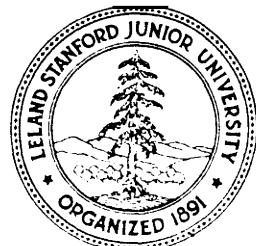# THE ART OF ARTIFICIAL INTELLIGENCE:
I. THEMES AND CASE STUDIES OF KNOWLEDGE ENGINEERING

by

E. A. Feigenbaum

COMPUTER SCIENCE DEPARTMENT
School of Humanities and Sciences
STANFORD UNIVERSITY

THE ART OF ARTIFICIAL INTELLIGENCE:

I.   Themes and Case Studies of Knowledge Engineering

STAN-CS-77-621
Heuristic Programming Project Memo 77-25

Edward A. Feigenbaum

Department of Computer Science
Stanford University
Stanford, California

ABSTRACT

The knowledge engineer practices the art of bringing the principles and tools of AI research to bear on difficult applications problems requiring experts' knowledge for their solution. The technical issues of acquiring this knowledge, representing it, and using it appropriately to construct and explain lines-of-reasoning, are important problems in the design of knowledge-based systems. Various systems that have achieved expert level performance in scientific and medical inference illuminates the art of knowledge engineering and its parent science, Artificial Intelligence.

THE ART OF ARTIFICIAL INTELLIGENCE:

I. Themes and case studies of knowledge engineering

Edward A. Felgenbaum

Department of Computer Science,
Stanford University,
Stanford, California, 94305.

Abstract

The knowledge engineer practices the art of bringing the principles and tools of AI research to bear on difficult applications problems requiring experts' knowledge for their solution. The technical issues of acquiring this knowledge, representatfng it, and using it appropriately to construct and explain lines-of-reasoning, are important problems in the design of knowledge-based systems. Various systems that have achieved expert level performance in scientific and medical inference illuminate the art of knowledge engineering and its parent science, Artificial Intelligence.

1 INTRODUCTION: AN EXAMPLE

This is the first of a pair of papers that will examine emerging themes of knowledge engineering, illustrate them with case studies drawn from the work of the Stanford Heuristic Programming Project , and discuss general issues of knowledge engineering art and practice.

Let me begin with an example new to our workbench: a system called PUFF, the early fruit of a collaboration between our project and a group at the Pacific Medical Center (PMC) in San Francisco.

A physician refers a patient to PMC's pulmonary function testing lab for diagnosis of possible pulmonary function disorder. For one of the tests, the patient inhales and exhales a few t'imes in a tube connected to an instrument/computer combination. The instrument acquires data on flow rates and volumes, the so-called flow-volume loop of the patient's lungs and airways. The computer measures certain parameters of the curve and presents them to the diagnostician (physician or PUFF) for interpretation. The diagnosis is made along these lines: normal or diseased; restricted lung disease or obstructive airways disease or a combination of both; the severity; the likely disease type(s) (e.g. emphysema, bronchitis, etc.); and other factors important for diagnosis.

PUFF is given not only the measured data but also certain items of information from the patient record, e.g. sex, age, number of pack-years of cigarette smoking. The task of the PUFF system is to infer a diagnosis and print it out in English in the normal medical summary form of the interpretation expected by the referring physician.

Everything PUFF knows about pulmonary function diagnosis is contained in (currently) 55 rules of the IF...THEN... form. No textbook of medicine current l y records these rules. They constitute the partly-publ ic, partly-private knowledge of an expert pulmonary physiologist at PMC, and were extracted and polished by project engineers working intensively with the expert over a period of time. Here is an example of a PUFF rule (the unexplained acronyms refer to various data measurements):

RULE 31

IF:
1) The severity of obstructive airways disease of the patient is greater than or equal to mild, and
2) The degree of diffusion defect of the patient is greater than or equal to mild, and
3) The tlc(body box)observed/predicted of the patient is greater than or equal to 110 and
4) The observed-predicted difference in rv/tlc of the patient is greater than or equal to 10

THEN:
I) There is strongly suggestive evidence (.9) that the subtype of obstructive airways disease is emphysema, and
2) It is definite (1.0) that "OAD, Diffusion Defect, elevated TLC, and elevated RV together indicate emphysema." is one of the findings.

One hundred cases, carefully chosen to span the variety of disease states with sufficient exemplary information for each, were used to extract the 55 rules. As the knowledge emerged, it was represented In rule form, added to the system and tested by running additional cases. The expert was sometimes surprised, sometimes frustrated, by the occasional **gaps** and inconsistencies in the knowledge, and the incorrect diagnoses that were logical consequences of the existing rule set. The interplay between knowledge engineer and expert gradually expanded the set of rules to remove most of these problems.

As cumulation of techniques in the art demands and allows, a new tool was not invented when an old one would do. The knowledge engineers pulled out of their toolkit a version of the MYCIN system (to be discussed later), with the rules about Infectious diseases removed, and used It as the inference engine for the PUFF diagnoses. Thus PUFF, like MYCIN, is a relatively simple **backward-**chaining inference system. It seeks a valfd line-of-reasoning based on **its** rules and rooted in the Instrument and patient **data.** With a little more work at fitting some existing tools together, PUFF will be able to explain this line-of-reasoning, just as MYCIN does.

As it is, PUFF only prints out the final interpretation, of which the following is an example:

_____   _____

PATIENT DATA:
The degree of dyspnear MODERATELY-SEVERE
The severity of coughing: MILD
Sputum production MODERATELY-SEVERE
The number of pack-years of **smoking:** 48
referral diagnosis: BRONCHITIS
IVC/IVC-predicted: 8 0
RV/RV-predic ted : 191
FVC/FVC-predicted: 8 7
TLC (body box) observed/predicted: 127
Predfcted **FEV1/FVC:** 83
TLC **(DLCO) observed/predicted:** 83
FEV1/FVC ratio: 50
RV/TLC Observed-Predicted: 21
MMF/MMF-predicted: 19
the-DLCO/DLCO-predicted: 48
The slope **(F50_obs-F25_obs)/FVC_obs:** 19

DEGREE OF OBSTRUCTIVE AIRWAYS DISEASE:

GAD degree by SLOPE: (MODERATELY-SEVERE 700)
OAD degree by **MMF:**   (SEVERE 900)
OAD degree by FEVI: **(MODERATELY-SEVERE 700)**

FINAL OAD DEGREE: **(MODERATELY-SEVERE 910) (SEVERE 900)**
No conflict.   Final degree:
(MODERATELY-SEVERE 910)

INTERPRETATION :

Obstruction is indicated by curvature of the flow-volume loop.
Forced Vital Capacity is normal and peak flow rates are reduced, suggesting airway obstruction.
Flow rate from 25-75 of expired volume is reduced, indicating severe airway obstruction.
OAD, Diffusion Defect, elevated TLC, and elevated RV together indicate emphysema.
OAD, Diffusion Defect, and elevated RV indicate emphysema.
Change in expired flow rates following bronchodilation shows that there is reversibility of airway obstruction.
The presence of a productive cough is an indication that the OAD is of the bronchitic type.
Elevated lung volumes indicate overinflation.
Air trapping is indicated by the elevated difference between observed and predicted RV/TLC ratios.
Improvement in airway resistance indicates some reversibility of airway
Airway obstruction is consistent with the patient's smoking history.
The airway obstruction accounts for the patient's dyspnea.
Although bronchodilators were not useful in this one case, prolonged use may prove to be beneficial to the patient.
The reduced diffusion capacity indicates airway obstruction of the mixed bronchitic and emphysematous types.
Low diffusing capacity indicates loss of alveolar capillary surface.
Obstructive Airways Disease of mixed types

_____   _____

150 cases not studied during the knowledge acquisition process were used for a test and validation of the rule set. PUFF inferred a diagnosis for each. PUFF-produced and **expert-**produced interpretations were coded for statistical analysis to discover the degree of agreement. Over various types of disease states, and for two conditions of match between human and computer diagnoses ("same degree of severity" and "within one degree of severity"), agreement ranged between approximately 90% and 100%.

The PUFF story **is** just **beginning** and will be told perhaps at the next IJCAI. The surprising punchline to my synopsis is that the current state of the PUFF system as **described** above was achieved In less than 50 hours of interaction with the expert and less than **10** man-weeks of effort by the knowledge engineers. We have learned much in the

past decade of the art of engineering **knowledge-**based intelligent agents!

In the remainder of this essay, I would like to discuss the route that one research group, the Stanford Heuristic Programming Project, has taken, Illustrating progress vith case studies, an&. discussfng themes of the work.

2 ARTIFICIAL INTELLIGENCE & KNOWLEDGE ENGINEERING,

The dichotomy that **was** used to classify the collected **papers** in the volume Computers and Thought still characterizes well the motivations and research efforts of the AI community. First, there are some who **work** toward the construction of Intelligent artifacts, or seek to uncover principles, methods, and techniques useful in such construction. Second, there are those who view artificial intelligence as (to use Newell's phrase) "theoretical psychology," seeking explicit and valid information processing models of human thought.

For purposes **of-this** essay, I **wish** to focus on the motivations of the first group, these days by far the larger of the two. I label these motivations "the Intel **ligent** agent viewpoint" and here is my understanding of that viewpoint:

'The potential uses of computers by people to accomplish tasks can be **'one-**dimensional **ized.** into a spectrum representing the nature of instruction that must be given the computer to do its job. Call It the WHAT-TO-HOW spectrum. At one extreme of the spectrum, the user supplies his intelligence to instruct the machine **with** precision exactly ROW to do his job, step-by-step. Progress in Computer Science can be seen as steps avay from the extreme 'HOW' point on the spectrum: the familiar panoply of assembly languages, subroutine libraries, compilers, extensible languages, etc. At the other extreme of the spectrum is the user **with** his real problem **(WHAT** he wishes the computer, as his instrument, to do for him). He aspires to **communicate** WHAT he wants done in a language that is comfortable to him (perhaps English) **;** via communication modes that are convenient for him (including perhaps, speech or pictures); with **some** generality, some vagueness, **imprecision,** even error ; without having to lay out in detail all necessary subgoals for adequate performance ─ with reasonable assurance that he is addressing an Intelligent agent that is using knowledge of his world to understand his intent, to fill in his vagueness, to make specific his abstractions, to correct his errors, to discover appropriate subgoals, and

ultimately to translate WHAT he really wants done into processing steps that define HOW it shall be done by a real computer. The research activity aimed at creating computer programs that act as "intelligent agents' near the WHAT end of the WHAT-To-HOW spectrum can be viewed as the long-range goal of AI research." **(Feigenbaum,** 1974)

Our young science is still more art than **science.** Art: 'the principles or methods governing any craft or branch of learning." Art: "skilled workmanship, execution, or agency. **"** These the dictionary teaches **us. Knuth** tells us that the endeavor of computer programming is an art, in just these ways. The art of constructing Intelligent agents is both part of and an extension of the programming art. It is the art of building complex computer programs that represent and reason with knovledge of the world. Our art therefore lives in symbiosis with the other worldly arts, whose practitioners -- experts of their art -- bold the knowledge ve need to construct intelligent agents. In most "crafts or branches of learning" what **we** call "expertise" is the essence of the art. And for the domains of knowledge that we touch with our art, it is the "rules of expertise" or the rules of "good judgment" of the expert practitioners of that domain that we seek to transfer to our programs.

2.1 Lessons of the Past

Two Insights from previous work are pertinent to this essay.

The first concerns the quest for generality and power of the Inference engine used In the performance of intelligent acts (what Minsky and **Papert** [see Goldstein and **Papert,** 1977] have labeled "the power strategy"). We must hypothesize from our experience to date that the problem solving power exhibited in an intelligent agent's performance is primarily a consequence of the specialist's knowledge employed by the agent, and only very secondarily related to the generalfty and power of the inference method employed. Our agents must be knovledge-rich, even If they are methods-poor. In 1970, reporting the first major **summary-of-results** of the DENDRAL program (to be dfscussed later), ve addressed this issue as follows :

"...general problem-solvers are too **weak** to be used as the basis for building high-performance systems. The behavior of the best general problem-solvers we know, human problem-solvers, is observed to be weak and shallow, except in the areas in which the human problem-solver is a specialist. And it is observed that the transfer of expertise betveen specialty

areas is slight. A chess master is unlikely to be an expert algebraist or an expert **mass** spectrum analyst , etc. In this view, the expert is the specialist, with a specialist's knowledge of his area and a specialist's methods and heuristics." (Feigenbaum, **Buc hanan** and Lederberg, 1971, p. 187)

Subsequent evidence from our laboratory and all others has **only** confirmed this belief.

AI researchers have dramatically shifted their view on generality and power in the past decade. In 1967, the canonical question about the DENDRAL program **was** : "It sounds like good chemistry, but **what** does it have to do with AI?' In 1977, Goldstein and **Papert** write of a paradigm shift in AI:

"Today there **has** been a shift in paradigm. The fundamental problem of understanding intelligence is not the identification of a few powerful techniques, but rather the question of how to represent large amounts of knowledge In a fashion that permits their effective use and interaction." (Goldstein and **Papert,** 1977)

The second insight from past **work** concerns the nature of the knowledge that an expert brings to the performance of a task. Experience has shown us that this knowledge is largely heuristic knowledge, experiential, uncertain -- mostly "good guesses" and "good practice," in lieu of facts and rigor. Experience has also taught us that much of this knowledge is private to the expert, not because he is unwilling to share publicly how he performs, but because he **is** unable. He knows more than he is aware of knowing. [Why else is the Ph.D. or the Internship a guild-like apprenticeship to a presumed "master of the craft?" What the masters really know is not written **in the** textbooks of the masters.] But ve have learned also that this private knowledge can be uncovered by the careful, painstaking analysis of a second party, or sometimes by the expert himself, operating In the context of a large number of highly specific performance problems. Finally, we have learned that expertise is **multi-** faceted, that the expert brings to bear many and var led sources of knowledge in performance. The approach to capturing his expertise must proceed **on** many fronts simultaneously.

2.2 The Knowledge Engineer

The knowledge engineer is that second party jus t discussed. [An historical note about the term. In the **mid-60s,** John McCarthy, for reasons obvious from his work, had been describing Artificial Intelligence as "Applied Epistemology.' When I first described the DENDRAL program to Donald **Michie** in 1968, he remarked that it was "epistemological engineering **,"** a clever but ponderous and unpronounceable turn-of-phrase that I simplified into "knowledge engineering."] She (in deference to **my** favorite knowledge engineer) works intensively with an expert to acquire domain-specific knowledge and organize it for use by a program. Simultaneously she is matching the tools of the AI workbench to the task at hand -- program organizations, methods of symbolic inference, techniques for the structuring of symbolic information, and the like. If the tool fits, or nearly fits, she uses it. If not, necessity mothers AI invention, and a new tool gets created. She builds the early versions of the intelligent agent, guided always by her intent that the program eventually achieve expert levels of performance In the task. She refines or reconceptualizes the system as the increasing amount of acquired knowledge causes the AI tool to "break" or slow down intolerably. She also refines the human interface to **the** intelligent agent with several aims: to make the system appear "comfortable" to the human user in his linguistic transactions with it; to make **the system's** inference processes understandable to the user; and to make the assistance controllable by the user when, in the context of a real problem, he has an Insight that previously was not elicited and therefore not incorporated.

In the next section, I wish to explore (in summary form) some case studies of the knowledge engineer's art.

3 CASES FROM THE KNOWLEDGE ENGINEER'S WORKSHOP

I will draw material for this section from the work of my group at Stanford. Much exciting **work** in knowledge engineering is going on elsewhere. Since my Intent is not to survey literature but to Illustrate themes, at the risk of appearing parochial I have used as case studies the work I know best.

My collaborators (Professors Lederberg and **Buc hanan)** and I began a series of projects, Initially the development of the DENDRAL program, in 1965. We had dual motives: first, to study scientific problem so **lv ing** and discovery, particularly the processes scientists do use or should use in inferring hypotheses and theories from empirical evidence; and second, to conduct this study in **such a** way that our experimental programs would one day be of use to working scientists, providing intelligent assistance on important and difficult problems. By 1970, we and

4

our co-workers had gained enough experience that we felt comfortable in laying out a program of research encompassing work on theory formation, knowledge utilization, knowledge acquisition, explanation, and knowledge engineering techniques. Although there were **some** surprises along the **way** (like the AM program), the general lines of **the** research are proceeding as envisioned.

THEMES

As a road map to **these** case studies, **it is** useful to keep in mind certain major themes:

Generation-and-test: Omnipresent in our experiments is the "classical" **generation-and-** test framework that has been the hallmark of AI programs for two decades. This is not a consequence of a doctrinaire attitude on our part about heuristic search, but rather of the usefulness and sufficiency of the concept.

Situation **=>** Action Rules: We have chosen to represent the knowledge of experts in this form. Making no doctrinaire claims for the universal applicability of ~**this** representation, we nonetheless point to the demonstrated utility of **the** rule-based representation. From this representation flow rather directly many of the characteristics of our programs: for example, ease of **modification of** the knowledge, ease of explanation. The essence of our approach is that a rule must capture a "chunk" of domain knowledge **that** is meaningful, in and of itself, to the domain specialist. Thus our rules bear only a historical relationship to the production rules used by Newell and Simon (1972) which **we** view as "machine-language programming" of a recognize **=>** act machine.

The Domain-Specific Knowledge: It plays a critical role in organizing and constraining search. The theme is that in the knowledge is the power. The interesting action arises from the knowledge base, not the inference engine. We use knowledge in rule form (discussed above), in the form of inferentially-rich models based on theory, and in the form of tableaus of symbolic data and relationships **(i.e.** frame-like structures). System processes are made to conform to natural and convenient representations of tbe domain-_ specific knowledge.

-Flexibility to modify the knowledge base: If the so-called "grain site" of the knowledge representation is chosen properly (i.e. small enough to be comprehensible but large enough to be meaningful to the domain specialist), then the rule-based approach **allows** great flexibility for adding, removing, or changing knowledge In the system.

Line-of-reasonfng: A central organizing principle in the design of knowledge-based Intelligent agents is the maintenance of a line-of-reasoning that is comprehensible to the domain specialist.

This principle is, of course, not a logical necessity, but seems to us to be an engineering principle of major importance.

Multiple Sources of Knowledge: The formation and maintenance (support) of the line-of-reasoning usually require the integration of many disparate sources of knowledge. **The** representational and inferential problems in achieving a smooth and effective integration are formidable engineering problems.

Explanation: The ability to explain the **line-of-** reasoning in a language convenient to the user is necessary for application and for system development **(e.g.** for debugging and for extending the knowledge base). Once again, **this** is **an** engineering principle, but very important. What constitutes "an explanation" is not a simple concept, and considerable thought needs to be given, in each case, to the structuring of explanations.

CASE STUDIES

In this section I will try to illustrate these themes with various case studies.

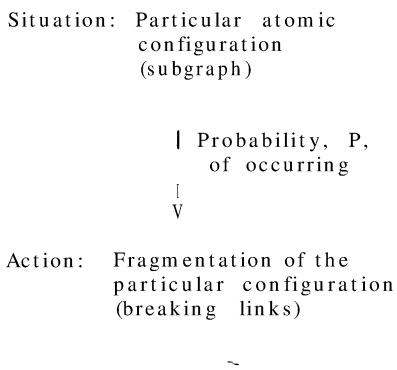### 3.1 DENDRAL: Inferring Chemical Structures

3.1.1 Historical Note

Begun in 1965, this collaborative project with the Stanford Mass Spectrometry Laboratory has become one of the longest-lived continuous efforts in the history of AI (a fact that in no small way has contributed to its success). The basic framework of generation-and-test and rule-based representation has proved rugged and extendable. For us the DENDRAL system has been a fountain of ideas, many of which have found their way, highly metamorphosed, into our other projects. For example, our long-standing commitment to **rule-** based representations arose out of our (successful) attempt to head off the imminent ossification of **DENDRAL** caused by the rapid accumulation of new knowledge in the system around 1967.

3.1.2 Task

To enumerate plausible structures (atom-bond graphs) for organic molecules, given two kinds of information: analytic instrument data from a **mass** spectrometer and a **nuc**lear magnetic resonance spectrometer; and user-supplied constraints on the answers, derived from **any** other source of knowledge (instrumental or contextual) available to the **user.**

3.1.3 <u>Representations</u>

Chemical structures are represented as node-link graphs of atoms (nodes) and bonds (links). Constraints on search are represented as subgraphs (atomic configurations) to be denied or preferred. The empirical theory of mass spectrometry is represented by a set of rules of the general form:

> Situation: Particular atomic
>   configuration
>   (subgraph)
>
> | Probability, P,
>   of occurring
> |
>   V
>
> Action: Fragmentation of the
>   particular configuration
>   (breaking links)

Rules of this form are natural and expressive to mass spectrometrists.

3.1.4 <u>Sketch of Method</u>

DENDRAL's inference procedure is a heuristic search that takes place in three stages, without feedback: plan-generate-test.

"Generate" (a program called CONGEN) is a generation process for plausible structures. Its foundation is a combinatorial algorithm (with mathematically proven properties of completeness and non-redundant generation) that can produce all the topological ly legal candidate **struc**tures. Constraints supplied by the user or by the "Plan" process prune and steer the generation to produce the plausible set (i.e. those satisfying the constraints) and not the enormous legal set.

"Test" refines the evaluation of plausibility, discarding less worthy candidates and rank-ordering the remainder for examination by the user. "Test " first produces a "predicted" set of instrument data for each plausible candidate, using the rules described. It then evaluates the worth of each candidate by comparing its predicted data with the actual input data. The evaluation is based on heuristic criteria of goodness-of-fit. Thus, "test" selects the "best" explanations of the data.

"Plan" produces direct (**i.e.** not chained) inference about likely substructure in the molecule from patterns in the data that are indicative of the presence of the substructure. (Patterns in the data trigger the left-hand-sides

of substructure rules). Though composed of many atoms whose interconnections are given, the substructure can be manipulated as atom-like by "generate." Aggregating many units entering into a combinatorial process into fewer higher-level units reduces the size of the combinatorial search space. "Plan" sets up the search space so as to be relevent to the input data. "Generate is the inference tactician; "Plan" is the inference strategist. There is a separate "Plan" package for each type of instrument data, but each package passes substructures (subgraphs) to "Generate." Thus, there is a uniform interface between **"Plan"** and "Generate." User-supplied constraints enter this interface, directly or from user-assist packages, in the form of substructures.

3.1.5 <u>Sources of Knowledge</u>

The various sources of knowledge used by the DENDRAL system are:

Valences (legal connections of atoms); stable and unstable configurations of atoms; rules for mass spectrometry fragmentations; rules for **NMR** shifts; expert's rules for planning and evaluation; user-supplied constraints (contextual).

3.1.6 <u>Results</u>

DENDRAL's structure elucidation abilities are, paradoxically, both very general and very narrow. In general, DENDRAL handles all molecules, cyclic and tree-like. In pure structure elucidation under constraints (without instrument **data)**,**CONGEN** is unrivaled by human performance. In **struc**ture elucidation with instrument data, DENDRAL 's performance rivals expert human performance only for a small number of molecular families for which the program has been given specialist's knowledge, namely the families of interest to our chemist collaborators. I will spare this computer science audience the list of names of these families. Within these areas of knowledge-intensive **specialization,** DENDRAL 's performance is usually not only much faster but also more accurate than expert human performance.

The statement just made summarizes thousands of runs of DENDRAL on problems of interest to our experts, their colleagues, and their students. The results obtained, along with the knowledge that had to be given to DENDRAL to obtain them, are published in major journals of chemistry. To date, 25 papers have been published there, under a series title "Applications of Artificial Intelligence for Chemical Inference: **<specif**ic subject>" (see references).

The DENDRAL system is in everyday use by Stanford chemists, their collaborators at other universities and collaborating or otherwise interested chemists in industry. Users outside

Stanford access the system over commercial computer/communications network. The problems they are solving are often difficult and novel. The British government is currently supporting work at Edinburgh aimed at transferring DENDRAL to industrial user communities in the UK.

3.1.7 Discussion

Representation and extensibility. The representation chosen for the molecules, constraints, and rules of instrument data interpretation is sufficiently close to that used by chemists in thinking about structure elucidation that the knowledge base has been extended smoothly and easily, mostly by chemists themselves in recent years. Only one major reprogramming effort took place in the last 9 years -- when a new generator was created to deal with cyclic structures.

Representation and the Integration of multiple sources of knowledge. The generally difficult problem of integrating various sources of knowledge has been made easy in DENDRAL by careful engineering of the representations of objects, constraints, and rules. We insisted on a common language of compatibility of the representations with each other and with the inference processes: the language of molecular structure expressed as graphs. This leads to a straightforward procedure for adding a new source of knowledge, say, for example, the knowledge associated with a new type of instrument data. The procedure is this: write rules that describe the effect of the physical processes of the instrument on molecules using the situation ▪> action form with molecular graphs on both sides; any special inference process using these rules must pass its results to the generator only(!) in the common graph language.

It is today widely believed in AI that the use of many diverse sources of knowledge in problem solving and data interpretation has a strong effect on quality of performance. How strong is, of course, domain-dependent, but the impact of bringing just one additional source of knowledge to bear on a problem can be startling. In one difficult (but not unusually difficult) mass spectrum analysis problem*, the program using its mass spectrometry knowledge alone would have generated an impossibly large set of plausible candidates (over 1.25 million!). Our engineering response to this was to add another source of data and knowledge, proton NMR. The addition on a simple interpretive theory of this NMR data, from which the program could infer a few additional constraints, reduced the set of plausible candidates to one, the right structure! This was not an isolated result but showed up dozens of times in subsequent analyses.

----

* the analysis of an acyclic amine with formula C20H45N.

DENDRAL and data. DENDRAL's robust models (topological, chemical, instrumental) permit a strategy of finding solutions by generating hypothetical "correct answers" and choosing among these with critical tests. This strategy is opposite to that of piecing together the implications of each data point to form a hypothesis. We call DENDRAL's strategy largely model-driven , and the other data-driven. The consequence of having enough knowledge to do model-driven analysis is a large reduction in the amount of data that must be examined since data is being used mostly for verification of possible answers . In a typical DENDRAL mass spectrum analysis, usually no more than about 25 data points out of a typical total of 250 points are processed. This important point about data reduction and focus-of-attention has been discussed before by Gregory (1968) and by the vision and speech research groups, but is not widely understood.

Conclusion. DENDRAL was an early herald of AI's shift to the knowledge-based paradigm. It demonstrated the point of the primacy of domain-specific knowledge in achieving expert levels of performance. Its developme t brought to the surface important problems of knowl edge representation, acquisition, and use. It showed that, by and large, the AI tools of the first decade were sufficient to cope with the demands of a complex scientific problem-solving task,or were readily extended to handle unforseen difficulties. It demonstrated that AI's conceptual and programming tools were capable of producing programs of applications interest, albeit in narrow specialties. Such a demonstration of competence and sufficiency was important for the credibility of the AI field at a critical juncture in its history.

3.2 META-DENDRAL: inferring rules of mass spectrometry

3.2.1 Historical note

The META-DENDRAL program is a case study in automatic acquisition of domain knowledge. It arose out of our DENDRAL work for two reasons: first, a decision that with DENDRAL we had a sufficiently firm foundation on which to pursue our long-standing interest in processes of scientific theory formation; second, by a recognition that the acquisition of domain knowledge was the bottleneck problem in the building of applications-oriented intelligent agents.

3.2.2 Task

META-DENDRAL's job is to infer rules of fragmentation of molecules in a mass spectrometer for possible later use by the DENDRAL performance

program. The Inference is to be made from actual spectra recorded from known molecular structures. The output of the system is the set of fragmentation rules discovered, summary of the evidence supporting each rule, and a summary of contra-indicating evidence. User-supplied constraints can also be input to force the form of rules along desired lines.

3.2.3 Representations

The rules are, of course, of the same form as used by DENDRAL that was described earlier.

3.2.4 Sketch of Method

MRTA-DENDRAL, like DENDRAL, **uses the** generation-and-test framework. The process is organ1 zed in three stages: Reinterpret the data and summarize evidence (INTSUM); generate plausible candidates for rules (RULEGEN); test and refine **the** set of plausible rules **(RULEMOD).**

INTSUM: gives every data point in every spectrum an interpretation as a possible (highly specific) fragmentation. It then **summar i ze**s statistically the "weight of evidence" for fragmentations and for atomic configurations that cause these fragmentations. Thus, the job of INTSUM is to translate data to DENDRAL subgraphs and bond-breaks, and to summarize the evidence accordingly.

RULEGEN: conducts a heuristic search of the space of all rules that are legal under the DENDRAL rule syntax and **the** user-supplied constraints. It searches for plausible rules, **i.e.** **those** for which positive evidence exists. A search path is pruned when there is no evidence for rules of the class just generated. The search tree begins with **the** (single) most general rule (loosely **put,** "anything" fragments from "anything") and proceeds level-by-level toward **mare** detailed specifications of the "anything." The heuristic stopping criterion measures whether a rule being generated has become too specific, in particular whether it is applicable to too few molecules of **the** input set. Similarly there is a criterion for deciding whether an emerging rule is too general. Thus, the output of RULEGEN is a set of **candidate** rules for which there is positive evidence.

RULEMOD: tests **the** candidate rule set using more **complex** criteria, including the presence of negative evidence. It removes redundancies in the candidate rule set; merges rules that are supported by the same evidence; tries further specialization of candidates to remove negative evidence; and tries further generalization that preserves positive evidence.

3.2.5 Results

META-DENDRAL produces rule sets that rival in quality those produced by our collaborating experts. In **some tests,** META-DENDRAL recreated rule, sets that we had previously acquired from our experts during the DENDRAL project. In a more stringent test involving members of a family of complex ringed molecules for which the mass spectral theory had not been completely worked out by chemists, META-DENDRAL discovered rule sets for **each** subfamily. The rules were judged by experts to be excellent and a paper describing them was recently published in a major chemical journal (Buchanan, Smith, et al, 1976).

In a test of **the** generality of the approach, a version of the META-DENDRAL program is currently being applied to **the** discovery of rules for the analysis of nuclear magnetic resonance data.

3.3 MYCIN and TEIRESIAS: Medical Diagnosis

3.3.1 Historical note

MYCIN originated in the Ph.D. thesis of **E.** Shortliffe (now Shortliffe, M.D. as well), in collaboration with the Infectious Disease group at the Stanford Medical School (Shortliffe, 1976). TEIRESIAS, **the** Ph.D. thesis work of **R.** Davis, arose from issues and problems indicated by the MYCIN project but generalized by Davis beyond **the** bounds of medical diagnosis applications (Davis, 1976). Other MYCIN-related theses are in progress.

3.3.2 Tasks

The MYCIN performance task is diagnosis of blood infections and meningitis infections and the recommendation of drug treatment. MYCIN conducts a consultation (in English) with a physician-user about a patient case, constructing **lines-of-** reasoning leading to the diagnosis and treatment plan.

The TEIRESIAS knowledge acquisition task can **be** d **escribed** as follows :

In the context of a particular consultation, confront the expert with a diagnosis with which **he** does not agree. Lead him systematically back through the line-of-reasoning that produced **the** diagnosis to the point at which he indicates the analysis went awry. Interact with the expert to modify offending rules or to acquire new rules. Rerun the consultation to test the solution and gain the expert's concurrence.

3.3.3 Representations:

MYCIN's rules are of the form:

IF <conjunctive clauses> THEN <implication>

Here is an example of a MYCIN rule for blood infections.

RULE 85

IF:
1) The site of the culture is blood, and
2) The gram stain of the organism is gramneg, and
3) The morphology of the organism is rod, and
4) The patient is a compromised host

THEN:
There is suggestive evidence (.6) that the identity of the organism is pseudomonas-aeruginosa

----

TEIRESIAS allows the representation of MYCIN-like rules governing the use of other rules, i.e. rule-based strategies. An example follows.

METARULE 2

IF:
1) the patient is a compromised host, and
2) there are rules which mention in their premise pseudomonas
3) there are rules which mention in their premise klebsiellas

THEN:
There is suggestive evidence (.4) that the former should be done before the latter.

3.3.4 Sketch of method

MYCIN employs a generation-and-test procedure of a familiar sort. The generation of steps in the line-of-reasoning is accomplished by backward chaining of the rules. An IF-side clause is either immediately true or false (as determined by patient or test data entered by the physician in the consultation); or is to be decided by subgoaling. Thus, "test" is interleaved with 'generation" and serves to prune out incorrect lines-of-reasoning.

Each rule supplied by an expert has associated with it a "degree of certainty" representing the expert's confidence in the validity of the rule (a number from 1 to 10). MYCIN uses a particular ad-hoc but simple model of inexact reasoning to cumulate the degrees of certainty of the rules used in an inference chain (Shortliffe and Buchanan, 1975).

It follows that there may be a number of "somewhat true" lines-of-reasoning -- some indicating one diagnosis, some indicating another. All (above a threshold) are used by the system as sources of knowledge indicating plausible lines-of-reasoning.

TEIRESIAS' rule acquisition process is based on a record of MYCIN's search. Rule acquisition is guided by a set of rule models that dictate the form and indicate the likely content of new rules. Rule models are not given in advance, but are inferred from the knowledge base of existing rules.

3.3.5 Language of Interaction

The language used looks like it might be English but is actually the dialect "Doctor-ese" used by members of the tribe of healing arts practitioners. Reasonably simple language processing methods suffice. When ambiguities of interpretation are encountered, they are fed back to the user for decisions about meaning.

3.3.6 Example diagnosis and treatment plan

The following is an example of a diagnosis reported by MYCIN to one of its users following a consultation. An example consultation will not be shown, but one can be found in Shortliffe's book (Shortliffe, 1976).

My therapy recommendations will be designed to treat for organisms that are either very likely or, although less likely, would have a significant effect on therapy selection if they were present. It is important to

cover for the following probable
Infection(s) and associated organ l **sm(s):**

INFECTION-1 is CYSTITIS
<Item 1> **PSEUDOMONAS-CEPACIA [ORGANISM-5]**
<Item 2> **CITROBACTER-DIVERSUS** (ORGANISM-4 l
<Item 3> **E.COLI [ORGANISM-3]**

INFECTION-2 is UPPER-RESPIRATORY-INFECTION
<Item 4> LISTERIA [ORGANISM-2 l

INFECTION-3 is BACTEREMIA
<I tern 5> ENTEROCOCCUS **[ORGANISM-1]**

[REC-1] My preferred therapy recommendation
is as follows:

In order to cover for **Items <1 2 3 4 5>:**
Give **the** following in combination:
1) KANAMYCIN
Dose: 750 **mg** (7.5 **mg/kg) q12h IM** (or IV)
    for 28 days
Comments: Modify dose **in** renal failure
2) PENICILLIN
Dose: **2,500,000** units (25000 units/kg)
    q4h IV for 28 days

3.3.7 Results

MYCIN is currently being prepared for
clinical evaluation. In a preliminary evaluation,
done outside of **the** clinical setting, a panel of
expert judges was asked to rate **MYCIN's**
performance along a number of dimensions (e.g.
infecting organism identification, organism
significance). In 90% of the cases submitted to
the judges, a majority of the judges said **that the**
program's decisions were the-same-as or as-good-as
the decisions **they** would have made.

3.3.8 Explanation

- MYCIN can **expose** its line of reasoning by
offering explanations of various kinds. "Why" and
"How" questions can be asked by the user during
the consultation. **"Why"** is the user's way of
inquiring, "Why do you want to know **that?" "How"**
is his way of asking, "How did you conclude
that...?" In **a post-consultation, he can** similarly
inquire about how the diagnostic conclusions were
reached, and inquire about **the** reasons why certain
other conclusions were ruled out. Examples of the
MYCIN-TEIRESIAS explanation capabilities follow:

MYCIN: Is the patient's illness with
ORGANISM-l a hospital-acquired infection?
USER: WHY
TEIRESIAS: **(i.e.** WHY is it important to
determine whether or not the infection with
ORGANISM-l was acquired while the patient
was **hospitalized?)**

[1.0] This will aid in determining the
category of ORGANISM-l.

It has already been established that

[1.1] the gram stain of ORGANISM-l is
g ramneg , and
[1.2] the morphology of ORGANISM-l is rod,
and
[1.3] the aerobicity of ORGANISM-l is
facultative

Therefore, if

[1.4] the infection with ORGANISM-l was not
acqui red while the patient was hospitalized

then

there is strongly **suggestive** evidence **(.8)**
that **the** category of ORGANISM-l is
enterobacteriaceae

Also: there is **weakly** suggestive evidence
(.1) that the identity of
ORGANISM-l is pseudomonas-aeruginosa
[RULE 37]

USER: HOW 1.3

TEIRESIAS: [i.e. HOW was it established
that the aerobicity of ORGANISM-l is
facultative?]

The following were used:

[3.1] RULE 27 indicated there is strongly
suggestive evidence **(.8)** that **the**
aerobicity of ORGANISM-l is facultative

[3.2] RULE 27 indicated there is weakly
suggestive evidence **(.2)** that the
aerobicity of ORGANISM-l is anaerobic

Since this gave a cumulative CF of **(.8)** for
facultative, and **(.2)** for anaerobic, it has
been established that the aerobicity of
ORGANISM-l is facultative.

USER: WHY DIDN'T YOU CONSIDER
STREPTOCOCCUS AS A POSSIBILITY FOR
ORGANISM-1?

TEIRESIAS: The following rule could have been used to determine that the identify of ORGANISM-l was streptococcus: RULE 33

But clause 2 ("the morphology of the organism is coccus") was already known to be false for ORGANISM-I, so the rule was never tried.

3.3.9 Discussion

A few short comments aimed at knowledge engineering are in order.

English is thalanguage most natural and comfortable for our experts and users, and was therefore the language chosen for interactive consultation, explanation, and external representation of the rules (the internal format is INTERLISP). This situation is not peculiar to doctors; in most areas of application of intelligent agents I believe that English (i.e. natural language) will be the language of choice. Programming an English language processor and front-end to such sys tems is not a scary enterprise because:

a) the domain is specialized, so that possible interpretati one are constrained

b) specialist-talk is replete with standard Jargon and stereotyped ways of expressing knowledge and queries -- just right for text templates, simple grammars and other simple processing schemes.

c) the ambiguity of interpretation resulting from simple schemes can be dealt with easily by feeding back Interpretations for confirmation. If this is done with a pleasant "I didn't quite understand you.. ." tone, it is not irritating to the user.

English may be exactly the wrong language for representation and interaction in some domains. It would be awkward, to say the least, to represent DENDRAL's chemical structures and knowledge of mass spectronetry in English, or to interact about these with a user.

Simple explanation schemes have been a part of the AI scene for a number of years and are not hard to implement. Really good models of what explanation is as a transaction between user and agent, with programs to implement these models, will be the subject (I predict) of much future research in AI.

Without the explanation capability' I assert, user acceptance of MYCIN would have been nil, and there would have been a greatly diminished effectiveness and contribution of our experts.

HYCIN was the first of our programs that forced us to deal with what we had always understood: that experts' knowledge is uncertain and that our inference engines had to be made to reason with this uncertainty. It is less important that the inexact reasoning scheme be formal, rigorous, and uniform than it is for the scheme to be naturnl to and easily understandable by the experts and users.

All of these points can be summarized by saying that MYCIN and its TEIRESIAS adjunct are experiments in the design of a see-through system, whose representations and processes are almost transparently clear to the domain specialist. "Almost" here is equivalent to "with a few minutes of introductory description." The various pieces of MYCIN -- the backward chaining, the English transactions' the explanations, etc. -- are each simple in concept and realization. But there are great virtues to simplicity in system design; and viewed as a total intelligent agent system, MYCIN/TEIRESIAS is one of the best engineered.

3.4 SU/X: signal understanding

3.4.1 Historical note

SU/X is a system design that was tested in an application whose details are classified. Because of this, the ensuing discussion will appear considerably less concrete and tangible than the preceding case studies. This system design vas done by H.P. Nil and me, and was strongly influenced by the CMU Hearsay II system design.

3.4.2 Task

SU/X's task is the formation and continual updating, over long periods of time, of hypotheses about the identity, location, and velocity of objects in a physical space. The output desired is a display of the "current best hypotheses" with full explanation of the support for each. There are two types of input data: the primary signal (to be understood); and auxiliary symbolic data (to supply context for the understanding). The primary signals are spectra, represented as descriptions of the spectral lines. The various spectra cover the physical space with some spatial overlap.

3.4.3 Representations

The rules given by the expert about objects, their behavior, and the interpretation of signal data from then are all represented in the situation => action form. The "situations" constitute invoking conditions and the "actions" are processes that modify the current hypotheses, post unresolved issues, recompute evaluations, etc. The expert's knowledge of how to do analysis in the task is also represented in rule form. These strategy rules replace the normal executive program.

The situation-hypothesis is represented as a node-link graph, tree-like In that it has distinct "levels ," each representing a degree of abstraction (or aggregation) that is natural to the expert in his understanding of the domain. A node represents an hypothesis; a link to that node represents support for that hypothesis (as in HEARSAY II, "support from above" or **"suppor** t from below"). "Lower" levels are concerned with the **specif ics** of the signal data. "Higher" levels represent symbolic abstractions.

3.4.4 Sketch of method ~.

The situation-hypothesis is formed incrementally. As the situation unfolds over time, the triggering of rules modifies or discards existing hypotheses, adds new ones, or changes support values. The situation-hypothesis is a common workspace ("blackboard," in HEARSAY jargon) for all the rules.

In general , the incremental steps toward a more complete and refined situation-hypothesis can be viewed as a sequence of local generate-and-test activities. Some of the rules are plausible move generators, generating either nodes or links. Other rules are evaluators, testing and modifying node descriptions.

In typical operation, new data is submitted for processing (say, N time-units of new data). This initiates a flurry of rule-triggerings and consequently rule-actions (called "events"). Some events are direct consequences of the data; other events arise in a cascade-like fashion from the triggering of rules. Auxiliary symbolic data also cause events, usually affecting the higher levels of the hypothesis. As a consequence, **support-** from-above for the lower level processes is made **avaflable;** and expectations of **possible** lower level events can be formed. Eventually all the relevant rules have their say and the system becomes quiescent, thereby triggering the input of new data to re-energize the inference activity.

The system uses the simplifying strategy of maintaining only one "best'* situation-hypothesis at **any** moment, modifying l t incrementally as **required** by the changing data. **This** approach is made **feasible** by several characteristics of the

domain. First, there is the strong continuity over time of objects and their behaviors (specifically, they do not change radically over time, or behave radically differently over short periods). Second, a single problem (identity, location and velocity of a particular set of objects) persists over numerous data gathering periods. (Compare this to speech understanding in ..which each sentence is spoken just once, and each presents a new and different problem.) Finally, the system's hypothesis is typically "almost right ," in part because it gets numerous opportunl t ies to refine the solution (i.e. the numerous data gathering periods), and in part because the availability of many knowledge sources tends to over-determine the solution. As a result of all of these , the current best hypothesis changes only slowly with time, and hence keeping only the current best is a feasible approach.

Of interest are the time-based events. These rule-like expressions, created by certain rules, trigger upon the passage of specified amounts of time. They implement various "wait **-and-see"** strategies of analysis that are useful in the domain.

3.4.5 Results

In the test application, using signal data generated by a simulation program because real data was not available, the program achieved expert level5 of performance over a span of test problems. Some problems were difficult because there was very little primary signal to support inference. Others were difficult because too much signal induced a plethora of alternatives with much ambiguity.

A modified **SU/X** design is currently being used as the basis for an application to the Interpretation of x-ray crystallographic data, the CRYSALIS program mentioned later.

3.4.6 Discussion

The role of the auxiliary symbolic sources of data is of critical importance. They **supply** a symbolic model of the existing situation that is used to generate expectations of events to be observed in the data stream. This allows flow of inferences from higher levels of abstractfon to lower. Such a process, so familiar to AI researchers, apparently is almost unrecognized among signal processing engineers. In the application task, the expectation-driven analysis **is** essential In controlling the combinatorial processing explosion at the lower levels,exactly the explosion that forces the traditional signal processing engineers to seek out the largest possible number-cruncher for their work.

The design of appropriate explanations for the user takes an interesting twist in **SU/X.** The

situation-hypothesis unfolds piecemeal over time, but the "appropriate" explanation for the user **is** one that focuses on Individual objects over time. Thus the appropriate explanation must be **synthesized** from a history of all the events that led up to the current hypothesis. Contrast this. with the MYCIN-TEIRESIAS reporting of rule Invocations in the construction of a reasoning chain.

Since its knowledge base and its auxiliary symbolic data give it a model-of-the-situation that strongly constrains interpretation of the primary data stream, **SU/X** is relatively unperturbed by errorful or missing data. These data conditions merely cause fluctuations in the credibility of individual hypotheses and/or the creation of the "wait-and-see" events. **SU/X** can be (but has not yet been) **used** to control sensors. Since Its rules specify what types and values of evidence are necessary to **establish** support, and since it is constantly processing a complete hypothesis structure, it can requeat "critical readings" from the sensors. In general, this al lows an efficient use of **limited** sensor bandwld th and data acquisition processing capability.

3.5 OTHER CASE STUDIES

Space does not allow more than just a brlef sketch of other intereating **projects** that have been completed or are in progress*

3.5.1 AM: mathematical discovery

AM is a knowledge-based **sys**tern that conjecture5 interesting concept5 in elementary mathematics. It **is** a discoverer of Interesting theorems to prove, not a theorem proving program. It was conceived and executed by **D. Lenat** for **his** Ph.D. thesis, and is reported by him in these proceedings ("An Overview of AM").

AM's knowledge **is** basically of two types: rules that suggest possibly interesting new concepts from previously conjectured concepts; and rule5 that evaluate the mathematical "interestingness" of a conjecture. These rules attempt to capture the expertise of the **professional** mathematician at the task of mathematical discovery. Though **Lenat** is not a professional mathematician, he was able successfully to serve as his own expert In the building of this program.

AM conducts a heuristic search through the space of concepts **creatable** from Its rules. Its **bas**ic framework is generation-and-test. The generation is plausible move generation, as indicated by the rules for formation of new concepts. The test is the evaluation of "interestingness." Of particular note **is** the method of tent-by-example that lends the flavor of

scientific **hypothesis** testing to the enterprise of mathematical discovery.

Initialized with concepts of elementary set theory, it conjectured concepts in elementary number theory, such as "add," "multiply" (by four distinct paths!), "primes **,"** the unique factorization theorem, and a concept similar to primes but previourly not much studied called "maximally divisible numbers."

3.5.2 MOLGEN: planning experiments In molecular genetics

MOLGEN , a collaboration with the Stanford Genetics Department, is work in progress. **MOLGEN's** task is to provide Intelligent advice to a molecular geneticist on the planning of experiments involving the manipulation of DNA. The geneticist has various kinds of laboratory techniques available for changing DNA material (cuts, joins, insertions, deletions, and 50 on); techniques for determining the biological consequences of the changes ; various instruments for measuring effects; various chemical methods for inducing, facilitating, or inhibiting changes; and many other tools.

HOLGEN will offer planning assistance in organizing and sequencing such tools to accomplish an experimental goal. In addition **MOLGEN** will check user-provided experiment plan5 for feasibility; and Its knowledge base will be a repository for the rapidly expanding knowledge of this specialty, available by Interrogation.

Current efforts to engineer a knowledge-base management system for MOLGEN are described by **Martin** et al in a paper in these proceedings. This subsystem uses and extends the techniques of the TEIRESIAS system discussed earlier.

In MOLGEN the problem of Integration of many diverse **sources** of knowledge **is** central since the essence of the experiment planning process is the successful merging of biological, genetic, chemical, topological, and Instrument knowledge. In MOLGEN the problem of representing processes is also brought into focus since the expert's knowledge of experimental **strategies** - - **proto**plans -- must also be represented and put to use.

3.5.3 CRYSALIS: inferring protein structure from electron density **maps**

CRYSALIS, too, is work in progress. Its task is to hypothesize the structure of a protein from a map of electron density that is derived from x-ray crystallographic data. The map is **three**-dimensional, and the contour information **is** crude and highly ambiguous. Interpretation is guided and supported by auxiliary Information, of which the amino acid sequence of the protein's backbone is the most important. Density map interpretation

is a protein chemist's art. As always, capturing this art in heuristic rules and putting it to use with an Inference engine is the project's goal.

The Inference engine for CRYSALIS is a modification of the **SU/X** system design described above. The hypothesis formation process must deal with many levels of possibly useful aggregation and abstraction. For example, the map itself can be viewed as consisting of "peaks," or "peaks and valleys," or "skeleton." The protein model has "atoms," "amide planes," "amino acid sidechains," and even massive substructures such as **"helices."** Protein molecules are so complex that a systematic generation-and-test strategy like **DENDRAL's** is not feasible. Incremental piecing together of the hypothesis using region-growing methods is necessary.

The CRYSALIS design (alias **SU/P)** is described in a recent paper by **Nii** and Feigenbaum (1977).

4 SUMMARY OF CASE STUDIES

Some of the themes presented earlier need no recapitulation, but I wish to revisit three here: generation-and-test; situation **=>** action rules; and explanations.

'4.1 Generation and Test

Aircraft come in a wide variety of sixes, shapes, and functional designs and they are applied in very many ways. But almost all that fly do so because of the unifyiag physical principle of lift by airflow; the others are described by exception. So it is with intelligent agent programs and, the information processing psychologists tell us, with people. One unifying principle of "intelligence" is **generation-and-test**. No wonder that It has been so thoroughly studied in AI research!

- In the case studies, generation is manifested in a variety of forms and processing schemes. There are legal move generators defined formally by a generating algorithm **(DENDRAL's** graph generating algorithm); or by a logical rule of **inference (MYCIN's** backward chaining). **When** legal- move generation **is** not possible or not efficient, there are plausible move generators (as in **SU/X** and AM). Sometimes generation is interleaved with testing (as in MYCIN, **SU/X,** and AM). In one case, all generation precedes testing (DENDRAL). One case **(META-DENDRAL)** is mixed, with some testing taking place during generation, some after.

Test also shows great variety. There are simple tests (MYCIN: "Is the organism aerobic?"; **SU/X:** "Has a spectral line appeared at position **P?")** Some tests are complex heuristic evaluations (AM: "Is the new concept 'interesting'?"; MOLGEN:

"Will the reaction actually take place?") Sometimes a complex test can involve feedback to modify the object being tested (as in **META-** DENDRAL).

The evidence from our case studies supports the assertion by Newell and Simon that **generation-and-test** is a law of our science (Newell and Simon, 1976).

4.2 Situation **=>** Action rules

Situation **=>** Action rules are used to represent experts' knowledge in all of the case studies. Always the situation part indicates the specific conditions under which the rule is relevant. The action part can be simple (MYCIN: conclude presence of particular organism; DENDRAL: conclude break of particular bond). Or it can be quite complex (MOLGEN: an experiential procedure). The overriding consideration in making design choices is that the rule form chosen be able to represent clearly and directly what the expert wishes to express about the domain. As illustrated, this may necessitate a wide variation in rule syntax and semantics.

From a study of all the projects, a regularity emerges. A salient feature of the Situation **=>** Action rule technique for representing expert's knowledge is the modularity of the knowledge base, with the concomitant flexibility to add or change the knowledge easily as the experts' understanding of the domain changes. Here too one must be pragmatic, not doctrinaire. A technique such as this can not represent modularity of knowledge if that modularity does not exist in the domain. The virtue of this technique is that it serves as a framework for discovering what modularity exists in the domain. Discovery may feed back to cause reformulation of the knowledge toward greater modularity.

Finally, our case studies have shown that strategy knowledge can be captured in rule form. In TEIRESIAS, the metarules capture knowledge of how to deploy domain knowledge; in **SU/X,** the strategy rules represent the experts' knowledge of "how to analyze" in the domain.

4.3 Explanation

Most of the programs, and all of the more recent ones, make available an explanation capability for the user, be he end-user or system developer. Our focus on end-users in applications domains has forced attention to human engineering issues, in particular making the need for the explanation capability imperative.

The Intelligent Agent viewpoint seems to us to demand that the agent be able to explain its activity; else the question arises of **who** is in

control of the agent's activity. The issue is not academic or philosophical. It is an engineering Issue that has arisen in medical and military applications of intelligent agents, and will govern future acceptance of AI work in applications areas. And on the philosophical level one might even argue that there is a moral fmperat ive to provide accurate explanations to end-users whose intuitions about our systems are almost nil.

Finally, the explanation capability is needed as part of the concerted attack on the knowledge acquisition problem. Explanation of the reasoning process is central to the interactive transfer of expertise to the knowledge base, and it is our most powerful tool for the debugging of the knovledge base.

5 EPILOGUE

What ve have learned about knowledge engineering goes beyond what is **discernible** in the behavior of our case study programs. In the next paper of this two-part series, I will raise and discuss many of the general concerns of knowledge engineers, including these:

What constitutes an "application" of AI techniques?

There is a difference between a serious application and an application-flavored toy problem.

What are some criteria for the judicious selection of an **application** of AI techniques?

What are some applications areas worthy of serious attention by knowledge eagiaeers?

For example, applications to sc fence, to signal interpretation, and to human interaction with complex systems.

How to find and fascinate an Expert.

The background and prior training of the expert.

The level of commitment that can be elicited.

Des igaing systerns that "think the way I do."

Sustaining attention **by** quick feedback and incrementa progress.

Focusing attention to data and specific problems.

Providing ways **to** express uncertainty of expert knowledge.

The side benefits to the expert of his investment in the knowledge engineering activity.

Gaining consensus among experts about the **knowledge** of a domain.

The consensus may be a more valuable outcome of the knowledge eagineerfng effort than the building of the program.

Problems faced by knowledge engineers today:

The l.**ack** o f **adequa**te and appropriate computer hardware.

The difficulty of export of systems to end-users, caused by the lack of **properly-**sized and -packaged combinations of hardware and sof tvare

The chronic absence of cumulation of AI techniques in the form of software packages that can achieve wide use.

The shortage of trained knowledge engineers.

The difficulty of obtaining and sustaining funding for interesting knowledge engineering projects.

7 REFERENCES

General

Felgenbaum. **E.A.** "Artificial Intelligence Research: What is It? What has it achieved? Where is it going?," invited paper, Symposium on Artificial Intelligence, Canberra, Australia, 1974.

Goldstein, I. and **S. Papert,** "Artificial Intelligence, **Language,** and the Study of Knowledge," Cognitive Science, **Vol.1, No.1,** 1977.

Gregory, R., **"On** How so Little Information Controls so Much Behavior," Bionics Research Report No. 1, Machine Intelligence Department, University of Edinburgh, 1968.

Newell, A. and H.A. Simon, Human Problem Solving, Prentice-Hall, 1972.

Newell, A. and H.A. Simon, "Computer Science as Empirical Inquiry: Symbols sad Search," **Com ACM,** 19, 3, March, 1976.

DENDRAL and **META-DENDRAL**

Feigenbaum, E.A., Buchanan, B.C. and **J.** Lederberg, "On Generality and Problem Solving: a Case Study Using the DENDRAL Program," Machine Intelligence **6,** Edinburgh Univ. Press, 1971.

Buchanan, B.G., Duffield, A.M. and A.V. Robertson, "An Application of Artificial Intelligence to the Interpretation of Mass Spectra," Mass Spectrometry Techniques and Applications, **G.W.A.** Milne, Ed., John Wiley & Sons, Inc., **p.** 121, 1971.

Michie, **D.** and B.C. Buchanan, "Current Status of the Heuristic DENDRAL Program for Applying Artificial Intelligence to the Interpretation of Mass Spectra," Computers for Spectroscopy, **R.A.G.** Carrington, ed., London: Adam Hilger, 1974.

Buchanan, B.G., "Scientific Theory Formation by Computer," Nato Advanced Study Institutes Series, Series E: Applied Science, **14:515, Noordhoff-** Leyden; 1976.

Buchanan, B.G., Smith, **D.H.,** White, W.C., Gritter, R.J., Feigenbaum, E.A., Lederberg, **J.** and **C.** Djerassi, "Applications of Artificial Intelligence for Chemical Inference XXII. Automatic Rule Formation in Mass Spectrometry by Means of the **Meta-DENDRAL** Program," Journal of the ACS, **98:6168,** 1976.

MYCIN

Shortliffe, E. Computer-based Medical Consultations: MYCIN, New York, Elsevier, 1976.

Davis, **R.,** Buchanan, B.C. and E.H. Shortliffe, 'Production Rules as a Representation for a Knowledge-Based Consultation Program," Artificial Intelligence, 8, 1, February, 1977.

Shortliffe, E.H. and B.G. Buchanan, "A Model of Inexact Reasoning in Medicine," Mathematical Biosciences, **23:351,** 1975.

TEIRESIAS

Davis, R., "Applications of **Meta** Level Knowledge to the Construction, Maintenance and Use of Large Knowledge Bases **,"** Memo HPP-76-7, Stanford Computer Science Department, Stanford, CA, 1976.

Davis, **R.,** "Interactive Transfer of Expertise I: Acquisition of New Inference Rules," these Proceedings.

Davis, **R.** and B.C. Buchanan, **"Meta-Level** Knowledge: Overview and Applications," these Proceedings.

SU/X

Nil, H.P. and **E.A.** Feigenbaum, "Rule Based Understanding of Signals," Proceedings of the Conference on Pattern-Directed Inference Systems, 1977 (forthcoming), also Memo HPP-77-7, Stanford Computer Science Department, Stanford, CA, 1977.

AM

Lenat, D., "AM: An Artificial Intelligence Approach to Discovery in Mathematics as Heuristic Search," Memo HPP-76-8, Stanford Computer Science Department, Stanford, CA, 1976.

MOLGEN

Martin, **N.,** Friedland, P., King, **J.,** and **M.** Stefik, "Knowledge Base Management for Experiment Planning in Molecular Genetics," these Proceedings.

CRYSALIS

Engelmore, **R.** and **H.P.** Nil, "A Knowledge-Based System for the Interpretation of Protein X-Ray Crystallographic Data," Memo HPP-77-2, Department of Computer Science, Stanford, CA, 1977.