known connectivities within the protein structure (i.e., the amino acid sequence, cofactors, disulfide bridges, and coordination bonds to prosthetic groups). The microstructure knowledge base contains known facts about protein molecules (e.g., the molecular geometry of the peptide bond and the amino acid side chains, hydrogen bonding properties, helix forming propensity, etc.). The macrostructure knowledge base contains stereotype templates for the plausible major components of the molecule (e.g., alpha helices and pleated sheets). Our strategy is to isolate an individual molecule within the map, then determine the path of the main chain by a skeletonizing procedure (as is done, for example, by J. Greer, J. Mol. Biol.(1974), vol.82, pp. 279-301). We will then parameterize the density along the backbone, identify the most obvious regions (heavy atoms, alpha helices, planar groups) and determine, by region growing and template matching, the identity and position of side chains.

2. Structure Determination in the Absence of Experimental Phase
   Information

X-ray crystallography is the primary experimental technique for investigating the 3-D structure of molecules. The data so obtained are a collection of intensity measurements at discrete directions with respect to the x-ray beam and crystal axes. These intensities are related to the positions of the atoms in the crystal lattice by a Fourier transformation. Thus, given the structure of the molecule, its orientation with respect to the crystal axes and the symmetry properties which determine the molecular stacking, one can calculate the intensities. However, given the intensities and the unit cell properties, one cannot go the other way. That is, the inverse Fourier transform cannot be calculated because the experiment measures only the amplitudes of the diffracted waves and not their phases -- the classical "phase problem" of x-ray crystallography.

During our first year we have been investigating and implementing various computational techniques for inferring trial structures, in the absence of phase information. Our aim has been to develop a system of computer programs which apply as much general and task-specific knowledge as possible to constrain the search for a plausible structure (or partial structure) consistent with the experimentally measured structure amplitudes.

A procedure well known to x-ray crystallographers investigating the structures of small molecules is that of Patterson search. Recently this technique has been shown to be effective in protein crystallography as well, when there exists a family resemblance between the molecule under investigation and a known protein. Patterson search is basically an image-seeking technique, where one searches for the "Patterson image" of an hypothesized molecular structure in the Patterson map derived from the experimental data. The Patterson function does not require any knowledge of phases.

Patterson search is our primary technique for inferring structures in the absence of experimental phase information. In order to resolve the ambiguities which often arise in Patterson search predictions, we are investigating the use of other knowledge sources, among which are

anomalous dispersion Paterson interpretation, Patterson search in reciprocal space, superposition and Fourier refinement methods. The integration of these diverse knowledge sources is a primary objective of the research.

C. Summary of Project Accomplishments

Our activities during the first year fall into three general categories: augmenting our stockpile of crystallographic computing tools, applying Patterson search methods to a solved protein structure (Cytochrome C2) and an unsolved protein structure (Cytochrome F), and initiating a new research objective.

1. Application of Patterson Search to Cytochrome F

A major goal of our first year of research has been to apply the method of Patterson search, in conjunction with other analytic techniques, to solve a real protein structure. Cytochrome F is an excellent candidate, because (1) phase information is not yet available, and (2) the protein's structure is expected to show a resemblance to other members of the cytochrome family. The current hope is that the family resemblance will be sufficiently strong that the complete structure can be solved by standard refinement techniques after one finds the correct orientation and position of a characteristic substructure. As of this writing a complete solution has not yet been obtained. A considerable effort, described in the remainder of this section has been invested in the pursuit of the correct orientations of the protein in the unit cell. A large number of Patterson search calculations were performed, exploring the effects of variations in the search structures, the selection of search vectors, the choice of measures of fit, and even in the primary data. It now remains to be seen if some of the candidate orientations proposed by the search calculations can be verified by other sources of crystallographic knowledge.

2. Selection of a new research objective.

Shortly after the inception of the project, the two collaborating groups agreed to the need for an additional scientist with an extensive knowledge of crystallography and crystallographic computing, and a serious interest in the application of AI techniques to his field of expertise. We were fortunate to induce Dr. Carroll Johnson of Oak Ridge National Laboratories, who well fulfills these qualifications, to join our project for a one-year period beginning September 1, 1975. His contributions have been instrumental in defining a new task area for the application of AI methodology to protein crystallography. After studying recent work in visual scene analysis, he noted the similarities of that AI application with the crystallographic problem of interpreting a 3-dimensional electron density map, i.e. deriving the coordinates for a trial structure, given the electron density function, the amino-acid sequence and the stereochemical principles and constraints known to apply.

The task so defined contains most if not all of the ingredients for the development of a knowledge-based system, in the mainstream of current AI research. The crystallographer integrates several sources of knowledge -- chemical, stereochemical, crystallographic -- as he builds a model of the protein which is consistent with the given data. He combines this knowledge with a rich set of heuristics for focussing his attention on promising regions of the map, for distinguishing characteristic features, for deciding at what level of detail to stop the interpretation in different regions, and for evaluating competing hypotheses.

The model builder's decision-making process is dynamic and flexible, driven at times by the need to reach specific subgoals, and at other times by the current state of the model or special features of the data. A computer program for interpreting the map will require a control structure which combines both goal-driven and event-driven elements. The design of a suitable control structure, and the implementation of a prototype program for performing the basic interpretive tasks are primary objectives for our second year of research.

3. Assembly of crystallographic computing tools.

With the assistance of several crystallographers around the country we have augmented our collection of crystallographic computing programs and systems (i.e., integrated collections of programs). Those programs we received and/or implemented on SUMEX include:

a) X-RAY 72. This is a large system of Fortran programs developed by J.M. Stewart (Univ. of Md.) and others. A version written for the DEC-10 at the University of Pittsburgh was kindly furnished to us by Steve Ernst. We have implemented some parts of the system as separate programs, including the Fourier transform, peak finding, and bond length and angle calculating programs.

b) Sequence-Structure Correlator. A program which predicts alpha helix and pleated sheet regions of a protein molecule from its amino-acid sequence was furnished to us by Ray Salemme (Univ. of Ariz.). The program was subsequently rewritten in both the SAIL and LISP languages. The algorithm is based on the rules developed by Chou and Fasman (Biochemistry, vol. 13, pp. 211-245 (1974)).

c) Oak Ridge Fast Fourier Program (ORFFP). A system of Fortran programs for generating, analyzing and plotting Fourier maps was obtained from Henry Levy of Oak Ridge and implemented on SUMEX. The plotting segment is the ORTEP program written by Carroll Johnson.

d) Greer Skeletonization Program. A Fortran program for reducing an electron density map of a protein to a set of connected line segments, following the algorithm proposed by Greer, has been written and is currently being debugged. This program will play a pre-processing role in the map interpretation problem, producing a highly abstracted representation of the map.

e) Huber Rotation and Translation Search Program. This is a Patterson

search program which, like our PSRCH, computes correlations of two vector sets in the vector space representation. The program, in Fortran, is on file but has not yet been tested at SUMEX.

f) ROTRAN. These programs were written by B.M. Craven (Univ. of Pittsburgh) and are designed to perform rotational and translational Patterson searching, employing a method developed by Crowther. At present we have only a listing of the Fortran program and instructions for its use.

D. Publications

Feigenbaum, E. A., Engelmore, R. S. and Johnson, C. K., A Correlation Between Crystallographic Computing and Artificial Intelligence Research, submitted for publication in Acta Crystallographica.

II. Interaction with the SUMEX-AIM resource

All program development, and most communications between the two collaborating groups are effected on the SUMEX computer. The UCSD group has a direct connection to SUMEX via the TYMNET computing network (UCSD lost its ARPANET connection during the past year). Routine daily communications now take place using the system's message facility. Program files are equally accessible from Stanford and UCSD, so that either group can construct, edit or exercise the programs. Large data files are transmitted on magnetic tape.

The greatest benefit of the interaction with the SUMEX-AIM resource is the opportunity to share ideas, programming experience and utility programs with other users in the community. The availability of a pool of INTERLISP programmers, for example, has been of great assistance in our initial efforts with the electron density map interpretation task. Members of the SUMEX staff have also been helpful and patient in solving some of the more mundane problems associated with any computational effort (e.g., reading magnetic tapes produced at other computer centers).

IV.A.2   NATIONAL USERS

IV.A.2.a   CHEMICAL SYNTHESIS PROJECT

Simulation and Evaluation of Chemical Syntheses (SECS)

W. Todd Wipke
Department of Chemistry
University of California
Santa Cruz, CA.   95064

I. Summary of Research Program

A. Technical Goals

The long range goal of this project is to develop the logical principles of molecular construction and to use these in developing practical computer programs to assist investigators in designing stereospecific syntheses of complex bio-organic molecules. Previously the focus of our work had been to represent as accurately as possible the fundamental chemical transformations and how steric, proximity, and electronic factors affect these reactions, going into great detail even involving analysis of three-dimensional models. The goals for this year focused on developing constraints to help guide the synthesis program in growing the tree of synthetic precursors. We wanted to utilize high level information about symmetry, and stereochemistry to set up strategies defining preferred orderings of making and breaking bonds. We also hoped to completely separate these strategies from the chemical transforms to allow experimenting with changing transforms keeping strategies constant and vice-versa. This separation was also deemed important for ease of maintenance of large transform libraries. Finally we hoped to use these strategies for guiding multistep lookahead so the user could see a sequence developed in the tree at one time.

B. Medical Relevance and Collaboration

The development of new drugs and the study of how drug structure is related to biological activity depends upon the chemist's ability to synthesize new molecules as well as his ability to modify existing structures, e.g., incorporating isotopic labels into biomolecular substrates. The Simulation and Evaluation of Chemical Synthesis (SECS) project aims at assisting the chemist in designing stereospecific syntheses of biologically important molecules. The advantages of this computer approach over a manual approach are manyfold: 1) greater speed in designing a synthesis; 2) freedom from bias of past experience and past solutions; 3) thorough consideration of all possible syntheses using a more extensive library of chemical reactions than any individual can remember; 4) greater capability of the computer to deal with the many

structures which result; and 6) capability of computer to see molecules in graph theoretical sense, free from bias of 2-D projection. SECS was designed to be able to apply any kind of chemical transformation, and because of this generality we see it finding application in biogenesis and metabolism (see section II A below).

C. Progress and Accomplishments

The environment of this project has changed dramatically in the past year with the move from Princeton to Santa Cruz. SECS was moved from a hands-on environment consisting of a KA PDP-10 with an LDS-1 graphics system and standard DEC software to a remote environment with access to a KI PDP-10 through a GT40 graphics terminal where the host monitor system is TENEX. The compatibility package of TENEX considerably eased the problems of conversion. Most problems resulted from differences in file handling, and differences in the Fortran operating system. Subtle problems arose from the fact that our files were organized by tapes and could not simply all be transfered to disk, because of space problems and naming conflicts. SECS was successfully converted to TENEX and the graphical interaction was modified for greater efficiency in our new remote low bandwidth mode of communication.

Progress in developing strategy includes creating a general goal list structure which allows complex logical combinations of goals to be expressed, for example, "(break bond 2 or break bond 3) and use group 1". Thus, instead of one set of "strategic bonds" to be broken, we now can express strategies involving pairs of bonds, or groups or atoms. We have succeeded in isolating strategy from the chemical transforms--strategies can only contain expressions which refer to structural units of the molecule or changes in those units, and may not refer to any transform by name. The transforms have been given "character" which describes the type of structural changes likely to occur if the transform is applied, e.g., cleaves ring, removes group, modifies stereochemistry, etc. The SECS strategy module first sets up standard goal lists based on graph-theoretic heuristics and then allows the user to view and modify the goal lists. In this way the user can place constraints on the syntheses generated, e.g., "don't modify this ring, instead, focus your attention on this part of the molecule." GOALTST was modified to interpret the new complex goals and also to test the achievement or violation of a goal as early as possible. Hence, there is testing before examination of the transform as well as after interpretation of the transform. The net result of this work on strategy is that the user can very closely constrain SECS now to work only in areas which the user decides is worthwhile, consequently fewer precursors are generated which the user would delete.

Significant progress was made in the recognition of symmetry and use of that information in SECS. A general algorithm based on SEMA, our canonical naming algorithm, was developed and implemented for generating the entire symmetry group of the molecule, using the stereochemical graph isomorphism group. We have applied this symmetry knowledge to make SEMA itself more efficient, and have combined it with symmetry of the transforms to make application of a transform generate a non-redundant set of precursors. Thus, if a double bond is introduced into cyclohexane,

only one cyclohexene is generated, not six. This algorithm takes into account all stereocenters in the molecule of both double bonds and saturated carbon. Addition of this algorithm reduces execution time of SECS on certain types of problems by a factor of up to six or more. We have not yet developed heuristics for creating strategies from this symmetry group.

Considerable improvement of aromatic chemistry has resulted from the addition of "character words" to the aromatic transforms and reorganization of the aromatic module. Electronic perception is only performed now if SECS is fairly certain that aromatic chemistry will be used and the user can prevent the MO calculations if he wishes. Work is still progressing on implementation of strategies to control when to apply aromatic transforms based on heuristics derived from an extensive literature study. Many other modifications have been made to improve the human engineering of SECS. Documentation of modules whose authors have left the project is still continuing. Now with an expanding users group, a good user's manual is required and is under revision.

D.   Current list of Project Publications

[1] W.T. Wipke and T. M Dyott, "Use of Ring Assemblies in a Ring Perception Algorithm," J. Chem. Info. and Computer Sci., 15, 140 (1975).

[2] T.M. Gund, P.v.R. Schleyer, P.H. Gund and W.T. Wipke, "Computer Assisted Graph Theoretical Analysis of Complex Mechanistic Problems in Polycyclic Hydrocarbons. The Mechanism of Diamantane Formation from Various Pentacyclotetradecanes," J. Amer. Chem. Soc., 97, 743 (1975).

Papers in Preparation:

[1] W.T. Wipke and P. Gund, "Simulation and Evaluation of Chemical Synthesis. Congestion: A Conformation Dependent Function of Steric Environment at a Reaction Center. Application with Torsional Terms to Stereoselectivity of Nucleophilic Additions to Ketones," J. Amer. Chem. Soc.

[2] W.T. Wipke, G. Birkhead, and T. Brownscombe, "Correlation of Congestion with Stereoselectivity of Olefin Epoxidation."

[3] W.T Wipke, C. Still, G. Grethe, T.M. Dyott, and P.E. Friedland, "ALCHEM: A Language for Representing Chemical Transforms. Application to Heterocyclic Chemistry."

[4] W. Todd Wipke and Hartmut W. Braun, "Graph-theoretical Perception of Molecular Symmetry.

[5] W.T. Wipke, G. Smith and H. Braun, "SECS-Simulation and Evaluation of Chemical Syntheses: Strategy and Planning," ACS Symposium Proceedings.

E. Funding Status.

    IBM Fellowship supporting S. Krishnan (postdoctoral)
        $4000. expires September, 1976

    Merck, Sharp and Dohme fellowship supporting Graham Smith
    (postdoctoral)
        $1000. expires July, 1976

    Sandoz Unrestricted grant to support computer synthesis
        $2000

    Proposal submitted 1 Mar 1976 to Division of Research Resources
    "Resource-Related Research: Biomolecular Synthesis"
        $391,532 for three years

II. Interactions with the SUMEX-AIM Resource

    A. Collaborations and Medical Use of Programs Through Networks.
Since SECS only recently has been operating on the SUMEX-AIM resource,
collaborations are just beginning. However demonstrations of SECS have
been given at the National Cancer Institute and collaboration with
Division of Chemical Carcinogenesis to try to use SECS for metabolism of
compounds to evaluate carcinogenic activity of metabolites is currently
under discussion. The National Library of Medicine toxicology program is
also interested in SECS and are planning to access SECS via SUMEX-AIM.
Dr. Steve Heller of the EPA and Dr. G.W.A. Milne of the National Heart
and Lung Institute are currently exploring the possibility of putting SECS
up on the Cyphernetics network. For the past year SECS has been available
over the TELENET from First Data Corporation. Squibb, Merck, FMC,
American Cyanamid, Pfizer and Searle pharmaceutical companies have used an
experimental version of SECS and have provided useful feedback to us about
problems they discovered. We expect increasing numbers of academic users
will be accessing SECS via SUMEX-AIM as they learn of its availability.

    The availability of SECS on SUMEX-AIM has also served health-related
research at the University of California, Santa Cruz. For example, model
building using the SECS model builders is being performed for Professor
Edward Dratz (UCSC) to generate conformations of fatty acids isolated from
visual membranes ("Structure and Function of Visual Photoreceptors",
EY00175-05), and for Professor Howard Wang (UCSC) to study how
conformations of steroids may affect the local anesthetic - membrane
interaction ("Role of Membrane Proteins in Local Anesthetic Action,"
GM22242-01).

    B. Cross Fertilization with other SUMEX-AIM projects. The SECS
project held joint research group meetings at Stanford with the DENDRAL
and AI groups to discuss common problems and research goals. This has
been very rewarding since the DENDRAL group has useful experience with
symmetry manipulation which SECS was getting into, and the SECS project
has useful experience with representing reactions, which DENDRAL/CONGEN
was getting into. These joint meetings also let the members meet in
person after having met on-line on the network. Last year's AIM

Conference at Rutgers was also a valuable experience, which allowed us to meet people interested in similar problems in different disciplines, and it also caused us to think about what we were doing in research with some new perspectives. We are looking forward to this year's AIM meeting.

We find the SUMEX-AIM network very well human engineered. The ability to leave messages on the network, and to LINK to other users on-line for advice has been extremely useful to us, since we were new to the TENEX operating system. But more than that, we have been able to utilize expertise of others which our group lacked, e.g., Trisha Davis (an undergraduate) has been writing a model builder and display program in SAIL although there is no SAIL expertise in the SECS group--that would not have been possible without the network communication features.

C. Critique of Resource Services. The SECS project finds the SUMEX-AIM staff and community extremely helpful, and anxious to extend themselves to meet our needs. SUMEX provided a leased line and modems to us and provided TYMNET access as well. Were it not for SUMEX, this research effort would have perished since there is no adequate computer facility on-campus.

We do find we are short of disk space and in our grant proposal we have requested funds for a disk drive to place at SUMEX to help resolve this problem. The response time during the day and sometimes even later is poor for interactive graphics, but hopefully the second processor being installed will help alleviate that difficulty. We have an additional problem that it is difficult to transfer files from TENEX to any other PDP-10 with the files retaining their filenames. This problem may also be resolved if we are able to write tapes locally from over the network. Basically we have found that SUMEX-AIM provides a productive and scientifically stimulating environment and we are thankful that we are able to access the resource and participate in its activities.

IV.A.2.b   INTERNIST (DIALOG) PROJECT

INTERNIST - Diagnostic Logic Program
Dr. H. Pople and J. Myers, M.D.
University of Pittsburgh

(Grant HEW MB-00144-01, 3 years, $167,168 last year)

I. SUMMARY OF RESEARCH PROGRAM

A. BACKGROUND AND OBJECTIVES

The principal objective of the MIS laboratory at the University of Pittsburgh is to develop, test, and implement a computer-based diagnostic consultation system for internal medicine. Considerable progress towards this goal had already been made prior to our receipt, in June, 1974, of a three-year $524,000 grant from the Bureau of Health Resources Development to establish a "Computer Laboratory Health Care Resource' at Pitt. At that time, the medical data base accessed by the internist (formerly DIALOG) program was estimated to comprise approximately twenty-five percent of the major diseases of internal medicine, and a number of case studies had been run illustrating the power of the INTERNIST heuristic process in dealing with a variety of complex clinical problems. Our research plan envisioned a five-year development effort, intended to yield:

(a) A four-fold expansion of the data base.

(b) Systematic field testing and evaluation of the system in actual clinical settings.

(c) Eventual implementation, making INTERNIST available for clinical use on a routine basis.

B. PROGRESS AND ACCOMPLISHMENTS

Shortly after award of the BHRD grant, arrangements were concluded permitting use of the SUMEX-AIM computer resource for INTERNIST research and development activities. Although the SUMEX-AIM computer is of the same genre as the one used in the original INTERNIST development work, differences in the LISP language supported necessitated major conversion efforts.

As mentioned in our last progress report,, the need for rapid access to large data files motivated the design of an interface between the INTERLISP host processor and a set of structured files containing the entire vocabulary and network of associations comprising the INTERNIST data base.

As of June, 1975, conversion of the data base had been completed and

the necessary interfaces had been established to enable INTERNIST diagnostic programs to work with these revised structures. design and implementation of an interactive data entry and editing system, was completed in December, 1975, enabling expansion of the on-line data base to its present size which is approximately 60% complete.

This newly expanded data base is currently being subjected to extensive testing in both typical examples of disease and difficult diagnostic problems. This procedure of systematically checking the entire clinical data base should be completed by late June, 1976. The planned field test and evaluation effort will then commence in early fall.

C. PUBLICATIONS

[1] Pople, H.E., "Artificial Intelligence Approaches to Computer-based Medical Consultation; Proceedings of IEEE Intercon, 1975, New York.

[2] Pople, H.E., Myers, J.D., and Miller, R.A., "DIALOG: A Model of Diagnostic Logic for Internal Medicine," Proceedings of Fourth International Joint Conference on Artificial Intelligence, Tbilisi, Georgia, USSR, 1975.

II. INTERACTIONS WITH SUMER-AIM RESOURCE

Because this year has been largely devoted to system development and checkout activities, there has been no real opportunity to engage in any meaningful collaboration via the communication networks associated with SUMEX-AIM. We fully expect to exploit this attractive feature of the resource, however, during the evaluation and field test studies planned for the coming year.

Concerning the service provided by SUMEX-AIM, our only complaint is the heavy loading during prime hours, which effectively prevents serious use of the INTERNIST diagnostic programs during certain portions of the day. We applaud and eagerly await the advent of the SUMEX-AIM dual processor.

IV.A.2.c   HIGHER MENTAL FUNCTIONS MODELING

HIGHER MENTAL FUNCTIONS MODELING (HMF)
Project Summary - 1976

Kenneth M. Colby, M.D.
Professor of Psychiatry, UCLA

(NIH MH-27132-01, 2 years, $67,000 this year)

Introduction.

One of the oldest and newest applications of computers in artificial intelligence is the simulation of human cognitive processes. The Higher Mental Functions project has been modelling belief systems and related psycho-pathological delusional systems for a number of years. The specific goal for the past two years has been to construct, test, and validate a computer simulation of paranoid processes. The development of such a model has clinical implications for the understanding, treatment, and prevention of paranoid disorders.

Recently we have been focussing on the origin of beliefs in belief systems and the criteria by which beliefs are significant to the entity we are modelling; i.e. the motivation for the beliefs. The motivation for an entity's purposive behavior is based in its affect or emotion system. We are currently formulating a theory of the motivational influence of affect on conative (volitional) and cognitive (inferential) processes, with the intent of implementing this background theory in a simulation model. By specifying the underlying theory of motivation we hope to make the theory of paranoia more explicit and the paranoid simulation model more adequate.

The strategy of computer simulation of mental processes can be characterized roughly by three phases: (1) identification and critical description of non-random patterns occuring in the phenomena under study, (2) explanation by postulation of underlying mechanisms which generate, produce, or are responsible for the non-random patterns, and (3) validation by repeated attempts to test the reality of the proposed theory or model. The construction and use of simulation models of mental processes closely parallels model-building in other sciences. An attempt is made to reproduce the relevant features of the patterns under study. This attempt produces simplification and idealization of the phenomena. Simplification implies that only centrally relevant variables are chosen for representation in the model. Idealization implies that exact classes and perfect properties are assumed in the implementation of the model. Still, the model can provide an explanation of underlying mechanisms which is useful in understanding and interpreting the observed patterns of phenomena. Finally, a model can be used in practical situations for prediction, and for providing suggestions to clinicians for potential control and change in the phenomena. Such technological purposes are important for models of mental disorders since the long-range goal of

mental health research is to prevent or reduce conditions of human mental suffering.

Simulations of cognitive processes are difficult for a number of reasons:

(1) The underlying generative mechanisms of human behavior are inaccessible to direct observation, and must instead be postulated as hypothetical constructs that may (possibly) account for the phenomena.

(2) A simulation must take into account the rich background of information that a human has available to apply to a contemporary situation.

(3) Human beings have internal needs which are a function of the immediate past and present, as well as the long-term past of the individual. These needs and past experiences color the human's response to the contemporary situation.

(4) Human linguistic behavior is the richest source of data for exploring cognitive processes as well as the most complex and therefore desirable behavior to simulate. At the same time, it is difficult due to the variety of behavior possible and the variety of explanations possible for one specific linguistic action.

(5) Once a simulation model is performing, it is difficult to show the subtleties of the model's generative mechanisms. Instead, some attempt is usually made to reveal the internal workings of the model and appeal to the observers' intuition and/or introspection.

Our overall purpose, then, is to develop theories of human mental processes, specifically psycho-pathological processes, and to implement these theories in computer simulation models. The simulation models help formalize and explicate their associated theories by forcing them into a single notation and requiring the theory-builders to specify the details of the theory. In addition, the models provide a testing ground for validating the theory. On the basis of such theories, we hope to explain the origin of psycho-pathologies and offer principles on which to base treatment and prevention.

Technical goals.

A. Expand the theory of paranoia. The theory implemented in the current model (PARRY2), the humiliation theory, postulates that informational inputs from other people activate a belief in the self's inadequacy. The paranoid mode then consists of strategies which forestall or ward off an impending negative affect experience of humiliation by negating this belief that the self is wrong and asserting the belief that the self is being wronged by others. The theory provides generative mechanisms for explaining the expression of a delusional system by a paranoid person, the chronic distress felt by paranoid persons, and for the sudden and extreme displays of fear and anger in interactions with other people.

We plan to extend the theory in two ways:

(1) To cover other paranoid phenomena, such as delusions of grandeur, the transformation of counter-evidence to evidence supportive of delusions, and retrospective misinterpretation of input expressions.

(2) To explain the genesis of paranoid patterns of thought; e.g., the manner in which: (a) normal strategies for dealing with the shame-humiliation affect are ineffective and paranoid strategies develop, (b) strategies are selected as being appropriate and are reinforced when they are successful, and (c) persecutory delusional systems develop and expand to include much of the paranoid personality's cognitive processing.

B. Expand the background theory and model of the motivation of cognitive processes. The more important characteristics for the model to have are:

(1) The top-level processes of the model should be purposive intentional processes guided by the affect system, rather than a question-answer loop or facsimile.

(2) Every action that the model performs should be motivated by an intention. These intentions may be explicit in the case of goals, or implicit in the case of an action appropriate to a situation but with no explicit end state represented.

(3) Each belief and intention should have an associated measure of its significance to the entity. The criteria for measuring the significance are based in the affect system.

(4) The model should have a number of coping mechanisms for avoiding or coping with distressing situations. These mechanisms can be reinforced or discarded as they are proved to be more or less useful to the model.

(5) The model should be able to change over time to show the development of psycho-pathologies. The most direct form of change is to the measures of significance attached to the model's beliefs and intentions.

C. Implement the theories in a simulation model. We hope to construct a model in such a way that the theories of motivation and paranoia can be represented explicitly, and therefore be open to inspection and modification. In addition, since we model paranoid behavior as expressed through linguistic actions, we hope to develop adequate natural language understanding programs for recognition and response in dialog situations.

D. Develop further techniques and methods for simulating cognitive processes. Specifically, we plan to explore human communication through natural language in dialog situations and human natural language interfaces with computers. Also, we will extend our previous attempts at finding stronger and more sophisticated tests for validation studies. Our

results should be applicable to other simulations of human mental activities.

Medical relevance.

The simulation model of paranoid processes that we are implementing has implications for the understanding, treatment, and management of paranoid disorders.

The shame-humiliation theory and its model suggest that paranoid phenomena be viewed as a consequence of intentionalistic information-processing strategies which attempt to avoid or minimize the distressing experience of humiliation. In trying to understand what is going on in a paranoid patient at a symbol-processing level, this perspective directs clinicians to look for humiliating and shame-engendering situations in the patient's experience. These may consist of a single, encompassing humiliating situation such as a demeaning job, or a series of esteem-damaging defeats such as repeated failures in disappointing love affairs.

Since activation of intense shame is posited to be the core process in paranoid disorders, implications for treatment involve trying to modify this central mechanism in some way. One method is to change the patient's distressing belief in his own inadequacy by exploring topics involving shame, esteem, and self-censure. Another is to desensitize the paranoid patient to shame experiences through behavior therapy involving a graded hierarchy of imagined distress situations and countering procedures. These treatment procedures may be deduced directly from mechanisms in the model, and the theory used to predict the outcomes of such procedures.

For management of the disorder, the model predicts that removal from situational humiliation, as in hospitalization, allows for repair from breakdowns occurring under repeated activations of shame-humiliation beliefs. Also, if the patient returns to unchanged situational humiliation, as in a distressing home life, he risks a relapse.

Current Status.

PARRY2 was completed a year ago and has been available for interviewing and validation tests on the SUMEX system for the past year. We are now in the process of writing a new version, PARRY3, incorporating the theoretical constructs presented in this report. The new version (which is being completely rewritten) contains mechanisms implementing the characteristics of models mentioned above with the exception of the ontogenesis of psycho-pathologies. In addition, it contains a new language recognizer capable of combining pattern-matching rules with parsing techniques, and explicit rules for recognizing and interpreting elliptical expressions in dialogs.

Current publications.

[1] Colby, K.M. Artificial Paranoia, Pergamon Press, New York, 1975.

[2] Faught, W.S. Affect as Motivation for Cognitive and Conative Processes, IJCAI Proceedings, September, 1975.

[3] Colby, K.M., Hilf, F.D., Wittner, W.K., Parkison, R.C. and Faught, W.S., A Note on Estimating Improvement in a Computer Simulation of Paranoid Processes. UCLA Department of Psychiatry, Memo ALHMF-2, April 1975.

[4] Parkison, R.C., Colby, K.M., and Faught, W.S., Conversational Language Comprehension Using Integrated Pattern-Matching and Parsing. UCLA Department of Psychiatry, Memo ALHMF-5, April 1976.

[5] Colby, K.M. Clinical Implications of a Simulation Model of Paranoid Processes. Archives of General Psychiatry, 1976.

[6] Faught, W.S., Colby, K.M., and Parkison, R.C., Inferences, Affects, and Intentions in a Model of Paranoia. Cognitive Psychology, 1976.

Funding Status.

Interactions with the SUMEX-AIM resource.

The SUMEX-AIM resource and its associated network connections make possible the merging of artificial intelligence techniques and technology in psychiatry and the resources of a west-coast center for psychiatric studies, the Neuro Psychiatric Institute (NPI) at UCLA. Access to SUMEX from the NPI has brought a new source of questions and viewpoints to research in mental disorders, as well as an opportunity for the model-builders to interact with clinicians in elaborating details of paranoid phenomena. In addition, the current simulation model is being explored for use as a training device for medical students and residents in the Department of Psychiatry.

Critique of resource services.

The resource itself has provided excellent facilities and service for our research needs. We have had almost no trouble developing simulations due to the SUMEX system itself, and the cooperation of the SUMEX staff has been excellent.

In spite of this, problems have arisen with our use of the resource, due almost entirely to the network aspect of our access. (We connect to SUMEX through both the ARPA net and TYMNET.) We see the problems of network use of the SUMEX-AIM facility as falling into four broad categories:

(a) the keyhole effect. The slow terminal rates typical of a network connection (due to phone lines or local computer delay) force the user to peer at his files and communicate with the computer through

a small data channel. Additionally, the network user may not have direct access to a high-speed printing device for listing the day's (or even the week's) work.

(b) the dropped connection. Typically, the network and/or local connection of the user to the net can drop, leaving the user's job in a dormant (non-running) state, forcing the user to reconnect (and perform the typically elaborate reconnect procedure) before his job will run again.

(c) slower interactive computer response due to network and local hardware, and subsequent greater amount of time necessary to get any work done.

(d) the difficulty of lobbying for system changes and/or additions from a distance. We are made acutely aware of this fact whenever notices are given over the system for classes or seminars explaining new system features.

Policies and programs that might be useful:

(a) more system programs designed with the network user in mind, such as status programs to report the status of detached or batch-run jobs, and cleaner detach and attach programs for reestablishing dropped connections.

(b) a service-level advantage to non-local users to put them on a more equal footing with local users.

(c) a special effort to elicit and implement improvements for network access.

In general, we have found the system reliable, and the staff courteous and helpful.

IV.A.2.d   LANGUAGE ACQUISITION MODELING PROJECT

Language Acquisition Modeling (ACT)

Dr. John Anderson
(Univ. of Michigan)

(Grant NIMH $20,000 this year)

I. Summary of Research Program

A. Technical goals:

To develop a production system that will serve as an interpreter of the active portion of an associative network. To model a range of cognitive tasks including memory tasks, inferential reasoning, language processing, and problem solving. To develop an induction system capable of acquiring cognitive procedures with a special emphasis on language acquisition.

B. Medical relevance and collaboration:

1. The ACT model is a general model of cognition. It provides a useful model of the development of and performance of the sorts of decision making that occur in medicine.

2. The ACT model also represents basic work in AI. It is in part an attempt to develop a self-organizing intelligent system.. As such it is relevant to the goal of development of intelligent artificial aids in medicine.

C. Progress and Accomplishments:

ACT provides a uniform set of theoretical mechanisms to model such aspects of human cognition as memory, inferential processes, language processing, and problem solving. ACT's knowledge base consists of two components, a propositional component and a procedural component.The propositional component is provided by an associative network encoding a set of facts known about the world. This provides the system's semantic memory. The procedural component consists of a set of productions which operate on the associative network. ACT's production system is considerably different than many of the other currently available systems (e.g., Newell's PSG). These differences have been introduced in order to create a system that will operate on an associative network and in order to accurately model certain aspects of human cognition.

A small portion of the semantic network is active at any point in time. Productions can only inspect that portion of the network which is active at the particular time. This restriction to the active portion of

the network provides a means to focus the ACT system in a large data base of facts. Activation can spread down network paths from active nodes to activate new nodes and links. To prevent activation from growing continuously there is a dampening process which periodically deactivates all but a select few nodes. The condition of a production specifies that certain features be true of the active portion of the network. The action of a production specifies that certain changes be made to the network. Each production can be conceived of as an independent "demon". Its purpose is to see if the network configuration specified in its condition is satisfied in the active portion. If it is, the production will execute and cause changes to memory. In so doing it can allow or disallow other productions which are looking for their conditions to be satisfied. Both the spread of activation and the selection of productions are parallel processes whose rates are controlled by "strengths" of network links and individual productions. An important aspect of this parallelism is that it is possible for multiple productions to be applied in a cycle through the set of productions. Much of the early work on the ACT system was focused on developing computational devices to reflect the operation of parallel, strength-controlled processes and working out the logic for creating functioning systems in such a computational medium.

We have successfully implemented a number of small-scale systems that model various psychological tasks in the domain of memory, language processing, and inferential reasoning. A larger scale effort is underway to model the language processing mechanisms of a young child. This includes implementation of a productions system to analyze linguistic input, make inferences, ask and answer questions, etc. Also a great deal of effort is being given to developing learning mechanisms that will acquire and organize the productions for this language processing. This learning program attempts to acquire procedures from examples of the computations desired of the procedures. For instance, the program learns to comprehend and generate sentences by being given sentences and picture representations of the meaning of the sentences(actually hand encodings of the pictures). Although this effort is focused on induction of linguistic procedures, the hope is to develop a general model of induction of cognitive procedures and not to place any language-specificity into the induction procedures.

D. Current list of project publications:

[1] Anderson, J. R. Induction of augmented transition networks. Cognitive Science, 1976, in press.

[2] Anderson, J. R. Language, Memory, and Thought. Lawrence Erlbaum Assoc., 1976, in press.

[3] Anderson, J. R. ,Klein, P., and Lewis, C. Language processing by production systems. To appear in P. Carpenter and M. Just(Eds.) Cognitive Processes in Comprehension,Lawrence Erlbaum Assoc.

E. Funding Status:

The research is currently being funded by a grant from NIMH for computer simulation of language acquisition. The level of funding for the year beginning May 1, 1976 has yet to be determined. It was $20,000 for the past year.

II. Interactions with SUMEX-AIM Resource.

Our period with the project has been too short to develop any significant interactions. The ACT program is currently being made a system which will be available to members of the SUMEX-AIM community.

IV.A.2.e   MEDICAL INFORMATION SYSTEMS LABORATORY

MISL Project

Dr. B. McCormick and M. Goldberg, M.D.
(Univ. of Illinois at Chicago Circle)

(Grant HEW MB-00114-02, 3 years, $248,793 this year)

I) Summary of research program

A) Technical goals

    The major goals of the Medical Information Systems Laboratory fall
into three broad categories, described briefly as follows:

  1) Construction of a database in ophthalmology. This will provide a
     trial setting for clinical decision-making research. Four major
     activities are involved: implementation of a clinical data network;
     design and on-going development of an Eye Outpatient Index; computer
     systems development; and installation of a glaucoma clinic satellite.

  2) Network-compatible database design. This will provide cost-effective
     distributed data management for clinical records. Current projects
     include: design of an intelligent coupler for the database system;
     continuing design of various levels of database software (a
     relational algebraic language -- RAIN, disk controller, database
     skeleton); large database / database software compatibility; design
     of a separate database computer; design of database network software.

  3) Clinical decision support

     a. Construction of a consultation system for use in the diagnosis and
        treatment of retinal/choroidal diseases. The immediate goal is
        the development of a system for giving advice about four diseases
        prevalent at the University of Illinois Eye and Ear Infirmary:
        histoplasmosis, central serous retinopathy, diabetic retinopathy,
        and sickle cell disease. Besides actual construction of a system,
        the project is concerned with theory of diagnosis and knowledge
        acquisition methodology.

     b. Interface between a pattern recognition system for Structured
        Analysis of the Retina (STARE) and the diagnostic model used by
        the retinal/choroidal consultation system.

     c. Initial development of a causal model for motility.

     d. Further inter-institutional communications in disease model theory
        and development.

B) Medical relevance and collaboration

We have chosen to explore inferential relationships between analytic clinical data and the natural history of glaucoma and selected retinal/choroidal diseases both in treated and untreated form. These investigations are intended to provide much more than a simple excursion of computer technology into ophthalmology. They address clinical problems of national interest, as indicated in the Report of the National Eye Advisory Council Vision Research Program Planning Committee (DHEW publication No. NIH-75-644).

Glaucoma, one of the major causes of blindness in the United States today, is difficult to diagnose in its early stages. Some recent evidence indicates that enlargement of the optic nerve cups may be the first sign of glaucoma's damage to the eye. One of the goals of the present project is the application of a newly developed technique for quantitative analysis of the optic nervehead. The technique is sufficiently simple to permit wide-spread adoption. If this technique is successful in identifying very early glaucomatous disk changes, it should permit institution of therapy at a very early stage, and thereby prevent serious glaucomatous damage from being done to the eye.

Diabetic retinopathy is another principal cause of blindness. Very little is known about its pathophysiology, and there are many gaps in our knowledge of its natural course. The present study is designed to elicit new information about this disease, using a series of new diagnostic tools which have been developed as part of a system of computerized retinal image analysis. The need here is great, because at present there is no proven satisfactory treatment.

Sickle hemoglobinopathy can cause ocular changes that lead to loss of vision and even total blindness. Little is known about the natural history of this problem, particularly during its early stages. The present project is ideally suited to assist in this study; the nation's first Sickle Cell Clinic has been established at the Illinois Eye and Ear Infirmary -- the site of our Ophthalmic Database System.

At present there is great demand in the United States for improved efficiency in the delivery of medical care. Two ways that this can be accomplished are: 1) by increased use of paramedical personnel to perform jobs currently being performed by physicians, and 2) by use of automated equipment to perform tasks previously performed by the physician. In the current project we utilize both these methods in the screening of new ophthalmic patients at the Illinois Eye and Ear Infirmary. If we can show that these methods are not only feasible, but also improve the efficiency and reliability of patient care, then a major contribution to ophthalmic care for patients in large ambulatory care centers around the country will have been made.

Modeling of clinical decision-making is best carried out in intimate association with an extensive referral clinic where a sufficient patient population can be accumulated to provide an adequate biostatistical sample. In the prescribed setting, the Illinois Eye and Ear Infirmary, the Medical Information Systems Laboratory has access to:

- clinical expertise provided by a house staff of 45;

- 28 residents, all of whom are required to undertake some research work as a requirement of their appointment, are available to explore latent contingencies of the database;

- an indigenous, relatively stable population (25% white, 75% black and other minorities) of a medically underserved portion of the inner city of Chicago; the clinic provides ophthalmic services to 50,000 patients per year.

Commonality of the diseases being studied assists construction of an adequate biostatistical sample. Roughly 2% of the general population exhibit symptoms of diseases treated at each of three specialty clinics (Glaucoma, Retina, and Motility), or allowing for multiple presenting of symptoms, approximately 5% by population.

Besides a strong clinical research orientation, the Medical Information Systems Laboratory brings to the study of disease a history of successful engineering-medical collaboration. MISL's sister project, "Image Processing in Clinical Ophthalmology," lists the development of the digital television ophthalmoscope as one of its achievements. This device will be a major source of clinical data for our Ophthalmic Database.

C) Progress and accomplishments (of the Clinical Decision
   Support activity only)

Interaction has continued over SUMEX-AIM with the authors of the Weiss/Kulikowski glaucoma modeling program. We are now entering cases into the glaucoma system at the rate of approximately 5 per week.

Work on the consultation system for retinal/choroidal diseases has progressed along two fronts. While interviewing an expert diagnostician, in order to build a knowledge base for the four diseases mentioned above, we have been piecing together a theory of diagnosis for ocular fundus diseases. We have attempted to incorporate pieces of the expert's knowledge in a "fuzzy" diagnostic model, based partly on multiple-cue probability theory, partly on fuzzy set and confirmation theory. The framework of the model is a hierarchy of disease categories, each with a significance tempered by functions built into the system. Our efforts have centered on the acquisition of categories for histoplasmosis, central serous retinopathy, and diabetic retinopathy, but should soon also include sickle cell disease.

Our experience in interviewing experts has pointed the way to a knowledge acquisition methodology that is compatible with our thoughts on diagnostic reasoning. Specifically we plan to store, for each disease category, a representation for the contexts (or frames) in which the disease's attributes apply. We conceive of a disease category as a "sphere" (embodying a structural model of the disease) in a hyperspace defined by dimensions on attributes. The significance of (or "belief in") the model of disease is modified by interactions between attributes. During acquisition, and after contexts have been defined, we can simulate clinical situations for the benefit of our expert, who indicates his level of belief in the disease model in the given situation. This we plan to do

with the help of plasma panels, for graphical presentations of the relevant contexts comprising each situation. This approach is especially convenient for specifying typical "default" situations, and for modeling the time course of disease (in terms of modifications on attributes).

Presently, while we continue interviews with our expert, we are formalizing our diagnostic model and expect shortly to finish an in-depth report.

D) Current list of project publications

[1] Chang Shi-Kuo, O´Brien M., Read J., Borovec R., Cheng W. H., Ke J. S. (1976) Design considerations of a database system in a clinical network environment. Accepted for 1976 National Computer Conference.

[2] Chang S. K. (1975) Preliminary report on the implementation of a relational data base management system with structurally decomposed relations. MISL internal report M.D.C. 1.1.3.

[3] Chang S. K. and McCormick B. H. (1975) An intelligent coupler for distributed database systems. MISL internal report M.D.C. 1.1.7.

[4] Malone J. (1975) User´s guide to uniclass cover synthesis. MISL report M.D.C. 4.4.1.

[5] Malone J. (1975) Addendum to AQVAL/1 (AQ7), part 1: User´s guide and program description. MISL report M.D.C. 4.4.1.

[6] Manacher G. K. (1975) On the feasibility of implementing a large relational data base with optimal performance on a minicomputer. Proceedings of the International Conference on very large data bases, Framingham, Mass.

[7] McCormick B. H. and Wilensky J. (1975) Clinical knowledge acquisition: design of a relational data base in ophthalmology. To be published in Proceedings Second Annual Medical Informations Systems Conference, Urbana, Ill.

[8] McCormick B. H., Goldberg M. F., and Read J. S. (1974) Clinical decision-making: design of a data base in ophthalmology. Proceedings First Annual Medical Information Systems Conference, Urbana Ill.

[9] Michalski, Ryszard S. (1975) On the selection of representative samples from large relational tables for inductive inference. MISL internal report M.D.C. 1.1.9.

[10] Murata T. (1976) Equations governing the behavior of E-nets. MISL internal report.

[11] Murata T. (1975) State equation, controllability, and maximal matchings of petri nets. MISL report M.D.C. 1.1.10.

[12] Murata T. and Church R. W. (1975) Analysis of marked graphs and petri nets by matrix equations. MISL report M.D.C. 1.1.8.

[13] Vere S. A. (1975) Induction of concepts in the predicate calculus.
    Proceedings of the Fourth International Joint Conference on
    Artificial Intelligence vol 1., Tbilisi, U.S.S.R.

[14] Vere S. A. (1975) Relational productions systems. MISL internal
    report M.D.C. 1.1.5.

E) Funding status

        Year 02 -- 6/30/75 - 6/29/76 : $248,793.

        Year 03 -- 6/30/76 - 6/29/77 : $228,000.

II) Interactions with the SUMEX-AIM Resource

A) & B) Collaboration, cross-fertilization

        Most of our interaction of late has involved the Glaucoma Network
fostered by the Rutgers Computers in Biomedicine group. This network has
made it especially convenient for our expert in glaucoma, Dr. Jacob
Wilensky, to maintain close contact with investigators around the country.

        In addition, monitoring of SUMEX-AIM system messages has helped us
keep abreast of developments in other projects. We have come to rely on
this facility as a vital source of up-to-date information.

C) Critique of resource services

        In our view SUMEX-AIM services are excellent. We have been very
pleased with the prompt and personal attention given to our requests by
the resource staff.