ARTIFICIAL INTELLIGENCE RESEARCH AT CARNEGIE-MELLON UNIVERSITY

Edited by Jaime Carbonell

AI research at CMU is closely integrated with other activities in the Computer Science Department, and to a major degree with ongoing research in the Psychology Department. Although there are over 50 faculty, staff and graduate students involved in various aspects of AI research, there is no administratively (or physically) separate AI laboratory. To underscore the interdisciplinary nature of much of our AI research, a significant fraction of the projects listed below are joint ventures between computer science and psychology.

The history of AI research at Carnegie-Mellon goes back twenty-five years. The early work was characterized by a focus on problem solving and by interaction with psychology. With the seventies came a major broadening to perception, starting with speech recognition and then vision. With the recent establishment of the Robotics Institute, additional broadening has occurred. (The work of the CMU Robotics Institute will be described in a separate report). In addition, a recent involvement in expert systems, focussed on industrial tasks is taking place. In consequence, research at CMU currently spans the full range of AI concerns.

We include in this report several recent projects that are not primarily AI research, but where the CMU AI community is rather heavily involved, such as the ZOG project exploring human interaction with computers an SPICE (Scientific Personalized Computing Environment), the Computer Science Department's effort to construct its future computing facilities.

In listing projects, we give (in alphabetical order) the faculty involved, including those in psychology. Though major contributors to the research, we mention graduate students only when appropriate to describing a specific project

Production Systems

John Anderson, Charles Forgy, Pat Langley, John McDermott, Allen Newell, Mike Rychenei, Herb Simon, Robert Thibadeau

Production systems have been studied at CMU for almost a decade. Interest in production systems first arose here because they presented an attractive model for the basic architecture of the human mind. However, as experience with production sytems was gained it became apparent that they could also serve as general AI programming languages. In recent years, both the view of production systems as cognitive models and the view of production systems as programming languages have been important in AI work at CMU.

The first production systems used at CMU were Newell's PS and PSG systems. Rychener developed his PSNLST system soon after PSG came into use. In 1975, Forgy, McDermott, Newell, and Rychener developed the first system in the OPS family. OPS has continued to evolve since that time, the latest version, OPS5, having been completed in 1979. OPS has a served as a basis for many programs, including Langley and Simon's BACON systems and McDermott's R1 expert system.

During the last few years, many production systems have been developed to serve as psychological models. (It is characteristic of psychological modeling that many systems must be implemented, for the testing of each new architectural feature requires that a system exhibiting that feature be implemented and tested.) Anderson combines production rules with a semantic-network memory in his ACT systems to model human memory phenomena. Recently, Langley implemented ACTG, a version of Anderson's model, which uses some of the implementation techniques of OPS. Robert Neches used ACTG to implement NEWHPM, a production system to simulate learning processes. It provides mechanisms for accessing production traces, performing goal-driven spreading activation of memory, and applying user-defined conflict resolution mechanisms. Further exploration is occuring. Rosenbloom has implemented a production system to model motor skills; this system, XAPS, fires multiple productions on each interpreter cycle. Thibadeau had develoed CAPS for use in a project that models eye-gaze durations in reading; CAPS is like OPS4 except that it relies on spreading activation and it allows multiple productions to fire on each cycle.

Expert Systems

Charles Forgy, John McDermott, Barbara Steele

As described above, there has been a long-standing effort at CMU to develop a highly efficient, general-purpose production system language. There now is such a language, OPS5, which we believe is a suitable tool for building large knowledge-based systems in a variety of domains. Two years ago McDermott began work on R1, a production system that configures VAX-11/780 computer systems. This system, which currently has a knowledge-base of almost 1000 rules, is being used on a regular basis by Digital Equipment Corporation's manufacturing organization.

Current expert-systems research has two thrusts. (1) We are using R1 as the core of two new systems: XSEL and XCITE. XSEL is a sales assistant; it will help a customer identify the computer system that best suits his needs. XCITE is a co-ordinator; it will provide assistance in managing the assembly and distribution of systems.

While XSEL and XCITE are being developed at CMU, the group at DEC responsible for the maintenance and development of R1 will extend R1's knowledge so that it can configure other computer systems manufactured by DEC. (2) In order to test the versatility of OPS5, we plan to develop a number of other expert systems in a variety of task domains. We are currently working on a system, in conjunction with Oak Ridge National Laboratory, that monitors the reprocessing of nuclear fuel. This system will use its knowledge of the process together with information from approximately 100 sensors to determine whether fuel is being diverted during reprocessing. In the near future, we will begin work on a system that assists in computer fault diagnosis.

Automated Knowledge and Skill Acquisition

John Anderson, Jaime Carbonell, Mark Fox, Martin Herman, Glenn Iba, Pat Langley, Raj Reddy, Herb Simon

Machine learning has experienced a vigorous resurgence, since interest waned after early AI work on perceptrons and neural nets CMU is heavily involved in this effort, with multiple research projects.

Pat Langley, Gary Bradshaw and Herb Simon have developed BACON 4, a production system that discovers empirical laws. The program represents information at varying levels of description, with higher levels summarizing levels below. BACON.4 employs a small set of data-driven heuristics to detect regularities in numeric and nominal data. These heuristics note constancies and trends, causing BACON 4 to formulate hypotheses, to define theoretical terms, and to postulate intrinsic properties. The introduction of intrinsic properties plays an important role in BACON.4's discovery of Ohm's law for electric circuits and Archimedes' law of displacement. When augmented with a heuristic for noting common divisors, the system is able to replicate a number of early chemical discoveries, arriving at Proust's law of definite proportions, Gay-Lussac's law of combining volumes, Cannizzaro's determination of the relative atomic weights, and Prout's hypothesis. The BACON.4 heuristics appear to be general mechanisms applicable to discovery in diverse domains.

Anderson and Neves are investigating how students learn to understand and formulate proofs in geometry. They combine psychological investigations with the construction of a production-system computer model based on the ACTG system. Some of the generalization, discrimination and composition mechanisms stemming from Anderson's earlier work in language learning are being successfully applied in this new domain.

Carbonell's META project investigates the acquisition of word meanings and new concepts from context and dialog structure as part of a natural language comprehension model. This work extends the learning-from-examples paradigm to include (1) memory reorganization, (2) constructing memory interconnections rather than "disembodied" concept acquisition, (3) interacting with a reactive environment (by example generation and experiment formulation), and (4) incorporating an ability to reason by analogy from known situations and processes. The central premise is that learning is a crucial, inalienable component of all human cognitive processes, and as such it cannot occur in the abstract or without processing significant amounts of information. The project-and-integrate algorithm and the principle of bounded generalization are two techniques under development based on propagating and transforming constraints in the learning process.

Pat Langley is investigating child language acquisition in his AMBER project. AMBER is a model of first language acquisition that improves its performance through a process of error recovery. The model is implemented in ACTG. AMBER starts with the ability to say only one word at a time, but adds rules for inserting additional words in the correct order, based on comparisons between predicted and observed sentences. These insertion rules may be overly general, leading to errors of commission; such errors lead to a call on the ACTG discrimination process, producing more conservative rules with additional conditions. AMBER's learning mechanisms account for many of the developments observed in children's speech.

There are other active learning projects, some are discussed in other sections (e.g., Fox's automatic classification techniques, and Herman's project of acquiring 3-D models from 2-D visual data). Other projects are just getting under way For instance, Glenn Iba, Pat Langley and David Nicholas hope to develop a flexible learning system that acquires and employs skill in the domains of puzzle-solving and game playing. More immediately, their objective is to explore the learning of "macro-operators" as a means of developing skill and reducing search in solving puzzles.

Natural Language Processing

John Anderson, Jaime Carbonell, Pat Carpenter, Phil Hayes, Marcel Just, Herb Simon, Robert Thibadeau

At CMU natural language has not been studied as a phenomenon in itself, but rather in its interaction with other cognitive processes, to gain a better understanding of human memory organization, problem-solving mechanisms and real-time perception. Carbonell is investigating self-extending language analyzers, requiring an investigation of metaphor comprehension, exploiting contextual and dialog-structure information. Simon's UNDERSTAND system investigated problem understanding as an integral part of a total problem solver. Phil Hayes is investigating flexible parsing techniques, such as coping with slightly inaccurate grammar in input sentences, or violations of case constraints.

Thibadeau, Just and Carpenter are investigating real-time reading behavior in humans. Thibadeau is building a system to simulate the time-course processing in human reading, as well as a parallel architecture for the comprehension of sentences. This work is based on accurate eye-movement and gaze-duration data collected and analyzed by Just and Carpenter. Don Kosy is investigating the problem of learning from a natural language programmed text, where the central problem lies in understanding natural language descriptions of new concepts and skills

Speech Recognition

Fil Alleva, Roberto Bisiani, Ron Cole, Raj Reddy, Alex Rudnicky, Alex Waibel

Speech understanding has a long tradition at CMU, from HEARSAY-I (the first connected-speech understanding system) through HARPY and its descendants. HEARSAY-II introduced the blackboard model to promote communication among a set of cooperating expert modules, and dynamic scheduling of modules based on expertise and data certainty. This computational paradigm has proven useful in AI applications outside speech recognition. HARPY introduced beam-search, a general technique added to the AI repertoire, while the LOCUST system (a HARPY descendant) demonstrated the feasibility of mini-computer-based large-vocabulary connected-speech understanding systems More recently, a multi-micro-processor version of LOCUST has been designed and implemented to provide real-time low-cost speech analysis.

A second project involves learning in speech-understanding systems Research is focused on identifying low-level speech knowledge whose selection will increase overall system performance. Two approaches are being taken. The first is a protocol analysis of people doing spectrogram analysis The second is a performance analysis of single-word and connected-speech systems. Further work will explore the learning of relevant higher-level knowledge, that is, lexical, syntactic, semantic, etc.

A voice message system is being constructed that will accept voice input and produce voice output directly from stored speech or synthesized from text The speech-understanding systems and the graceful interaction project (described below) will be integrated into this system

Vision

Jerry Agin, Martin Herman, Takeo Kanade, John Kender, Dave McKeown, Hans Moravec, Raj Reddy

Research in vision at CMU is quite diverse. Some projects focus on low-level issues, such as Kender's work on recovering shape from texture Other projects focus on high-level scene analysis. Kanade extends Waltz-Huffman-Clowes labeling and combines it with geometric search-constraint heuristics Herman investigates the learning of 3-D object descriptions. Some vision research takes place under the auspices of the Robotics Institute and is closely coupled with work on manipulation systems.

Kender continues to extend his work on how object shape can be determined from textural and other object regularities The work is based on a predictive, computational paradigm that handles both orthographic and perspective imaging. Under study are the application of the paradigm to regularities perceived under motion and binocularity. Also being investigated are the relative advantages of various representations for surface orientation. Lastly, the integration of these methods into a full image understanding system is being pursued.

Kanade has developed a theory of the Origami world (a generalization of Waltz's tetrahedral world), and has shown how it can be used to recover 3-D shape from 2-D images. In an implementation of Kanade's theory, image regularities are mapped into shape constraints, which allow quantitative shape recovery.

The goal of Herman's research is to develop a system that learns concept-models of 3-D objects from examples in the form of 2-D images. The first step toward this goal is for the system to learn automatically the 3-D shape of an object from the images. The approach is to represent objects in terms of Origami surfaces, and generate partial shape descriptions from each of several 2-D views of the object. The partial 3-D descriptions will be combined by a process of projecting some of them along the directions of other views, and then seeing how well the projections match the descriptions derived from the latter views. The best-matching descriptions, when integrated into a coherent model, form the complete 3-D description.

Current research by McKeown and colleages in the design and implementation of a multi-sensor image database to support research in interpretation of aerial and satellite imagery has focused on the issues of using map knowledge of a particular task area to guide the interpretation processing. The goal is to produce an interactive aid for photo interpreters that enables users to build textual descriptions of scene content, perform simple measurements, and relate regions of the image to specific man-made and natural features. One key area of the research is how to structure map knowledge, which represents a static and idealized view of the world, to allow for symbolic indexing into an image database and the selection and manipulation of landmarks and cultural features.

Problem Solving and Search

Hans Berliner, Jaime Carbonell, James Greeno (U Pittsburgh), Jill Larkin, John McDermott, Allen Newell, Herb Simon

The study of human problem solving provided much of the original impetus in early AI research at CMU. For instance, automatic protocol analysis was developed, models of problem solving were built (e.g., originally Newell, Simon and Shaw's GPS, and much later Simon's UNDERSTAND system), and production systems were developed primarily to advance such studies.

Presently, Simon, Larkin, Greeno and McDermott are investigating human expertise in problem solving (in the domain of mechanics problems in physics). These investigations have led to production-system models of expert and novice problem solvers. The primary difference between the two models appears to be that the expert works forward from the premises, knowing which rules or formulae to apply at each step, whereas the novice works backwards from the goal, applying means-ends analysis. Under current investigation is the problem of how expertise is acquired in a particular domain; i.e., how to transform gradually a general means-ends analysis mechanism into a constrained forward-chaining inference discipline.

Berliner developed the B* search algorithm for game-playing situations. The algorithm is capable of producing optimal search strategy given reliable evaluation functions and a uniform distribution of values in the range of uncertainty. Now, Andrew Palay is extending the method, making it applicable to more general probability distributions. This has the potential of improving the algorithm considerably when information is available to postulate specific non-uniform distributions.

Carbonell developed the DELTA-MIN algorithm to guide backtracking decisions. In essence, when best-first search is not possible (due to lack of a meaningful global evaluation function), but local decisions can be ranked as a function of expected reliability, DELTA-MIN provides a method of backtracking to the least-certain decision points, with the option of propagating reasons for failure.

The notion that problem solving requires search in an appropriate problem space has become a reasonably well-accepted thesis in AI. Newell is attempting to extend this thesis, by exploring whether all symbolic cognitive activity, elementary and routine as well as problematical, takes place in a problem space.

Attention has focussed on the nature of evaluation functions as well as on the nature of search algorithms. Berliner has done extensive work on non-linear context-sensitive evaluation functions. This work was done in the game-playing domain, culminating in BKG 9.8, a backgammon program based solely on a context-sensitive evaluation function (no forward minimax search) that defeated the world champion in an exhibition match. The principle underlying context-sensitive evaluation functions is the introduction of application coefficients, i.e., variables that change continuously (and gradually) as a function of measurable parameters of the global context. The effect obtained is that knowledge comes into play gradually when called for, but does not interfere when the situation changes and it becomes less relevant. Currently, Berliner is applying his techniques to develop a version of BKG that teaches backgammon to a human student.

As part of the POLITICS project, Carbonell continues to work on strategy-based counterplanning in multi-actor planning scenarios, where the planner must contend with the actions of adversaries and allies. In order to improve performance and robustness, Carbonell developed two simple principles of evidentiary combination. These principles, analogous to Berliner's application coefficients, are applied in the context of a rule-based, forward-chaining inference engine. In essence, the principles specify ways to combine relevance and certainty measures of information at hand, so that the system can make judgements based on indicative, corroborating or contradictory evidence in the absence of more reliable direct causal evidence.

Automatic Programming

Elaine Kant

Automatic programming is one domain for expert systems in which the builders of the system can be their own experts. Kant's research covers a range of techniques involved in the automatic design, construction, and analysis of algorithms and programs. The overall framework is based on heuristic search in a space of alternatives generated by design and implementation rules. Search is controlled by explicit rules that include both heuristics and an evaluation function based on program costs. Current work includes (1) a study of the design techniques used in computational geometry algorithms and (2) the refinement of representations for information about programming and design techniques. The representations are expressed with flexible inheritance techniques for relating different programming (and domain-dependent) constructs, implementation rules, and search strategies. Some intended uses for the representations are to facilitate constructing good user interfaces that help explain what a synthesis system is doing and to help in adding new knowledge and modifying or adapting exisiting information. Earlier, Kant developed the LIBRA system for automatic cost-evolution and search control in PSI, the automatic programming project at Stanford.

Memory Representation

John Anderson, William Chase, Scott Fahlman, Mark Fox

Since knowledge representations are central to most other AI research, many ongoing projects have a sizable component dedicated to developing and improving representation and memory access mechanisms. For instance, Kant's work in automatic programming requires a flexible inheritance mechanism for a network of rules and implementation techniques. All of the work on production system architectures can be seen as an exploration into specific memory representations. For instance, the work of Anderson focusses on a semantic-net-like memory; the work of Newell and Rychencr focusses on long term memory in pure procedural (production) form. However, some projects are dedicated exclusively to the development and improvement of knowledge representation methods.

Fahlman's NETL project is concerned with using massive parallelism to obtain human-like speed in accessing large volumes of real-world knowledge. This effort currently has two foci: completing the design of a million-element NETL machine in 1981, and improving the representational conventions of NETL to remove long-standing problems, which plague current semantic network systems. Current work is investigating how NETL's marker-passing parallelism differs from more traditional parallelism.

Fox's work lies in the intersection of knowledge representation and knowledge acquisition. The learning component consists of a mechanism for acquiring new classification hierarchies and causal networks. The approach taken is that structure discovery is teleological, where the expected use of the information guides its structuring. A system is being constructed that takes information and a structuring criterion (goal) and discovers one or more classification hierarchies. Research is also focused on the construction of new types of inheritance mechanisms. In particular, inheritance concepts allow the definition of new inheritance relations, and their context-dependent specialization. Fox is also examining the problem of generating procedural code from a declarative definition of a function concept hierarchy.

Finally, Chase is investigating spatial knowledge and its representation, especially where differences in representation relate to differences in skill on spatial tasks (such as navigation).

Graceful Interaction

Gene Ball, Phil Hayes, Raj Reddy, Mike Rychenei

Both this project and the next belong to an area of computer science that is not normally included in AI, namely user-computer interaction. However, we include them, because at CMU many of those involved are primarily AI researchers, and bring to the problems an AI approach.

The aim of the Graceful Interaction project is to develop interactive user interfaces that will be significantly more flexible, friendly, and pleasant to use than those available at present. The approach is based on the notion of an intelligent user agent, which intercedes between the user and a (rather stupid) functional tool system he wishes to use. Because of the amount of effort required to construct it, the user agent is tool-independent, and derives its knowledge of the tool from a declarative data base we call the tool description. The user agent must (among other things) parse input in natural or more cryptic language and integrate it with pointing input; follow the user's focus of attention and use it to resolve anaphoric and elliptical inputs; interact with the user to resolve ambiguities or contradictions.

Hayes is focusing his efforts on flexible parsing, incorporating the ability to analyze somewhat ungrammatical sentences, and on the design of a robust but friedly user interface that minimizes prerequisite user knowledge of the underlying system. Mike Rychener is investigating issues of personalizing interfaces to accord with preferences and needs of individual users.

ZOG

Mark Fox, Don McCiacken, Allen Newell, George Robeitson, Kamila Robeitson, Roy Taylor

ZOG is a rapid-response, large-network, menu-selection interface. Each node of the network is a display and the user moves through the network by making repeated selections. ZOG's primary purpose is to provide a powerful and versatile vehicle for exploring the man-machine interface. It is currently being used primarily to construct a research test-bed aboard the USS Carl Vinson, a new nuclear aircraft carriar. The system will be used to hold an online manual of operations and procedures, and to facilitate top-level task scheduling.

Fox has developed a variant of the ZOG systems, called BROWSE, which provides online access to the CMU Computer Science library. BROWSE uses an off-line data base to generate the appropriate ZOG network of frames, used for browsing and searching the catalogue of entries.

In addition, completion of a User Studies Laboratory now facilitates analysis of man-machine interaction in much greater detail than previously possible.

SPICE

Eugene Ball, Scott Fahlman, Sam Harbison, Petei Hibbard, Rick Rashid, George Robertson

The SPICE project (Scientific Personalized Integrated Computing Environment) is a large 5-year effort to develop a high-quality computing environment on single-user machines, connected via an ether-like network,

permitting access to a (logically) centralized file system and to other specialized and experimetal computing facilities. In addition to a very large address space and high-resolution color display, each machine will have roughly the computing power of a KI-10. SPICE will support a multi-language environment (with interlanguage communication via interprocess communication). There will be two primary programming environments, one based on the algebraic language ADA and the other on LISP.

Scott Fahlman is supervising the development of a complete LISP environment, first by emulating a virtual machine to speed software development, and later experimenting directly with the available hardware.

SPICE involves the entire CMU Computer Science Department, both in its development and in its promise of providing the primary computing facility for the department. Though the technical problems are primarily in the area of software systems, we describe it here because a number of the CMU AI researchers are deeply involved in it and because it will form the basis of the AI computing evironment in the middle-range future.

Scientific Relevance of Robotics

Many other things happen in universities that increase this diversity of goals Visitors are welcomed into the environment, often with their own support, hence independence. They attack the problems they see as important from their own interests, which do not coincide entirely either with the educational or the research interests of their host. The result is an increase in

diversity and an increase in the opportunities for science to emerge.

What I claim here is hardly novel. Yet it is important to restate it on this occasion. What often passes for organizational noise in a university effort, is precisely what permits the emergence of the truly fundamental. The noise cannot be too great, but neither can it be too quiet

Conclusion

Let me conclude. A birth -- or perhaps we should call this a christening -- is a time for good wishes and good hopes. So everyone gets to have his own It is my good fortune to be a speaker this morning, and thus to share my hopes with all of you Of all the hopes I could have for the Robotics Institute, the one that is most precious to me -- hence the one I hereby share -- is that it provide the place where some fundamental advances will occur in our scientific understanding of intelligent action. The whole of my remarks have been addressed to showing you why I think that hope is far from being in vain

I understand fully why others should have other hopes. Hope that the Robotics Institute adds intellectual excitement to the campus Hope that it creates a vortex of interdisciplinary activity Hope that the Institute itself, by its very form and function, helps to pioneer a new era of cooperation between university and industry. Hope that it help in our national attempt to become a more productive society. All these hopes, and others too, I also hold. But I must confess, I like mine best ∎

Artificial Intelligence Proceedings

1980: NCAI

Proceedings of the First Annual National Conference on Artificial Intelligence held at Stanford University, sponsored by the American Association for Artificial Intelligence. (In one volume.) Order from:
> Garcia Robinson, Inc.
> 301 Menlo Oaks
> Menlo Park, California 94025

1979: IJCAI6

Held in Tokyo, Japan (In two volumes.) Order from:
> IJCAI6
> Computer Science Department
> Stanford University
> Stanford, California 94305

1977: IJCAI5

Held at MIT, Cambridge, Massachusetts. (In two volumes) Order from:
> IJCAI5
> Department of Computer Science
> Carnegie-Mellon University
> Pittsburgh, Pennsylvania 15213

1975: IJCAI4

Held in Tbilisi, Georgia, USSR (In two volumes) Order from:
> IJCAI4
> MIT-AI Laboratory
> 545 Technology Square
> Cambridge, Massachusetts 02139

1973: IJCAI3

Held at Stanford University. (In one volume) Order from:
> Artificial Intelligence Center
> J2044
> SRI-International
> Menlo Park, California 94025