI circuits has been demonstrated and results presented in major publications and conferences.

   [See HPP technical memos HPP-83-31, HPP-83-39, HPP-83-45, HPP-83-46, HPP-83-47, HPP-84-3, HPP-84-5]

8. *Diagnosis*: How can we build a diagnostic system that reflects any of several

diagnostic strategies? How can we use knowledge at different levels of abstraction in the diagnostic process?

Research on using causal models in a medical decision support system (NESTOR) was largely completed and will be published in the coming year. A second medical diagnosis program that uses causal models of renal physiology (AI/MM) was also substantially completed and will be published soon. We are investigating the process of diagnosis in electronics as well as in medicine. The major thrust of this work has been integrating causal models about, and the structure of, a computer system or systems of the human body.

[See HPP technical reports: HPP-83-32, HPP-83-37, HPP-83-40, 84-7]

D. Relevant Publications

HPP-83-26 Michael R. Genesereth, *"MRS Casebook"*, May 1983.

HPP-83-27 Thomas D. Dietterich and Ryszard S. Michalski, *"Discovering Patterns in Sequences of Objects"*, May 1983.

HPP-83-28 Michael R. Genesereth, *"A Meta-level Representation System"*, May 1983.

HPP-83-29 M. Grinberg, *"MRS Installation Instructions"*, May 1983. This report available only to those who have purchased the software system MRS.

HPP-83-30 Barbara Hayes-Roth, *"The Blackboard Architecture: A General Framework for Problem Solving?"* May 1983.

HPP-83-31 Harold Brown, Christopher Tong, Gordon Foyster, *"Palladio: An Exploratory Environment for IC Design"*, June 1983.

HPP-83-32 John Kunz, E.A.Feigenbaum, Bruce G. Buchanan, E.H. Shortliffe, *"Comparison of Techniques of Computer-Assisted Decision Making in Medicine"*. Submitted for publication in the **Pure and Applied Biostructure**. World Press, Singapore (1983).

HPP-83-33 Nelleke Aiello, *"A Comparative Study of Control Strategies for Expert Systems: AGE Implementation of Three Variations of PUFF"*, June 1983.

HPP-83-34 Jock Mackinlay, *"Intelligent Presentation: The Generation Problem for User Interfaces"*, March 1983.

HPP-83-36 Russell Greiner and Michael R. Genesereth, *"What's New? A Semantic Definition of Novelty"*, June 1983.

HPP-83-37 Robert Joyce, *"Reasoning About Time-dependent Behavior in a System for Diagnosing Digital Hardware Faults"*, August 1983.

HPP-83-38 Barbara Hayes-Roth, *"The Blackboard Model of Control"*, June 1983.

HPP-83-39 Jerry Yan, Gordon Foyster, Harold Brown, *"An Expert System for Assigning Mask Levels to Interconnect in Integrated Circuits"*, October 1983.

HPP-83-40 Benoit Mulsant and David Servan-Schreiber, *"Knowledge Engineering: A Daily Activity on a Hospital Ward"*, October, 1983.

HPP-83-41 (working paper) Diane Warner Hasling, *"Strategic Explanations for a Diagnostic Consultation System"*, in AAAI Proceedings 1983 pp. 157-161.

HPP-83-42 Wm. J. Clancey, *"GUIDON"*, November 1983.

HPP-83-43 Narinder Singh, *"MARS: A Multiple Abstraction Rule-Based System"*, December 1983.

HPP-83-44 H.Penny Nii, *"Signal-to-Symbol Transformation: Reasoning in the HASP/SIAP Program"*, December 1983.

HPP-83-45 (working paper) Christopher Tong, *"A Framework for Circuit Design"*, December 1983.

HPP-83-46 (working paper) J.J. Finger, Michael Genesereth, *"RESIDUE - A Deductive Approach to Design"*, December 1983.

HPP-83-47 (working paper) J.J. Finger, Michael Genesereth, *"Planning to Gather Information"*, December 1983.

HPP-84-1 Michael R. Genesereth, *"Partial Programs"*, January 1984. (Replaces HPP-81-6)

HPP-84-2 (working paper) Wm. J. Clancey, *I"Acquiring, Representing, and Evaluating a Competence Model of Diagnostic Strategy"*, February 1984.

HPP-84-3 (working paper) Gordon Foyster, *"A Knowledge-Based Approach to Transistor Sizing"*, March 1984.

HPP-84-4 (working paper) Jock Mackinlay, Michael R. Genesereth, *"Implicit Language "*,March 1984.

HPP-84-5 Jeffrey Rosenschein, Michael R. Genesereth, *"Communication and Cooperation"*, March 1984.

HPP-84-6 D.E. Smith, Michael R. Genesereth, *"Controlling Recursive Inferences"*, March 1984.

HPP-84-7 William J. Clancey, *"Classification Problem Solving"*, March 1984.

HPP-84-8 (author), *"The Role of Abstractions in Understanding Analogy"*, April 1984.

E. Funding Support

We are pursuing a broad core research program on basic AI research issues with support from not only SUMEX but also DARPA, NASA, NSF, and ONR. SUMEX provides some salary support for staff and students involved in core research and invaluable computing support for most of these efforts. Additional salary support comes from the sources listed below.

Agency: National Library of Medicine; 5 P01 LM 03395
Project Title: Biomedical Knowledge Representation

Principal Investigator: Edward A. Feigenbaum
Amount: $95,424 (Direct Costs only)
Period Covered: 7/1/83 - 6/30/84

Agency: Defense Advanced Research Projects Agency; N00039-83-C-0136
Project Title: Heuristic Programming Project
Principal Investigators: Edward A. Feigenbaum and Bruce G. Buchanan
Amount: $3,354,493
Period Covered: 10/1/82 - 9/30/85

Agency: Defense Advanced Research Projects Agency; N00014-81-K-0303
Project Title: Intelligent Agents
Principal Investigator: Edward A. Feigenbaum
Award Amount: $484,652
Period Covered: 3/1/81 - 2/28/84
(the follow-on is merged with N00039-83-C-0136)

Agency: Defense Advanced Research Projects Agency/Martin Marietta; (pending)
Project Title: Intelligent Task Automation
Principal Investigators: Michael R. Genesereth
Amount: $297,626
Period Covered: 10/1/83 - 2/28/85

Agency: Office of Naval Research; N00014-79-C-0302
Project Title: Recognizing and Articulating Diagnostic Skills in an Intelligent Tutoring System
Principal Investigator: Bruce G. Buchanan
Award Amount: $1,110,447
Period Covered: 3/15/79 - 3/14/85

Agency: Office of Naval Research; N00014-80-C-0609
Project Title: Automatic Induction of Strategic Rules
Principal Investigator: Douglas B. Lenat
Award Amount: $108,000
Period Covered: 6/1/82 - 5/31/84

Agency: Office of Naval Research; N00014-81-K-0004
Project Title: Research on Introspective Systems
Principal Investigator: Michael R. Genesereth and Edward H. Shortliffe
Award Amount: $511,748
Period Covered: 1/1/84 - 12/31/86

Agency: NASA Goddard Space Flight Center; NAG 5-261
Project Title: Planning in Uncertain and Unforgiving Situations
Principal Investigators: Bruce G. Buchanan (and Thomas O. Binford)
Award Amount: $55,029
Period Covered: 9/1/83 - 8/31/84

Agency: NASA-AMES Research Center; NCC 2-220
Project Title: Research on Advanced Knowledge-based System Architectures
Principal Investigator: Edward A. Feigenbaum
Award Amount: $90,000
Period Covered: 1/1/84 - 11/30/84 (support level pending for future years)

Agency: NASA-AMES Research Center; NCC 2-274
Project Title: Research on Knowledge Representation
Principal Investigator: Bruce G. Buchanan
Award Amount: $50,000
Period Covered: 10/1/83 - 12/31/84 (support level pending for future years)

Agency: National Science Foundation; IST-83-12148
Project Title: Information Structure and Use in Knowledge-Based Expert Systems
Principal Investigator: Bruce G. Buchanan and Edward H. Shortliffe
Award Amount: $330,138
Period Covered: 3/15/84 - 2/28/87

Agency: IBM; IBM/Stanford Joint Study
Project Title: The Use of Design Models in the Diagnosis of Computer Hardware
Principal Investigator: Edward A. Feigenbaum
Award Amount: $660,000
Period Covered: 10/1/82 - 9/30/85

II. INTERACTIONS WITH THE SUMEX-AIM RESOURCE

We rely on the central SUMEX facility as a focal point for all the research within the HPP, not only for much of our computing, but for communications and links to our many collaborators as well. As a common communications medium alone, it has significantly enhanced the nature of our work and the reach of our collaborations. As SUMEX and the HPP acquire a diversity of hardware, including LISP workstations machines and smaller personal computers, we rely more and more heavily on the SUMEX staff for integration of these new resources into the local network system. The staff has been extremely helpful and effective in dealing with the myriad of complex technical issues and leading us competently into this world of decentralized, diversified computing.

III. RESEARCH PLANS

A. Project Goals and Plans

The Core Research Project focuses on understanding the roles of knowledge in symbolic problem solving systems ¨ its representation in software and hardware, its use for inference, and its acquisition. We are continuing to develop new tools for system builders and to improve old ones. The research crosses a number of application domains, as reflected in the subprojects discussed earler, but the main issues that we are addressing in this research are those fundamental to all aspects of AI. We believe this core research is broadening and deepening the groundwork for the design and construction of even more capable and effective biomedical systems.

As mentioned above, although our style of research is largely empirical, the questions we are addressing are fundamental. The three major research issues in AI have, since its beginning, been knowledge representation, control of inference (search), and learning. Within these topics, we will be asking the following kinds of questions and as our work progresses, we hope to leave behind several prototype systems that can be developed by others in the medical community.

1. Knowledge Representation -- How can we represent causal models and structural information? What are the relative benefits of logic-based, rule-based, and frame-based systems? How can we represent temporal relations and events so that reasoning over time is efficient?
2. Knowledge Acquisition -- How can an expert system acquire new knowledge without consuming substantial time from experts? Can we improve the knowledge engineering paradigm enough to make a difference? Can automatic learning programs be designed that will work across many disciplines? Will cooperative man-machine systems be able to open the communication channel between expert and expert system?
3. Knowledge Utilization -- By what inference methods can a variety of sources of knowledge of diverse types be made to contribute jointly and efficiently toward solutions? What is the nature of strategy and control information?

Plans for the Coming Year -- Several systems have been developed in recent years to serve as vehicles for knowledge engineering and research on knowledge representation and its use. Knowledge acquisition (including machine learning) and advanced architectures for AI will be the two areas of most new activity in the coming year. Research on these topics obviously must draw on on-going work in representation and control.

In particular, we will focus on

- Inductive learning of MYCIN-like rules from case data in the domain of diagnosing disorders where the chief complaint is jaundice;
- Learning from experience in domains where the means for interpreting new data are largely contained in the emerging (and thus incomplete and not wholly correct) theory;
- Learning by watching a medical expert diagnose cases presented by NEOMYCIN;
- Investigating complex signal understanding systems for ways to exploit and represent concurrency with a view toward hardware and software architectures that may be capable of several orders of magnitude improvement in performance.

B. Justification and Requirements for SUMEX Use

Core research is essential to the vitality of a national resource for artificial intelligence applications in biomedicine. It provides the new ideas and tools to address the limitations of existing experimental systems. We believe that the technical reports and programs produced as part of our continuing scientific efforts are received with interest by the AIM, and larger AI, research communities.

We require a stable source of computing cycles and substantial file space for the myriad of sub-projects that make up HPP/SUMEX core research. We anticipate no special needs beyond those in evidence this past year.

C. Computing Resources Outside of SUMEX-AIM

For some of the research reported here, we use Xerox-1100 series Lisp workstations, some of which were purchased by the NIH for SUMEX use. We have also purchased additional computing resources for the community with DARPA and HPP gift funds, including a VAX 11/780, a VAX 11/750, a Symbolics LM-2, 4 Symbolics 3600's, a Xerox Dorado, 2 Xerox Dandelions, and overflow cycles on the SCORE 2060. We expect to purchase additional Lisp workstations with similar funding over the next year and a half.

II.A.1.4. MOLGEN Project

MOLGEN - Applications of Artificial Intelligence to Molecular Biology: Research in Theory Formation, Testing, and Modification

Prof. E. Feigenbaum and Dr. P. Friedland
Department of Computer Science
Stanford University

Prof. Charles Yanofsky
Department of Biology
Stanford University

I. SUMMARY OF RESEARCH PROGRAM

A. Project Rationale

The MOLGEN project has focused on research into the applications of symbolic computation and inference to the field of molecular biology. This has taken the specific form of systems which provide assistance to the experimental scientist in various tasks, the most important of which have been the design of complex experiment plans and the analysis of nucleic acid sequences. We are now moving into a new phase of research in which we explore the methodologies scientists use to modify, extend, and test theories of genetic regulation, and then emulate that process within a computational system.

Theory or model formation is a fundamental part of scientific research. Scientists both use and form such models dynamically. They are used to predict results (and therefore to suggest experiments to test the model) and also to explain experimental results. Models are extended and revised both as a result of logical conclusions from existing premises and as a result of new experimental evidence.

Theory formation is a difficult cognitive task, and one in which there is substantial scope for intelligent computational assistance. Our research is toward building a system which can form theories to explain experimental evidence, can interact with a scientist to help to suggest experiments to discriminate among competing hypotheses, and can then revise and extend the growing model based upon the results of the experiments.

The MOLGEN project has continuing computer science goals of exploring issues of knowledge representation, problem-solving, discovery, and planning within a real and complex domain. The project operates in a framework of collaboration between the Heuristic Programming Project (HPP) in the Computer Science Department and various domain experts in the departments of Biochemistry, Medicine, and Biology. It draws from the experience of several other projects in the HPP which deal with applications of artificial intelligence to medicine, organic chemistry, and engineering.

B. Medical Relevance and Collaboration

The field of molecular biology is nearing the point where the results of current research will have immediate and important application to the pharmaceutical and chemical industries. Already, clinical testing has begun with synthetic interferon and human growth hormone produced by recombinant DNA technology. Governmental reports estimate that there are more than 200 new and established industrial firms already undertaking product development using these new genetic tools.

The programs being developed in the MOLGEN project have already proven useful and important to a considerable number of molecular biologists. Currently several dozen researchers in various laboratories at Stanford (Prof. Paul Berg's, Prof. Stanley Cohen's, Prof. Laurence Kedes', Prof. Douglas Brutlag's, Prof. Henry Kaplan's, and Prof. Douglas Wallace's) and over 400 others throughout the country have used MOLGEN programs over the SUMEX-AIM facility. We have exported some of our programs to users outside the range of our computer network (University of Geneva [Switzerland], Imperial Cancer Research Fund [England], and European Molecular Biology Institute [Heidelberg] are examples). The pioneering work on SUMEX has led to the establishment of a separate NIH-supported facility, BIONET to serve the academic molecular biology research community with MOLGEN-like software.

C. Highlights of Research Progress

C.1 Accomplishments

The current year has seen the completion of the previous grant's research on experiment design and debugging and the beginning of our new work on theory formation. The highlights of this work are summarized in several categories below.

C.1.1 Cloning Experiment Design

The cloning advisory system is now operational. It utilizes the following basic strategy or skeletal plan for the design of all experiments: First, isolate the piece of DNA you wish to clone, second, select a vector to carry the clone, third, insert the DNA into the vector, fourth, select a host for expression of the hybrid molecule, fifth, insert the hybrid into the host, and sixth, select for the protein or nucleic acid product that was the eventual goal of the cloning experiment. Following this skeletal plan, the cloning knowledge base contains information on DNA isolation methods, cloning vectors, insertion methods, hosts, host insertion methods, and selection methods.

This knowledge base has been tested on a wide range of cloning experiments in various laboratories. Dr. Rene' Bach finished work on the knowledge base by concentrating on two areas: vector selection and simulation of biological operations. He researched and described the criteria needed to make expert choices among several dozen different DNA cloning vectors, viewing that choice as being the "key" decision in the skeletal plan that would constrain and motivate the other decisions. He also did extensive work on describing the procedural knowledge necessary to accurately model the changes to DNA structures that take place during the course of a cloning experiment. This modeling serves to make decision-making during plan refinement more accurate and is also an important part of the experiment debugging system described below.

C.1.2 Experiment Debugging Research SPEX (the name given to the current version of our skeletal planning system) keeps complete records of all decisions made during the course of designing an experiment. These include strategic decisions as to which general planning heuristics to employ and which domain-specific skeletal plans to use, as well as tactical decisions made in the course of choosing specific operators to instantiate a plan step. In addition, SPEX keeps a dynamic model of the world state as assumed after the execution of each plan step. During the last year, Mr. Armin Hakin made use of this comprehensive information to extend the SPEX system to include experiment debugging facilities.

Experiment designs fail for one of three major reasons: a technical mistake in the laboratory (added too much salt, stopped a reaction too soon, etc.), a knowledge base mistake in technique selection (for example, the wrong enzyme was chosen for a cutting step), or a strategic error--all of the steps work individually, but the design as a whole is in error. Our experiment debugging system has demonstrated an ability to cope well with errors of the first two types, and partially with errors of the final type.

The system works by first acquiring a description of the failed experiment and its goals from a scientist. This is done through a special experiment editing and description component that was added to the Unit System. The debugging system then queries the user to determine the skeletal plan that led to the creation of the particular experiment design; this step may involve the creation of a new skeletal plan (thereby serving as a useful aid to knowledge acquisition) or it may be that an existing skeletal plan will serve. If it is a new skeletal plan, then the system tries to find errors of the third type from above by utilizing some general skeletal plan design heuristics (e.g. making sure appropriate preconditions are established).

The system refines the skeletal plan given the goals and conditions of the experiment in question. It compares its choices with those actually selected by the scientist. When the debugging system's choices differ from those of the scientist, the system determines whether the difference indicated a fatal flaw in the scientist's plan or merely reflected different optimality criteria among nearly equal possibilities.

Finally, the system examines its model of what changes should occur in the laboratory environment during the course of the experiment. It informs the scientist when measurable changes should occur and asks him to compare those to actual changes. When a step is found whose "before" and "after" states do not correspond to predicted changes, then that step is pointed out as being suspect to a technical error of type 1 above.

C.1.3 Research in Theory Formation, Modification, and Testing

The first goal of our new work in scientific theory discovery was to extensively study an existing example of the process. Professor Charles Yanofsky's work in elucidating the structure and function of regulation in the trp operon of E. coli provided us with an excellent subject that spanned twelve years of research, dozens of collaborators, and almost one hundred research papers.

We have conducted extensive interviews with Professor Yanofsky and many of his former students and collaborators. We have examined most of the relevant research papers. We believe we now have a good understanding of the three major classes of knowledge that were important in the discovery of the theory of regulation in the trp operon: knowledge about the relevant biological objects, knowledge about the techniques used to elicit new information, and discovery heuristics used to build new models. The major stages in the discovery process have been mapped out, and work has begun on constructing a knowledge base that will represent the state of the world at the beginning of the trp operon research.