AUTOMATED REASONING FOR BIOLOGY AND MEDICINE

Eric J. Horvitz
Stanford University

During the last decade, computer scientists have made significant progress in developing techniques for storing and retrieving information, and for solving difficult inferential problems with computer-based reasoners. The growth in the power of computer processors, and the parallel decline of the cost of computer memory, has catalyzed the development of innovative software for problem solving. In particular, there have been promising advances in computational methods for acquiring, representing, and manipulating biological and medical information. I will present key concepts of automated reasoning investigated in the computer-science subdiscipline called *artificial intelligence* (AI). I will frame my discussion in terms of the genesis and maturation of AI and related subdisciplines that were spawned shortly after the development of electronic computers, and will review key themes that have dominated research over the last three decades.

IN PURSUIT OF A CRISP DEFINITION

What is artificial intelligence? It is often difficult to construct a definition of a discipline that is satisfying to all of its practitioners. AI research encompasses a spectrum of related topics. Broadly, AI is the *computer-based* exploration of methods for solving challenging tasks that have traditionally depended on people for solution. Such tasks include complex logical inference, diagnosis, visual recognition, comprehension of natural language, game playing, explanation, and planning.

In *Advances in Computer Methods for Systematic Biology: Artificial Intelligence, Databases, and Computer Vision*, Johns Hopkins University Press, 1993. *From opening talk, Conference on AI in Systematic Biology, Napa Valley, California, September 1990.*

I have highlighted "computer-based" to stress the centrality in AI of the use of computers to experiment with implementations of abstract theories of reasoning. Experiments with computational models play an important role in distinguishing AI research from other inquiries into reasoning and problem solving, including philosophy and theoretical psychology. The sculpting of complex computational models gives AI researchers an experimental basis for testing theories of problem solving and for comparing alternative solution strategies. Implementing real systems also provides researchers with opportunities to explore unanticipated or counterintuitive behaviors of complex models.

Several AI endeavors have led to the development of distinct approaches for performing well-defined tasks such as medical diagnosis or species identification from observations. These methods are sometimes referred to as "AI techniques," especially when they are available as software packages for use by nonspecialists. Such AI reasoning techniques provide a means of representing knowledge and a set of inference procedures for manipulating or drawing conclusions from that knowledge.

FORMAL VERSUS DESCRIPTIVE AI RESEARCH

There are different methodological approaches to AI research. Two large, coexisting phyla of investigators in the AI world are the *formal* and *heuristic* groups. The scientific methodology of the formalists is analogous to that of early aerodynamics scientists attempting to construct theoretical foundations for building a flying machine. In contrast, researchers in the heuristic group are like early airmen, enthusiastically attempting to take to the air, sometimes with flimsy assemblies of wood and fabric engineered to reflect intuitions drawn from their observations of birds in flight.

The formalists study systems based on logic or other persuasive sets of axioms, such as probability theory. Such axiomatic systems define consistent theories of reasoning. Formalists are committed to the view that reasoning systems based on axiomatic principles can explain and generate intelligent behaviors, much as the set of astronomical laws developed by Newton, Tycho, and Kepler provides a satisfactory explanation of the complex motions of heavenly bodies.

Investigators in the heuristic group reject the requirement for using reasoning techniques that follow from a set of mathematical principles, often noting that "scaling up" the formal methods to problems of realistic size can lead to computationally intractable problems (Simon 1969; Szolovits 1982; Buchanan and Shortliffe 1984). These scientists have also argued that the parsimonious axiom-based methods lack the richness or *expressiveness* needed for intelligent behavior (Gorry 1973; Davis 1982).

Instead of relying on mathematical principles, investigators in the heuristic group apply intuitively attractive methods they hope can provide a more direct and tractable approach. These intuitive, and often ill-characterized, procedures are referred to as heuristic methods. They are based typically on observations and introspection about the ways in which people seem to solve problems. Heuristic approaches that explicitly attempt to model human problem solving are called *descriptive* in that they attempt to describe how people solve problems. Descriptive theories of problem solving are analogous to the epicycle machines used to describe the motions of the planets before modern astronomy offered satisfying theoretical explanations for observations.

Although there have been debate and tension between people in the formal and heuristic groups, both approaches have advantages and disadvantages. Formal models can help in developing a clear understanding of reasoning methods. A successful formal theory for solving a class of problems can be useful for creating computational methods that are easy to modify and to extend for application in diverse areas. Heuristic approaches are especially useful where no computationally tractable formal method is available for solving a problem. Indeed, heuristic models and solutions are often a precursory step in the development of new theories. Investigators in the heuristic group have also suggested that the heuristic approaches are more natural, and thus make it easier to acquire knowledge from experts for use in computer-based reasoning systems, and to explain the conclusions of machine reasoning to people.

THE GENESIS OF AI

Several factors coincided to create the research environment that spawned a set of information-oriented disciplines in the mid-1940s. Key forces that created and shaped the new disciplines were the pressure for solving complex tactical problems during World

War II, the development of electronic computers, and the development of the conceptual framework of subjective expected utility.

Under the pressures of war, blueprints for general electronic computing systems were implemented and rapidly refined. The need to solve critical warfighting problems and the availability of computers stimulated research on *operations analysis.* Operations analysis focused on the design of complex systems, processes, and deployments, through optimization and control. Briefly, the solutions to operations-analysis optimization problems are optimal values for variables that define the behavior of a system. These optimal values maximize some measure of benefit or minimize some measure of cost, such as dollars, time, distance, or lives.

Operations analysis matured into the modern discipline of *operations research* (OR). Following World War II, OR methods were used to analyze a a broad spectrum of civilian problems, such as determining an ideal schedule for maintenance of a fleet of airplanes, purchasing an optimal number of ambulances for a county, or minimizing the number of oil tankers to meet a fixed schedule (Dantzig and Fulkerson, 1954). Investigators in the OR community developed a set of numerically intensive methods, often referred to as "OR techniques," for modeling and solving optimization and scheduling problems (Dantzig 1963). These OR techniques included linear programming, nonlinear programming, integer programming, dynamic programming, queuing theory, and control theory.

In 1947, von Neumann and Morgenstern developed an axiomatic framework called *utility theory*, founded on the preexisting axioms of probability theory. Briefly, utility theory provides a formal definition of preference and of rational decisions under uncertainty, that is, decisions taken in light of uncertain knowledge about a situation (von Neumann and Morgenstern 1947; Raiffa and Schlaifer 1961). The axioms of utility theory define a measure of preference called *utility*. Utility theory dictates that people should make decisions that have optimal average, or *expected,* utility.

Under uncertainty, we do not know exactly how an action will affect the world. The fundamental procedure in determining a course of action that has optimal expected utility is to consider the probabilities of alternate outcomes associated with a possible action.

The expected value of each action is computed by summing the value attributed to each possible outcome multiplied by the probability of that outcome.

Expected value, teamed with the concept of personal or *subjective probability*, provide decision analysts with a powerful framework for maximizing expected utility. Many people that have been trained in classical statistics are not familiar with the interpretation of probability as a subjective measure of belief. That is, the most widespread interpretation of probability is that of a measurable frequency of events determined from repeated experiments. For example, probability is viewed typically as the number of times a coin (fair or biased) lands on one side, given some large number of coin flips. This view has been called the *frequentist* interpretation.

A different perspective on probabilities, known as the *subjectivist*, or *Bayesian*, interpretation, is that a probability is a measure of a person's degree of belief in an event, given the information available (Savage 1972; Hacking 1975). According to this conception, probabilities refer to a *state of knowledge* held by an individual, rather than to the properties of a sequence of events. Subjective probabilities abide by the same set of axioms as do classical probabilities or frequencies. The subjectivist approach is a generalization of the more popular notion of a probability as a long-run frequency of a "repeatable" event. The use of subjective probability in calculations of the expected value of actions is called *subjective expected utility* (SEU). SEU allows personal opinions and hard data to be integrated in a theoretically sound manner.

The development of SEU contributed to the maturation of OR, and to modern economics and psychology. In addition, it provided the basis for the emergence, in the early 1950s, of management science and decision analysis, disciplines that are conceptually and academically related to OR. Decision analysts study techniques for determining ideal personal decisions under uncertainty (Raiffa and Schlaifer 1961; Howard 1966; Howard 1968; Keeney and Raiffa 1976). Investigators in management science develop and test mathematical models of production and organizational decision making.

THE 1950s: EARLY DAYS OF THE DISCIPLINE

In the early 1950s, AI research evolved separately from OR, and from the related disciplines of management science and decision analysis. Rather than relying on numerical analyses, early AI investigators turned their attention to the automated processing of abstract symbols (Simon 1972). Herbert Simon, one of the founding fathers of AI, has described how the concept of manipulating sophisticated symbols with computers was stimulated by work with computer systems developed for performing complex numerical analyses (Simon 1987). In the early electronic computers, simple mathematical functions such as ADD and MULTIPLY were encoded in the form of binary numbers. That is, these mathematical *concepts* were represented in the same way as the numbers they operated on. If simple mathematical notions could be manipulated by computers, perhaps, by analogy, more sophisticated and abstract concepts could be processed.

AI research was born with a great enthusiasm for the possibilities of harnessing computer-based manipulation of abstract symbols to perform "intelligent" reasoning tasks. Early on, the field became dominated by studies of computational inference using principles of logic. AI researchers sought to encode knowledge *about* problem solving in the form of explicit, or *declarative*, representations of objects and relationships in the world, in constrast to the traditional *procedural* approach as typified by OR solution methodologies (Barr and Feigenbaum 1982; Cohen and Feigenbaum 1982).

AI researchers also dismissed numerical methods as relatively unimportant to decisions made by agents with the abilty to manipulate abstract symbols. In particular, methods for optimizing the expected value of action under uncertainty seemed inadequate for explaining cognition and intelligent problem solving. Early investigators noted that people successfully address challenges in the world without using numerical optimization procedures. In particular, number-intensive decision-analysis and OR procedures seemed to require inappropriate detail and analytic complexity.

Early investigators pointed out that, in contrast to the crisply defined problems amenable to OR or decision-analytic techniques, almost all real-world problems faced by people in daily life are intrinsically *ill-defined*. Thus, rather than pursue an optimal solution to well-defined numerical problems, many AI researchers have worked to develop techniques for finding suboptimal yet satisfactory behaviors and decisions, known as *satisficing* solutions (Simon 1969). Many AI investigators believed that, beyond being inappropriate, attempting to apply numerical techniques to real-world problem solving would require extraordinary amounts of computation (Feigenbaum 1964).

In fundamental work in the mid-1950s, Newell and Simon built systems that could perform logical reasoning. The first logical theorem-proving programs were created to derive simple mathematical proofs. However, the primary goal of the investigators was to identify general principles of intelligent problem solving. Newell and Simon constructed a program called the Logic Theorist, and later built a descendant system called the General Problem Solver (GPS) (Newell 1958; Newell and Simon 1963). GPS employed a computer representation of logical axioms. The system derived new logical relationships by repeatedly applying one of several rules of logical implication, or *operators,* to the axioms, or to other logical statements stored in the system. The system could generate long chains of inference by applying logical operators to statements stored initially or generated later by the system.

GPS's structure and capabilities highlighted several important concepts. The system set a precedent for separating inference machinery and capabilities from domain knowledge. That is, domain-specific knowledge was stored in a separable, or modular, *knowledge base.* Research on GPS also pioneered the use of *means-ends analysis* to guide the selection of operators.

Means-ends analysis is an approach for repeated selection of operators that transform the current version, or *state,* of the solution to a new subproblem that is closer to a final solution, or *goal*. (Nilsson 1980). Operators are selected and applied until a solution is reached. With GPS, the goal was to prove a logical theorem from a set of initial axioms and logical statements. As an analogy, a sequence of operators for traveling to Stanford University from M.I.T. is "walk to Technology Square," "automobile to Logan airport," "plane to San Francisco," "walk to taxi," "taxi to Stanford." In this example, subproblems are different positions in space that are progressively closer to a goal. In logical theorem proving, as in actual real-world problems, there are typically many possible choices, or sequences of operators, that might be applied to any stage of a solution. Thus, we need reasoning techniques that can identify the best next operator. In the work on logical reasoning, operators were selected by heuristics that considered the relative ability of operators to minimize the conceptual distance between the current state and a goal or a necessary subgoal.

In l959, the term *artificial intelligence*, as the name of a discipline charged with the study of computer-based problem solving, was first used at a Dartmouth conference by John McCarthy, a computer scientist and mathematician interested in the philosophical foundations of computational models of intelligence. The name stuck, and the conference participants formed the core of an evolving field, distinct from other areas of computer science, dedicated to furthering our understanding of intelligence as the processing of symbols.

THE 1960s: MATURATION OF LOGICAL REASONING METHODS

In the late 1950s and early 1960s, investigation of techniques for reasoning with logical methods continued to mature. Theorem proving was applied to solve mathematical problems beyond logic. For example, Gelernter built a system for performing geometrical proofs (Gelernter 1959). *Search* evolved as a central notion in intelligent problem solving. Systems must often rely on the generation and evaluation of large sets of alternatives to solve logical reasoning problems. In GPS, for example, there are many different ways to generate subgoals and goals from an initial problem statement and set of axioms.

To visualize the large *search trees* that a reasoning system must contend with in solving an inference problem, consider each subgoal to be a point in a conceptual space that separates an initial state from a final answer or goal. We can represent the fact that one subgoal can be transformed to another subgoal with an operator by drawing an arc between the *parent* and *descendant* subgoals. Problem solving can be viewed as choosing a path through a large branching tree created by the following procedure, or *algorithm*: (1) Generate a set of subproblems by first applying all valid operators to an initial problem, and (2) perform the same subproblem "expansion" on each subproblem recursively, until we reach goal states or cannot apply any more operators. The subgoals in such a search tree define the *search space* for a theorem-proving problem.

Consider, as another application, the classic search problem, called the traveling salesperson problem (TSP) (Dantzig et al. 1954; Christofides 1979; Lawler et al. 1985). The TSP requires us to identify the shortest path connecting a set of cities each of which the salesperson must visit once before returning to a home base. In generating the search tree for the TSP, we apply a MOVE operator to generate all possible trips from the current location to another city. We apply the MOVE operator again to each new location, making sure each time that we move only to unvisited cities. The process continues for each path through the tree of cities until we have visited all cities and returned to the starting position. Alternative orderings of cities compose a set of *tours*. Search algorithms for TSP identify different partial sequences of subgoals on the path to complete tours. The shortest circuit, or tour, is the optimal answer to a TSP problem instance.

Finding the optimal solution to a TSP problem is difficult; TSP and a variety of other difficult problems constitute a family of problems called *NP-complete*, in a taxonomy of problem complexity developed by theoretical computer scientists to classify the fundamental difficulty of problems (Garey and Johnson 1979; Papadimitriou and Steiglitz 1982; Aho et al. 1983). With high likelihood, the time required to solve NP-complete problems exactly grows exponentially with the size of the problem (for example, the number of cities). That is, the amount of time required to solve these problems grows geometrically with linear increases in problem size. A great number of problems that are of interest both to AI researchers and biologists are in the NP-complete class. For example, many taxonomists are probably familiar with the NP-complete *Steiner Graph Reduction* problem, used to build plausible taxonomic trees from data. Because NP-complete problems are difficult to solve exactly, numerous algorithms have been developed to generate approximate solutions with far less effort than the work required for a naive search of all possibilities (Papadimitriou and Steiglitz 1982).

Intelligent Versus Brute-Force Search

Without forethought, we might attempt simply to generate all possible sequences of operators until we find an answer to a search problem. Such *brute-force* searches require a great amount of computation. However, there are many ways to navigate through a large tree. AI investigators have developed search techniques for guiding search more intelligently. Frequently, these procedures can be used to solve a problem more quickly than the worst-case situation described by NP-complete brute-force methodologies.

The design of intelligent search procedures often hinges on the identification or development of an *evaluation function* that is used to assign a score to each intermediate state along the path to an answer (Nilsson 1980). Some search techniques rely on logical analyses of the dominance of one or more states to identify irrelevant paths through a tree. Such dominance information comes from the use of deterministic evaluation functions for identifying the value of a subgoal. For example, in one approach, called *branch-and-bound*, paths through a search tree are pruned from consideration if subgoals generated by those paths will be less valuable than the best approach discovered so far.

Often, branch-and-bound is not applicable because we do not have a deterministic measure of value for each subgoal. Instead, for many problems, researchers have relied on heuristic estimates of the value of the "goodness" of an answer. For example, the GPS means-ends analysis depends on the use of a heuristic evaluation function to control the selection of a best operator at each step of an analysis.

Sometimes we can evaluate the behavior of a heuristic search method in terms of an optimal analysis or answer. For example, for some search techniques, investigators have developed formal proofs that show how we can characterize the error in an approximate answer in terms of an exact answer (for example, the shortest tour possible through a set of cities), or characterize the amount of computation required by an approximate analysis as some fraction of the work required by an exact solution.

Search research spans several types of problem solving beyond theorem proving and tour identification. An active area of search research is called game playing. Several research teams have studied computer-based reasoning for competing in games such as checkers and chess (Samuel 1959; Samuel 1967; Good 1977; Barr and Feigenbaum 1982; Berliner 1980). In the case of chess, a search space is generated by applying the rules for moving chess pieces from an initial legal position. AI investigators have designed chess programs that assess the promise of the legal moves that can be made from a current board position by evaluating paths of moves in large trees of successive moves that are made possible by each candidate move. The large trees of moves are generated by considering a volley of moves and countermoves to some depth of player interaction, or "lookahead."

Production Systems for Capturing Expertise

In the late 1960s, many AI investigators continued to study the application of logical techniques similar to the methods employed by Simon and Newell in GPS. However, instead of remaining within the confines of logical theorem proving, investigators began to examine the manipulation of symbols and knowledge in specialty areas, with the goal of building computer systems to assist people with decision making in scientific and professional endeavors. Work began on the construction of knowledge bases that could be used to bring expertise to decision makers in such professions as chemistry and medicine.

Systems based on the use of sets of logical rules of the form IF A THEN B, and on the use of logical-chaining techniques for deriving conclusions from the rules, came to be called "production systems," and later "rule-based systems" (Davis et al. 1977). One of the earliest knowledge-intensive production systems was a program for performing difficult symbolic mathematical inference for engineering applications. In this work, mathematicians constructed knowledge bases of mathematical identities that were used for the symbolic integration and differentiation of equations. This work led to one of the most widely used AI programs in the world today, called MACSYMA. MACSYMA is used by scientists and engineers for performing complex symbolic integration.

Another early application of production systems to real-world problems was the Dendral project at Stanford (Lindsay et al. 1980). For this project, expert rules describing chemical structure and decomposition were used to reason about mass-spectroscopy data. At the core of the DENDRAL system was a large knowledge base composed of rules that described how molecules, and portions of molecules (as subgoals), might decompose in a mass spectrometer. Experts in organic chemistry collaborated with computer scientists to build DENDRAL's knowledge base and to evaluate the functioning of the system.

Planning Research

The late 1960s and early 1970s saw a marked growth of interest in an area of research called *planning*. To this day, planning research continues to be a hotbed of AI research. In contrast to the study of production systems to automate expert reasoning and decision making within the confines of a specialty area, planning research has dwelled on *commonsense* reasoning, that is, reasoning whose goals are routinely achieved by people in the course of daily life. Investigators interested in planning have developed logical-reasoning systems similar to the older theorem-proving systems. However, instead of storing abstract logical operators and statements in the system's knowledge base, the investigators encoded operators and logical rules that capture basic physical and mechanical properties of a domain, such as the information that two objects cannot be in the same place at the same time. The rules also represent the preconditions and effects of *actions*. Also in contrast to theorem proving, planning problems require investigators to consider the effects of actions on the world.

Planning research has demonstrated that the simplest goal-oriented behaviors involve complex chains of reasoning about goals and subgoals. In an early planning project, called STRIPS, investigators tackled tasks in a Lilliputian world of blocks on a surface (Fikes and Nilsson 1971). In this world, which came to be known as *Blocksworld*, the goals were simple tasks such as changing one configuration of stacked blocks to another. The STRIPS reasoner demonstrated the complexity of using computers to perform tasks that appear quite natural and simple to people. For example, at first blush, it might appear relatively straightforward for a computer to stack a block, say Block B, on Block C, starting from an initial state of affairs in which Block D is on Block C and Block B is beneath Block A. Such tasks exposed problems such as costly interactions that occur in making moves to attain subgoals. That is, a careless or shortsighted action, seemingly on the way to a goal, might make it difficult or impossible to reach that goal.

Vision Research

During the 1960s and 1970s, great strides were made in the development of algorithms for identifying shapes and objects in a visual field. Like planning research, automated vision research highlighted the difference between tasks that are easy for people and tasks that are easy for computers. It quickly became clear that computer programs have great difficulty recognizing patterns and objects that people recognize with ease (Winston 1975).

Researchers studying computer-based vision sought early on to extract features through "global processing," the homogeneous application of simple mathematical transformations to an entire scene. Such processing was applied to generate information about classes of features, including the presence of edges, and the shading and textures of surfaces (Marr 1975; Ullman 1979). The earliest vision-identification techniques were supplemented by increasingly sophisticated methods for the *segmentation* of an image into its chief components (Marr and Hildreth 1980; Marr 1982). Search-based methods were developed for building objects out of sets of segmented features. In one set of approaches, a search is applied to identify a feasible *labeling* of edges in a scene as different possible boundaries of an object, and to build up sets of edges into three-dimensional objects. In some of this work, identification procedures draw on a database of objects for information about what might be present in a visual field.

A trend in vision research from the 1960s to the present has been to use progressively richer models of objects in the world to reduce the computational complexity of visual identification (Binford et al. 1989). Toward the late 1960s, techniques were developed for modeling three-dimensional objects. In one such approach, structures are identified by mapping a set of cylindroid structures to a visual scene. In many current vision projects, relatively detailed three-dimensional models are used to choose among alternate interpretations of a scene. For example, three-dimensional anatomical models have been used to assist in the automated interpretation of medical ultrasound and radiographic images (Brinkley 1985).

THE 1970s: KNOWLEDGE REPRESENTATION AND EXPERT SYSTEMS

Many areas of reasoning, including the inferential principles and techniques developed in the 1950s and 1960s, received continuing and intensive attention in the 1970s. The 1970s were marked by growing interest in the representation of knowledge, and by a surge of enthusiasm for the development of production systems to assist professionals and scientists. Other areas matured in the 1970s, including the study of methods for parsing and understanding natural language, for controlling the focus of attention of computer-based reasoning systems, and for learning from experience. I will discuss work on representations of knowledge and the use of these representations in expert systems.

Knowledge Representation and Problem-Solving Efficiency

A key theme of AI research in the 1970s was the importance of identifying expressive and efficient representations of knowledge. Early in the decade, investigators studied and discussed the best ways to codify knowledge for computational inference (Amarel 1968; Cohen 1977; Korf 1980). In an interesting piece of research, Saul Amarel, of the AI group at Rutgers, investigated the sensitivity of the complexity of solving a problem to details of the representation of that problem (Amarel 1968). He used examples from a classic computer-science problem called the "Towers of Hanoi." To solve this problem, a sequence of moves must be generated to transfer a set of discs of different sizes (resembling a Buddhist temple) from one pole to another, given a set of rules about legal moves. Amarel showed how changing the initial formulation of the Towers of Hanoi problem to a new representation could increase the efficiency of computational problem solving.

The studies of Amarel and others on the efficiency of representation came at a time of growing interest in reasoning systems that could perform inference on very large knowledge bases. There was a proliferation of alternative representations of knowledge in reasoning systems.

Expert Systems

Projects such as MACSYMA and DENDRAL were developed to explore how computer-based reasoning might assist people with difficult inference problems. However, it was not until the early 1970s that the phrase "expert systems" was first used to refer to computer programs that could draw conclusions by performing logical inference on a large knowledge base of information acquired from experts (Buchanan 1982). The expression was made popular in the medical domain.

In the 1970s, a great amount of work in AI was performed on automated medical diagnosis. A large community of investigators with a primary interest in AI in medicine

(AIM) coalesced as a bona fide subdiscipline of AI. Many AIM researchers endeavored to build systems to assist with diagnosis, given information about a patient's symptoms. Expert-systems studies included the MYCIN project at Stanford University, the Present Illness Program (PIP) at the Massachusetts Institute of Technology (Pauker et al. 1976), CASNET at Rutgers University, and INTERNIST-1 at the University of Pittsburgh (which continues today as the Quick Medical Reference (QMR) project at the University of Pittsburgh).

Attempts to build expert systems for medical reasoning highlighted the inadequacy of straightforward applications of logic for reasoning about large and complex domains. Logical-reasoning systems assign values of complete truth or complete falsity to possible states of the world and to the effects of actions. However, physicians are rarely certain about the state of a patient's physiology or about the effects of alternate therapies.

As work on expert systems in medicine progressed, investigators found that the deterministic nature of these systems did not allow experts to express uncertain relationships. A theme of 1970s work on representation was the modification of logical techniques, developed for automating logical theorem proving, to handle the more general situation of uncertainty. In one aspect of this work, investigators extended the logic-based production systems of the late 1960s with heuristic techniques for capturing uncertainty.

Popular representations of knowledge for systems that considered large quantities of expert knowledge about uncertain relationships included production rules (Davis et al. 1977), *semantic nets*, *causal networks* (Weiss et al. 1978), *frames* (Minsky 1974; Lindberg et al. 1980), and several mixtures of these representations. I will briefly describe these knowledge representations in the context of several medical expert systems developed in the 1970s.

MYCIN. MYCIN is a production system that was developed at Stanford to study the representation of medical knowledge (Shortliffe 1976; Buchanan and Shortliffe 1984). MYCIN reasons about the diagnosis and treatment of bacterial infections. The system uses a set of logical rules. Logical inference in the system is designed to follow a specific search strategy called *backward chaining.* With backward chaining, a system chains together a set of rules that creates an inference path, or *chain*, from causative agents to observed symptoms. *Forward chaining* refers to chaining from observations in the world to possible causes.

MYCIN investigators explored the extension of logical reasoning to address uncertainty by developing a heuristic calculus, called the *certainty-factor model* (Shortliffe and Buchanan 1975). With this approach, measures of the certainty or "strength" of a rule, called *certainty factors,* were assessed from experts. For example, consider the following rule from MYCIN:

> If the infection is a primary bacteremia, and the site of the culture is one of the sterile sites, and the suspected port of entry of the organism is the gastrointestinal tract, then there is suggestive evidence (certainty factor = 0.7) that the identity of the organism is Bacteroides.

MYCIN contains a knowledge base including hundreds of such rules. When rules are chained together logically, a certainty factor for the logical result is computed with a certainty-factor combination scheme. This numerical calculus propagates uncertainty to logical conclusions by considering the certainty factors associated with the rules that comprise an inference chain.

CASNET. Investigators at Rutgers studied the use of semantic networks with a system for the diagnosis of of glaucoma, called CASNET (Weiss et al. 1978). Semantic networks are graphs that consist of a set of directed arcs between concepts. In CASNET, the directed arcs describe relationships at three levels: observations, pathophysiological states, and diseases. Arcs called *associational links* are used to connect observations to pathophysiological states; arcs called *classification links* connect pathophysiological states to disease categories. Observations reported to CASNET activate a subset of pathophysiologic states; these, in turn, activate disease categories. The CASNET team experimented with the use of probabilistic information that describes the strength of association among interdependent concepts.

Present Illness Program. The Present Illness Program (PIP) for reasoning about renal disease was constructed in the 1970s by a team at M.I.T. (Pauker et al.1976). The system uses expert knowledge represented in the form of files of structured records, called *frames*. A separate frame was created for each disease considered by PIP. Each frame contains several classes of knowledge that experts have identified as being useful in considering the presence of a disease, given a set of observations. The classes of knowledge include typical findings or symptoms associated with a disease, criteria for making decisions, and relationships between the disease under consideration and the symptoms and diseases described in other frames.

The structure, or *architecture,* of PIP is based on descriptive intuitions about human cognition. The heuristic machinery used in PIP to draw conclusions from the knowledge stored in its frames includes a *supervisory program,* a *working memory,* and a *long-term memory*. In reaction to a set of symptoms observed in a patient, the PIP supervisory program "activates" a set of diseases by introducing these diseases from long-term memory to the active, working memory. Diseases linked to the activated diseases via relationships specified in the frames are added to the working memory. Finally, questions about symptoms of the activated diseases are asked to gather additional information about the patient. The process of refinement continues until a single disease, or a small set of diseases, remains in active memory.

INTERNIST-1. The INTERNIST-1 project was initiated at the University of Pittsburgh, in the 1970s, to construct a system for assisting physicians with diagnosis in the broad domain of internal medicine (Miller et al. 1982; Pople 1982). The INTERNIST-1 reasoning system relies on a numerical scoring approach for assigning belief to alternate disease hypotheses, given a patient's symptoms. It uses several classes of numbers obtained from experts. For example, one set of numbers is used to represent the degree of belief in the presence of different individual diseases, given a symptom. Another class of numbers characterize how often, for each disease, a symptom appears. INTERNIST-1 applies heuristic scoring rules to combine the numbers associated with different symptoms into comprehensive measures of belief in alternative disease entities. These quantities are used to generate a list of diseases, each with a heuristic measure of likelihood. Such a list of diseases ordered by likelihood, or *differential diagnosis,* is analyzed to select questions about the best next tests to perform to narrow the list of possible diseases.

INTERNIST-1 uses an iterative reasoning cycle called *hypothetico-deductive* reasoning. With this approach, a few salient observations are input to the system and a measure of likelihood is assigned to each possible disease. Based on this list of hypotheses, new questions and tests are recommended. After new information is input to the system, the system performs the entire cycle again, rescoring all of the remaining diseases and recommending new questions based on the revised disease list. Answers to the new questions are considered in conjunction with the initial salient features to generate a revised set of hypotheses, with revised associated likelihoods. This cycling process continues until only one disease remains or until the cost of gathering additional information outweighs the benefits of further refinement.

INTERNIST-1 continues to this day as the Quick Medical Reference (QMR) project at the University of Pittsburgh (Miller et al. 1986). Its knowledge base has continued to grow since the inception of the Internist-1 project in 1973. Today, QMR is one of the largest expert systems in the world.

The evolution of INTERNIST-1 to QMR illustrates how, for some applications, the power of commonly available computer hardware has grown more quickly than the sophistication, and concomitant thirst for memory and computation of our software. INTERNIST-1 was implemented on a large mainframe computer at Stanford, accessed by Pittsburgh researchers via a computer network. Today, QMR operates comfortably on inexpensive MS-DOS--based personal computers.

THE 1980s: MATURATION OF REASONING UNDER UNCERTAINTY

In the l980s, several rule-based expert-system architectures and representations matured to the point of commercialization. Several companies were founded with the goal of providing "expert-system shells:" software packages that could be used by investigators in specialty areas, outside of AI and computer science, to build their own expert systems. However, although the popularity and use of rule-based systems grew throughout the 1980s, there was a concurrent growth of interest in alternatives, especially in sound methods for addressing uncertainty.

Some AI investigators began to reexamine concepts and reasoning methodologies developed in decision analysis and operations research in order to develop more accurate reasoning methods. These investigators began to borrow ideas that would allow them to define the concept of a theoretically correct answer to a problem, or best action to take, given uncertainties about actions and outcomes. In particular, interest began to grow in the use of SEU to tackle difficult AI reasoning problems (Horvitz et al. 1988). In some of this work, probability and utility were applied at the metalevel to address decisions *about* the best ways to build a reasoning system, or to control reasoning to optimally solve a problem, given limited computational resources (Horvitz 1988; Russell and Wefald 1989).

I will focus on the development, in the 1980s, of SEU-related methods for using probability and utility to represent knowledge and reason under uncertainty; these techniques are particularly relevant to the subject of this volume.

Sufficiency of Heuristics for Reasoning Under Uncertainty

Throughout the 1980s, increasing attention was given to the integration of probabilistic and decision-theoretic reasoning methods with AI approaches. Some of this interest was stimulated by attempts to apply computer-based reasoning methods to progressively more difficult decision-making tasks. The importance of uncertainty, and of developing formal reasoning techniques, was underscored by the complexity and the high stakes of application areas such as medicine and aerospace.

Research had never halted on the use of probability and utility in the realm of expert systems and AI. In fact, several groups, working to the side of mainstream AI, had continued to explore the use of simplified forms of probabilistic and decision-theoretic reasoning for medical diagnosis (Ledley and Lusted 1959; Warner 1961; Gorry and Barnett 1968; Gorry 1973, Gorry et al. 1973). Typically, assumptions were imposed on these systems to make the representation and reasoning with uncertain information tractable. For example, most researchers investigating the use of probability and utility in the 1960s and 1970s assumed conditional independence among observations, and assumed that there was an exhaustive set of mutually exclusive diseases. By assuming conditional-independence, the researchers considered the probability of each observation to be unaffected by the presence or absence of other observations. Although the early probabilistic systems performed well in small subdomains, AI investigators showed little interest in them, beyond pointing out their limitations.

Most AI investigators understood that probability and decision theories served as axiomatic bases for ideal reasoning and decision making under uncertainty; however, they were not impressed by the simple probabilistic systems for several reasons. Many were convinced that it is difficult to obtain probabilistic information and to reason efficiently with probability and utility. Also, AI investigators studying heuristic techniques for reasoning under uncertainty in expert systems justified the need for heuristic techniques by focusing on the invalidity of the independence assumptions made by the probabilists to render their systems tractable (Buchanan and Shortliffe 1984). The researchers suggested that attempts to relax such assumptions would lead to unmanageable combinatorial increases in the computation time and in the numbers of probabilities required by the systems.

Interest in probabilistic inference in the middle to late 1980s was invigorated in part by theoretical work that demonstrated clear parallels between the shortcomings of several heuristic calculi and the simplified probabilistic inference. Several investigators demonstrated that the heuristic methods did not provide a means for escaping from the invalidity of the independence assumptions made in the simple probabilistic systems. In fact, for some of the heuristic methods, the assumptions were shown to be even more restrictive and, moreover, dangerous because they were implicit. For example, the MYCIN certainty-factor model and the INTERNIST-1 scoring scheme were shown to be roughly equivalent to the use of simplified and constrained probabilistic reasoning (Heckerman 1986; Horvitz et al. 1988).

AI scientists were also stimulated to reinvestigate the applicability of probabilistic reasoning by the development of analyses of the epistemological inadequacy of rule-based methodologies for capturing uncertain knowledge (Heckerman 1986; Pearl 1988). One study showed that most classes of uncertain knowledge cannot be represented efficiently in rule-based systems (Heckerman and Horvitz 1987). Other discussions demonstrated that probability theory uniquely satisfies a set of essential desirable properties of uncertain belief (Cox 1946; Tribus 1969; Horvitz et al. 1986).

Belief Networks and Influence Diagrams

Beyond theoretical analyses, the most important stimulus to renewed work on probability theory and SEU in AI was the development of efficient and expressive representations of probabilistic knowledge called *belief networks* (Pearl 1988; Howard 1989) and *influence diagrams*. Belief networks and influence diagrams allow people to express qualitative, in addition to quantitative, knowledge about beliefs, preferences, and decisions.

Belief Networks. A belief network is a graphical representation that allows a computer scientist, working with an expert, to efficiently encode expert knowledge about probabilistic dependencies among important distinctions in a domain. More important, belief networks allow an expert to specify *independence* in a domain. After the construction and assessment of a belief network, algorithms can be applied to the network to assign probabilities to alternative hypotheses, given a set of observations.

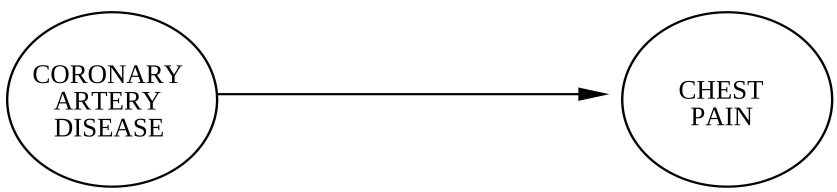

Figure 1. A small belief network that expresses the dependence of chest pain on coronary artery disease.

In the language of computer science, a belief network is a directed acyclic graph (DAG) that contains nodes representing propositions (for example, hypotheses and observations), and arcs representing probabilistic dependencies among nodes. Each node representing a proposition is associated with an exhaustive set of mutually exclusive values that represent alternative possible states or events.

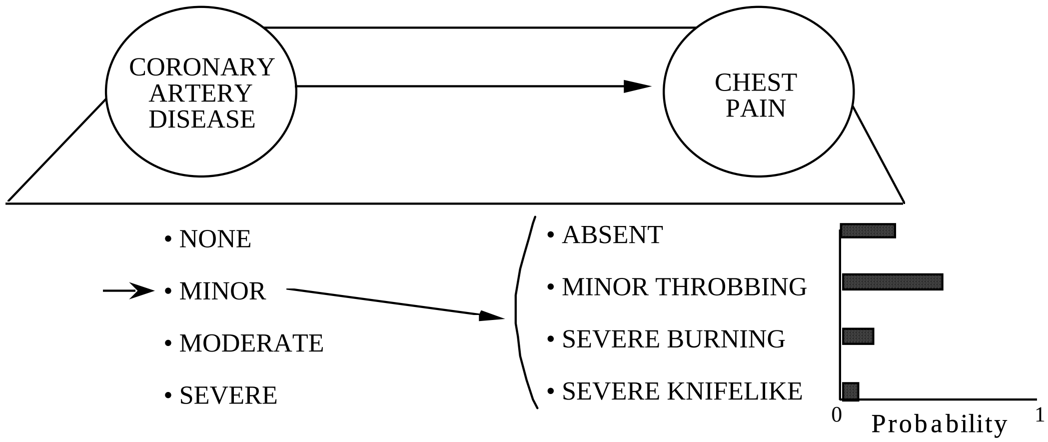

Figure 2. The qualitative dependency structure that defines the top-level structure of a belief network is captured by nodes and arcs. However, additional information lies beneath the top-level structure. For each node, there is a mutually exclusive and exhaustive set of values. Also, conditional probability distributions are assessed for each value of the predecessor node. Here, the probability of alternative forms of chest pain, given minor coronary artery disease, is depicted with a bar graph.

To construct a belief network, an expert defines nodes that represent important distinctions about the world. These distinctions include hypotheses, important states of interest (for example, diseases in a patient or disorders in a jet engine) that may not be confirmed directly, and observations that are useful for discriminating among the alternate hypotheses. The expert also provides information about the probabilistic dependencies among nodes.

Figure 1 displays a simple belief network, showing a dependency between coronary artery disease and chest pain. A directed arc from the node CORONARY ARTERY DISEASE to the node CHEST PAIN, means that an expert asserts that coronary artery disease affects the *probability distribution* over the possible values of chest pain. As indicated in Figure 2, beyond constructing the high-level dependency graph, an expert specifies the possible values of each node. For example, the expert might specify that coronary artery disease is either NONE, MINOR, MODERATE, or SEVERE, and that chest pain on exertion may be either ABSENT, MINOR THROBBING, SEVERE BURNING, or SEVERE KNIFELIKE. An arc between the two nodes means that the expert believes that the probabilities of having the different kinds of chest pain depend on the severity of the artery disease.

As part of building a belief network, we would need to assess the *conditional probability* that we would see the different types of chest pain when coronary artery disease was at each of its possible values. Given the direction of the arrows, conditional probabilities in this case have the form: "The probability of observation *A*, given the presence of hypothesis *H*, is *x*." We say that the probability of *A* is *conditioned* on *H*. As an example, for the sample cardiac belief network, the direction of the arrow indicates that we need to assess the probability that chest pain is knifelike given that coronary artery disease is minor." We write this conditional probability statement as, p(CHEST PAIN = KNIFELIKE | CORONARY ARTERY DISEASE = MINOR, ξ). The portion of the probability statement that follows the vertical line is called the *conditioning clause.* The symbol ξ in the clause refers to the general context, or *background state of information* that is not explicitly listed in the conditioning clause. Figure 2 uses a bar graph to denote the probability distribution over chest pain, given minor coronary artery disease. Figure 3 depicts the different probability distribution over different values of chest pain, given a patient has severe coronary artery disease.

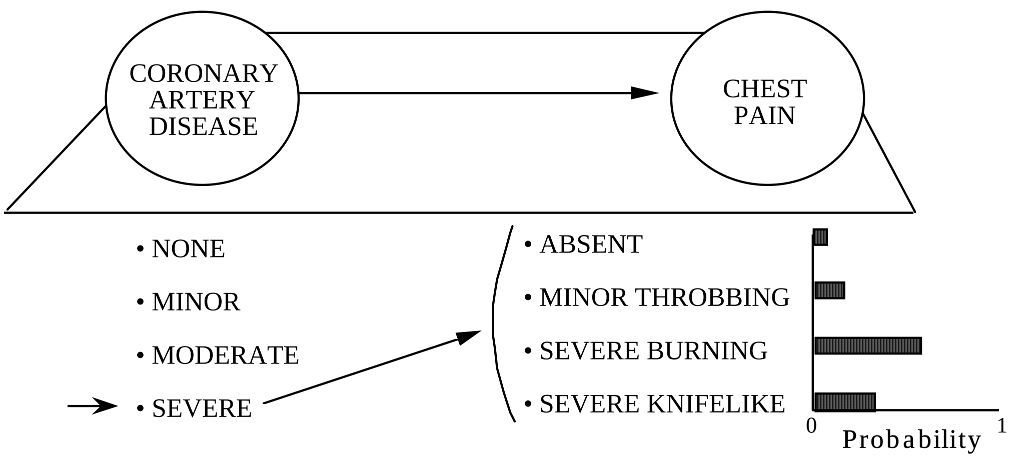

Figure 3. Separate distributions are stored for each value of the predecessor node. The probability distribution over different forms of chest pain, given that a patient is afflicted with severe coronary artery disease, is displayed with a bar graph.

Of course, belief networks can be more complicated. Figure 4 displays how we can represent multiple causal dependencies with a belief network. In this case, an expert diagnostician has added a node and arc representing his knowledge that gastric ulcer can lead to reports of chest pain. To assess the probability distributions over alternate degrees of chest pain, we must now consider all combinations of the values of gastric ulcer and coronary artery disease. In general, we can add many more manifestation and hypothesis

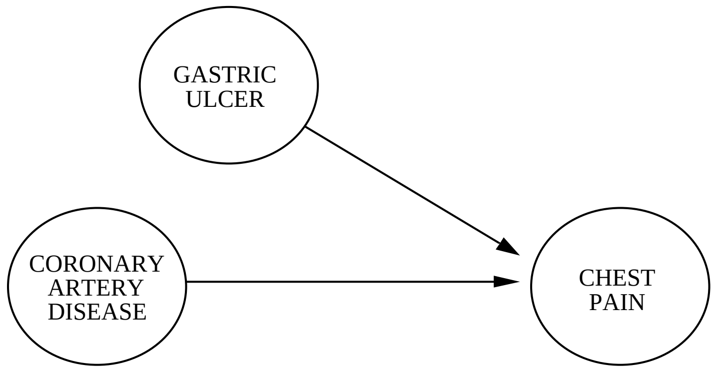

Figure 4. We can represent multiple causes with a belief network. Here, we capture a physician's knowledge that gastric ulcer, as well as coronary artery disease, can cause chest pain.

nodes to a belief network. Decision-support systems have been constructed which rely on belief networks consisting of hundreds of interconnected nodes.

Typically, we assess belief networks in the *causal* direction. That is, we draw arrows and assess conditional probabilities in terms of the likelihood of an observation, given a causal hypothesis. However, we often wish to use this information by performing inference in the *diagnostic* direction; that is, we would like to determine the probabilities of alternative hypotheses, given the observation (or assumption of the truth) of one or more symptoms. To do this, we apply belief-network inference algorithms. For example, we can apply an inference algorithm to determine the probability of various levels of severity of coronary artery disease when we hear the age of a patient and also hear a particular complaint about chest pain on exertion. A number of exact and approximate algorithms have been developed for revising the assignment of likelihoods to alternate hypotheses in a belief network, given observations of some evidence (see reviews in Pearl 1988 and Horvitz et al. 1988).

Influence Diagrams. Belief networks capture knowledge only about probabilistic dependencies among hypotheses and observations. An influence diagram (Howard and Matheson 1981; Olmsted 1983) is an extension of belief networks which, in addition to probabilistic-dependency information, represents alternative actions and outcomes, as well as information about a decision maker's preferences for different outcomes. The influence diagram is in many ways a more compact representation of the more familiar "decision-tree" representation of actions, events, and outcomes.

An influence diagram is a directed acyclic graph similar to a belief network. However, an influence diagram contains several types of nodes. Chance nodes, which are the only nodes in a belief network, define distinctions about states of the world. Decision nodes represent alternative actions; and value nodes represent the utility of different outcomes. For clarity, nodes are drawn as different shapes (typically, chance nodes are round, decision nodes are square, and value nodes are diamond-shaped).

Figure 5 displays an influence diagram that represents the problem of deciding whether to perform an angiogram test on a patient with chest pain. In determining the expected value of taking, versus not taking, the costly and potentially dangerous test, we consider how different outcomes of the test will affect a decision to have heart surgery. We also consider such factors as how the test, and future surgery will effect future chest pain, and the possibility of future myocardial infarction, or heart attack.

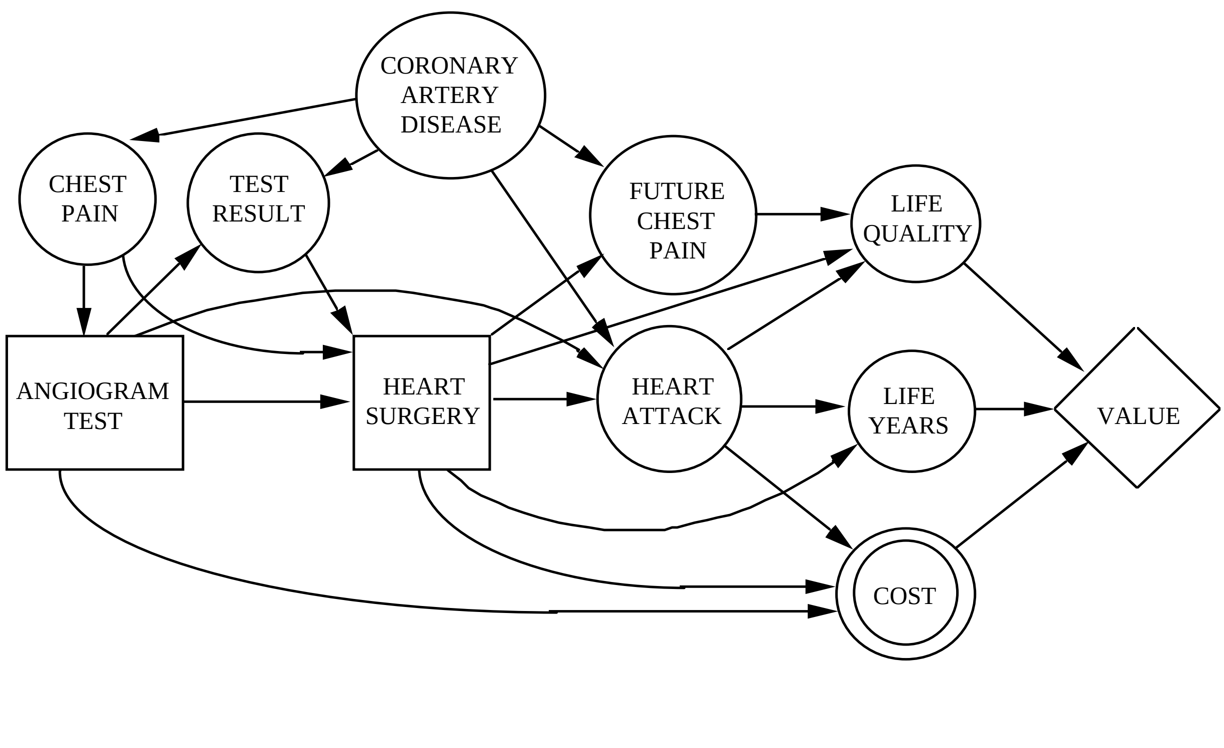

Figure 5. An influence diagram extends a belief network with decision nodes (squares), and a value node (diamond). Here, we represent the problem of whether or not a patient should receive an angiogram, given chest pain (from Horvitz, et al 1988).

All arcs point into a set of nodes that represent fundamental attributes of value to the patient. The value node represents a function that combines the value of the basic attributes into a scaler utility. This influence diagram also depicts another type of node, called a deterministic node (double-circled). The value of a deterministic node is a deterministic function of the inputs to the node. In this case, a physician has specified that the monetary cost is a deterministic function of the values of its predecessors (whether the angiogram is given, whether surgery is performed, and whether the patient has a myocardial infarction).

Algorithms have been developed that perform inference directly on influence diagrams through direct manipulation of the graphical structure of the representation (Shachter 1986; Shachter 1988). The output of such decision-theoretic inference is the best decision to make, given a decision maker's preferences, the uncertainties about the world, and the information that is available.

Acquiring Probabilities from Experts and from Data

Typically, the qualitative dependency relationships defined by the structure of a belief network or influence diagram require hundreds, or even thousands, of conditional probabilities to be gathered, or *assessed*, from experts. Some investigators have been concerned about the difficulty of acquiring these conditional probabilities, and a specific class of probabilities, called the *prior probabilities* of alternate hypotheses. A prior probability is the prevalence of a state of the world, conditioned only on the background state of information. For example, the prior probability, p(CORONARY ARTERY DISEASE = SEVERE|ξ), is the prevalence of severe coronary artery disease in a population.

Conditional and prior probabilities can be gathered through new statistical studies, or through extraction of data from preexisting studies. Most often, however, we find that the distinctions used in a belief network have not been defined and analyzed before. Thus there is little statistical information available on dependencies in the network. Moreover, we generally do not have the resources to gather information about the frequencies of thousands of events.

The subjective probability methodology has proved extremely successful for gathering the probabilities needed by an expert system. Our work on the construction of belief networks with expert diagnosticians from the fields of medicine, aviation, and electric-power generation has demonstrated that many experts can assess prior probabilities and conditional probabilities with confidence and ease.

One of the systems we have constructed, called PATHFINDER, performs pathology diagnosis given information about several hundred features that might be seen under a microscope (Heckerman et al. 1989; Heckerman et al. 1990; Horvitz et al. 1989b). PATHFINDER uses a belief network containing about 30,000 probabilities assessed from an expert pathologist with many years of experience. To construct a large knowledge base for the system's domain of lymph-node pathology, we assessed numbers from experts such as the probabilities that MONOCYTOID CELLS are ABSENT, MONOCYTOID CELLS are PRESENT, and MONOCYTOID CELLS are PROMINENT, given NODULAR-SCLEROSING HODGKIN'S DISEASE. Such assessments were made for all diseases in the domain. The assessment procedures begin only after such distinctions as ABSENT, PRESENT, and PROMINENT for monocytoid cells are defined precisely. Assessments can be more complex: In cases where an expert has identified dependencies among features, probabilities must be conditioned on the values of other features, in addition to the presence of a disease.

Studies with the chief expert on the PATHFINDER project showed that probability assessments are quite stable and repeatable. Perhaps more important, we found that the diagnostic accuracy of the system is not very sensitive to small variations in individual conditional or prior probabilities. Typically, it is the *order of magnitude* of a probability that is most important. Regarding error in prior probabilities, we have found that for diagnosing disorders within the pathology application area, the influence of a prior probability is often dominated quickly by probabilistic updates with evidence. That is, small errors in prior probabilities are quickly "washed out" by the updates based on observations. PATHFINDER has been shown in formal trials to perform at the level of expert pathologists. More detailed clinical trials are under way to ascertain the value of PATHFINDER as a diagnostic aid for non-experts.

Well-defined techniques have been developed for helping experts with the assessment of probabilities for use in belief networks. For example, methods have been developed (Spetzler and Stael von Holstein 1975) to help experts avoid biases in probability assessment that have been identified by cognitive psychologists (Tversky and Kahneman 1974). Other methods test an expert's consistency in probability assessment. The expert's confidence can be measured in terms of a *second-order probability distribution*, a probability distribution that describes the probability of alternative probabilities. Techniques have been developed to simplify the assessment of second-order distributions and for ideally updating probabilities given new data. Also, software tools have been developed that allow an expert to avoid redundancies in the assessment of similar probabilities. These tools enable a user to combine diseases into groups and to perform assessment at greater levels of abstraction. For example, Heckerman recently developed a graphical software tool that streamlines the probability-assessment process through abstraction and pairwise comparisons (Heckerman 1990).

Because subjectivist and frequentist probability are defined by the same system of axioms, we can refine an expert's initial assessment by performing statistical studies. There are well-characterized techniques for updating an expert's beliefs with statistical data, based on the assessment of the expert's confidence in probabilities, in addition to the probabilities themselves. Furthermore, we can apply formal cost-benefit analyses to direct the collection of better subjective assessments or to supplement the subjective probabilities with statistical information, depending on the importance of the probabilities and on the confidence of the expert. Such analyses draw on techniques in decision science known as "sensitivity analyses." Sensitivity analyses are used to identify the probabilities that have the greatest impact on the accuracy of a system's conclusions.

Belief-Network Tools and Applications

Belief networks hold promise as a representation for helping biologists to acquire and reason with knowledge about the likely identity of organisms given incomplete sets of observed features. Belief networks also can provide a coherent means of integrating diverse classes of knowledge to solve taxonomic problems. For example, investigators might use belief networks to reason about the relative likelihood of alternative phylogenetic trees by considering genetic, embryogenetic, ontogenetic, and paleontologic data.

Belief-network programming shells (Andreassen et al. 1987), as well as stable expert systems, have been produced commercially, ranging in application from sleep disorders to jet-engine repair (Henrion et al. 1991). Perhaps the most widespread application has been in the area of pathology, stemming from the work on PATHFINDER. Pathologists at over 300 sites throughout the world are using multimedia expert systems that are based on belief networks integrated with videodisc systems. These systems allow pathologists to employ one of several belief networks, each representing the leading expertise on the histological analysis of tissue from a major human organ system.

Interest in belief networks has been growing in a variety of areas of AI, beyond diagnostic problem solving. For example, there has been recent work in the field of machine learning on the automatic construction of belief networks from data. Belief networks have also been used recently in systems for performing visual identification (Binford et al. 1989) and for understanding natural language and identifying plans (Charniak and Goldman 1989).

Other Developments

The last decade has seen a broad spectrum of advances in automated reasoning methods, beyond techniques for reasoning under uncertainty. I will briefly mention a few areas of research with application to systematic biology.

Given the explosion of biomedical literature and information, it is becoming difficult and time-consuming for investigators to access and integrate information that is relevant to their problems. Thus, tools for managing this information and helping biologists, expecially systematists, to search through a number of large knowledge bases will become increasingly useful. Techniques have recently been developed to assist researchers in the design of databases (Barsalou 1989). Work has also increased on methods for performing intelligent searches in databases (Frisse 1988). In this realm, tools have been developed that allow a user to search more efficiently through disparate knowledge bases and to generate query results that take into account the preferences and interests of that user.

Another area of research with relevance to the study of biological processes is work on *qualitative reasoning.* In this area, investigators have worked to capture the essential aspects of physical systems with abstract qualitative models (Forbus 1981). These models are used to predict behaviors of a system without using numerical information, by considering the orders of magnitude of effects, and modeling the competition and synergy between basic influences in a system. Several projects have focused on the construction of qualitative models for reasoning about biological systems (Karp 1988). Related research has also been done on qualitative probabilistic reasoning with belief networks. This work centers on abstracting the networks to qualitative networks that consider only positive and negative influences on probabilities (Wellman 1988).

Exciting research is under way on computer-based methods for understanding and assisting scientific investigation. Studies of computers in scientific reasoning include work on tools for assisting with the discovery of new theories and with validation, or *confirmation,* of competing theories of scientific phenomena. For example, in recent work, computer models and programs have been developed to perform probabilistic *meta-analyses* (Lehmann 1991). Such analyses supplement standard statistical analyses to determine the effects of an experimental outcome on the probability of alternate hypotheses. In other work, planning techniques have been applied to the task of designing experiments in molecular biology (Friedland 1979; Stefik 1981). That is, given the experimental goals, a computer-based planner determines a sequence of steps that will perform the analysis.

Another area of growing interest is the control of tradeoffs in reasoning systems. The goal of this work is to optimize the value of computing systems, given limitations in the time available for computation imposed by a scientific investigator, a physician, or an industrial process. Several studies have examined methods for trading off the precision or accuracy of a computational result in exchange for the timeliness of that result (Horvitz 1987; Dean and Boddy 1988). By applying decision-analytic techniques to control computation in belief networks and influence diagrams, we can develop a model of rational belief and action under bounded reasoning resources. As an example of this work, we explored a computational model of bounded rationality with the PROTOS system (Horvitz et al. 1989a; Horvitz and Rutledge 1991). PROTOS considers the ideal tradeoff between the precision of probabilistic inference and the timeliness of action for medical intensive-care decision problems.

CONCLUSION

I have reviewed advances in automated reasoning that I believe are most relevant to the problems that challenge systematists and biomedical scientists in related disciplines. I organized the presentation of fundamental areas of research in terms of the major research themes that have characterized three decades of AI research. In my attempt to make the best use of the few pages allotted to this chapter, I have unfortunately had to omit broad areas of research on automated reasoning. I trust readers will seek more comprehensive understanding of the topics they find most interesting.

References

(Aho et al. 1983)
Aho, A., Hopcroft, J., and Ullman, J. (1983). *Data Structures and Algorithms*. Addison-Wesley, Menlo Park, California.

(Amarel 1968)
Amarel, S. On representations of problems of reasoning about actions. *Machine Intelligence 3*, pages 131-171.

(Andreassen et al. 1987)
Andreassen, S., Woldbye, M., Falck, B., and Andersen, S., MUNIN: A causal probabilistic network for interpretation of electromyographic findings. In *Proceedings of the Tenth International Joint Conference on Artificial Intelligence*, Milan, Italy, pages 366-372. Morgan Kaufmann, San Mateo, CA.

(Barr and Feigenbaum 1982)
Barr, A. and Feigenbaum, E., *The Handbook of Artificial Intelligence*, volume 2. William Kaufmann.

(Barsalou 1989)
Barsalou, T. An object-based architecture for biomedical expert database systems. In *Proceedings of the Twelveth Symposium on Computer Applications in Medical Care*, Washington, DC. IEEE Computer Society Press, Los Angeles, CA.

(Berliner 1980)
Berliner, H. J., Backgammon program beats world champion. *Artificial Intelligence*, 14:205-220.

(Binford et al. 1989)
Binford, T., Levitt, T., and Mann, W. Bayesian inference in model-based machine vision. In Kanal, L., Lemmer, J., and Levitt, T., editors, Uncertainty in Artificial Intelligence 3, pages 73--96. North Holland, New York.

(Brinkley 1985)
Brinkley, J.F. Knowledge-driven ultrasonic three-dimensional organ modelling, *IEEE Trans. Pattern Analysis and Machine Intelligence* , 7(4): 431-441.

(Buchanan and Shortliffe 1984)
Buchanan, B. and Shortliffe, E., editors, *Rule-Based Expert Systems: The {MYCIN Experiments of the Stanford Heuristic Programming Project.*
Addison-Wesley, Reading, MA.

(Buchanan 1982)
Buchanan, B. G., Research on expert systems.
In J. Hayes, D. Michie, Y. H. P., editor, *Machine Intelligence*,
10:269-299. Ellis Howard Ltd., Chichester, England.

(Charniak and Goldman 1989)
Charniak, E. and Goldman, R., A semantics for probabilistic quantifier-free first-order languages, with particular application to story understanding. In *Proceedings of the Eleventh International Joint Conference on Artificial Intelligence*, Detroit, Michigan. International Joint Conferences on Artificial Intelligence.

(Christofides 1979)
Christofides, N., The travelling salesman problem. N. Christofides, M. Mingozzi, P. Toth, C. Sandi, eds. *Combinatorial Optimization*, Wiley, Chichester, pages 131-149.

(Cohen 1977)
Cohen, B., The mechanical discovery of certain problem properties. *Artificial Intelligence*, 8:119-131.

(Cohen and Feigenbaum 1982)
Cohen, P. and Feigenbaum, E., *The Handbook of Artificial Intelligence*, volume 2. William Kaufmann.

(Cox 1946)
Cox, R., Probability, frequency and reasonable expectation. *American Journal of Physics*, 14:1-13.

(Dantzig 1963)
Dantzig, G.B., *Linear Programming and Extensions*, Princeton University Press, Princeton, New Jersey.

(Dantzig and Fulkerson 1954)
Dantzig, G.B. and Fulkerson, D.R., *Minimizing the number of tankers to meet a fixed schedule*. Naval Research Logistics Quarterly (1):217:222.

(Dantzig, et al 1954)
Dantzig, G.B., Fulkerson, D.R., and Johnson, S.M. Solution of a large-scale traveling-salesman problem. *Operations Research (2):393-410.*

(Davis 1982)
Davis, R., Consultation, knowledge acquisition, and instruction. In Szolovits, P., editor, *Artificial Intelligence In Medicine*, pages 57-78. Westview Press, Boulder, CO.

(Davis et al. 1977)
Davis, R., Buchanan, B., and Shortliffe, E., Production rules as a representation for a knowledge-based consultation program. *Artificial Intelligence,* 8:15-45.

(Dean and Boddy 1988).
Dean, T. and Boddy, M., An analysis of time-dependent planning. In *Proceedings AAAI-88 Seventh National Conference on Artificial Intelligence*, pages 49-54. American Association for Artificial Intelligence.

(Feigenbaum 1964)
Feigenbaum, E., *Computers and Thought*.. McGraw-Hill, New York.

(Fikes and Nilsson 1971)
Fikes, R. E. and Nilsson, N. J., STRIPS: a new approach to the application of theorem proving to problem solving. *Artificial Intelligence*, 2:189-208.

(Friedland 1979)
Friedland, P. E., *Knowledge-based experiment design in molecular genetics.* PhD thesis, Computer Science Department, Stanford University. Rep. No. STAN-CS-79-771.

(Frisse 1988)
Frisse, M., Searching for information in a hypertext medical handbook. *Communications of the Association for Computing Machinery*, 31(7):880-888.

(Garey and Johnson 1979)
Garey, M. and Johnson, D., *Computers and Intractability: A Guide to the Theory of NP-Completeness*. W.H. Freeman and Company, New York.

(Gelernter 1959)
Gelernter, H., Realization of a geometry theorem-proving machine. In *Proceedigns of the International Conference on Information Processing*, Paris, pages 273-282. UNESCO House.

(Good 1977)
Good, I., Dynamic probability, computer chess, and the measurement of knowledge. In Elcock, E. and D., M., editors, *Machine Intelligence*, pages 139-150. Wiley, New York.

(Gorry 1973)
Gorry, G., Computer-assisted clinical decision making. Methods of Information in Medicine, 12:45-51.

(Gorry and Barnett 1968)
Gorry, G. and Barnett, G., Experience with a model of sequential diagnosis. *Computers and Biomedical Research*, 1:490-507.

(Gorry et al. 1973)
Gorry, G., Kassirer, J., Essig, A., and Schwartz, W., Decision analysis as the basis for computer-aided management of acute renal failure. *American Journal of Medicine*, 55:473-484.

(Hacking 1975)
Hacking, I., *The Emergence of Probability*. Cambridge University Press, Cambridge.

(Heckerman 1986)
Heckerman, D., Probabilistic interpretations for MYCIN's certainty factors. In Kanal, L. and Lemmer, J., editors, *Uncertainty in Artificial Intelligence*, pages 167-196. North Holland, New York.

(Heckerman 1990)
Heckerman, D., *Probabilistic Similarity Networks*. PhD thesis, Stanford University, Stanford, CA.

(Heckerman and Horvitz 1987)
Heckerman, D. and Horvitz, E., On the expressiveness of rule-based systems for reasoning under uncertainty. In *Proceedings AAAI-87 Sixth National Conference on Artificial Intelligence*, Seattle, WA, pages 121-126. Morgan Kaufmann, San Mateo, CA.

(Heckerman and Horvitz 1988)
Heckerman, D. and Horvitz, E., The myth of modularity in rule-based systems. In Lemmer, J. and Kanal, L., editors, *Uncertainty in Artificial Intelligence 2*, pages 23-34. North Holland, New York.

(Heckerman et al. 1989)
Heckerman, D., Horvitz, E., and Nathwani, B., Update on the Pathfinder project. In *Proceedings of the Thirteenth Symposium on Computer Applications in Medical Care*, Washington, DC, pages 203--207. IEEE Computer Society Press, Los Angeles, CA.

(Heckerman et al. 1990)
Heckerman, D., Horvitz, E., and Nathwani, B., Toward normative expert systems: The Pathfinder project. *Methods of Infomation in Medicine* 1992. *In press*.

(Heckerman and Miller 1986)
Heckerman, D. and Miller, R., Towards a better understanding of the INTERNIST-1 knowledge base. In *Proceedings of Medinfo*, Washington, DC, pages 27-31. North Holland, New York.

(Henrion et al. 1991)
Henrion, M., Breese, J., Horvitz, E., Decision Analysis and Expert Systems. *AI Magazine* 12(4):64-91. AAAI Press, Menlo Park, CA.

(Horvitz,1987).
Horvitz, E., Reasoning about beliefs and actions under computational resource constraints. In *Proceedings of Third Workshop on Uncertainty in Artificial Intelligence*, Seattle, WA, July 1987, pp. 429-444. Association for Uncertainty and Artificial Intelligence. Also in L. Kanal, T. Levitt, and J. Lemmer, eds., *Uncertainty in Artificial Intelligence 3*, Elsevier 1989, pages 301-324.

(Horvitz 1988)
Horvitz, E., Reasoning under varying and uncertain resource constraints. In *Proceedings AAAI-88 Seventh National Conference on Artificial Intelligence*, Minneapolis, MN, pages 111-116. Morgan Kaufmann, San Mateo, CA.

(Horvitz et al. 1988)
Horvitz, E., Breese, J., and Henrion, M., Decision theory in expert systems and artificial intelligence. *International Journal of Approximate Reasoning*, 2:247-302.

(Horvitz et al. 1989a)
Horvitz, E., Cooper, G., and Heckerman, D., Reflection and action under scarce resources: Theoretical principles and empirical study. In *Proceedings of the Eleventh International Joint Conference on Artificial Intelligence,* pages 1121-1127. International Joint Conference on Artificial Intelligence.

(Horvitz and Heckerman 1986)
Horvitz, E. and Heckerman, D., The inconsistent use of measures of certainty in artificial intelligence research. In Kanal, L. and Lemmer, J., editors, *Uncertainty in Artificial Intelligence*, pages 137--151. North Holland, New York.

(Horvitz et al. 1986)
Horvitz, E., Heckerman, D., and Langlotz, C., A framework for comparing alternative formalisms for plausible reasoning. In *Proceedings AAAI-86 Fifth National Conference on Artificial Intelligence*, Philadelphia, PA, pages 210--214. Morgan Kaufmann, San Mateo, CA.

(Horvitz et al. 1989b)
Horvitz, E., Heckerman, D., Ng, K., and Nathwani, B., Heuristic abstraction in the decision-theoretic Pathfinder system. In *Proceedings of the Thirteenth Symposium on Computer Applications in Medical Care*, Washington, DC, pages 178--182. IEEE Computer Society Press, Los Angeles, CA.

(Horvitz 1990)
Horvitz, E.J., *Computation and Action Under Bounded Resources*, Ph.D. Thesis, Stanford University 1990.

(Horvitz and Rutledge 1991)
Horvitz, E. and Rutledge, G., Time-dependent utility and action under uncertainty. In *Proceedings of Seventh Conference on Uncertainty in Artificial Intelligence*, Los Angeles, CA. Morgan Kaufmann, San Mateo, CA.

(Howard 1966)
Howard, R., Decision analysis: Applied decision theory. In Hertz, D. and Melese, J., editors, In *Proceedings of the Fourth International Conference on Operational Research*, pages 55-77.

(Howard 1968)
Howard, R., The foundations of decision analysis. *IEEE Transactions on Systems Science, and Cybernetics*, 4:211-219.

(Howard 1989)
Howard, R., Knowledge maps. *Management Science*, 35(8):903--922.

(Howard and Matheson 1981)
Howard, R. and Matheson, J., Influence diagrams., In Howard, R. and Matheson, J., editors, *Readings on the Principles and Applications of Decision Analysis*, volume 2, pages 721-762. Strategic Decisions Group, Menlo Park, CA.

(Karp 1988)
Karp, P. , A process-oriented model of bacterial gene regulation. Technical Report KSL-90-75, Knowledge Systems Laboratory, Stanford University, Stanford, CA.

(Keeney and Raiffa 1976)
Keeney, R. and Raiffa, H., *Decisions with Multiple Objectives: Preferences and Value Tradeoffs*.Wiley and Sons, New York.

(Korf 1980)
Korf, R., Towards a model of representation change. *Artificial Intelligence,* 14(1):41-78.

(Lawler et al. 1985)
Lawler, E., Lenstra, J., Kan, A. R., and Shmoys, D., *The Traveling Salesman Problem*. John Wiley and Sons, New York.

(Ledley and Lusted 1959) Ledley, R. and Lusted, L., Reasoning foundations of medical diagnosis. *Science,* 130:9-21.

(Lindberg et al. 1980) Lindberg, D., Sharp, G., Kingsland, L., Weiss, S., Hayes, S., Ueno, H., and Hazelwood, S., Computer-based rheumatology consultant. In *Proceedings of Medinfo*, pages 1311-1315. North Holland, New York.

(Lindsay et al. 1980) Lindsay, R., Buchanan, B. G., Feigenbaum, E. A., and Lederberg, J., *DENDRAL*. McGraw-Hill, New York.

(Marr 1975)
Marr, D., Analyzing natural images: A computational theory of texture vision. *Technical Report Artificial Intelligence Memo 334*, M.I.T. Artificial Intelligence Laboratory, Cambridge, MA.

(Marr 1982)
Marr, D., *Vision*, W.H. Freeman, San Francisco, CA.

(Marr and Hildreth 1980)
Marr, D. and Hildreth, E. Theory of edge detection. *Proceedings of the Royal Society B,* 207:187-280.

(Miller et al. 1986)
Miller, R., McNeil, M., Challinor, S., Masarie, F., and Myers, J., The INTERNIST-1/Quick Medical Reference project--status report. *Western Journal of Medicine,* 145:816-822.

(Miller et al. 1982)
Miller, R., Pople, E., and Myers, J. INTERNIST-1: An experimental computer-based diagnostic consultant for general internal medicine. New England Journal of Medicine, 307:476-486.

(Minsky 1975)
Minsky, M., A framework for representing knowledge. In Winston, P., editor, *The Psychology of Computer Vision*, pages 211-277. McGraw-Hill, New York.

(Newell 1958)
Newell, A., *Report on a general problem-solving program*. Technical Report P-1584, Rand Corporation, December 1958.

(Newell and Simon 1963)
Newell, A. and Simon, H., GPS, a program that simulates human thought. In *Computers and Thought*. McGraw-Hill, New York.

(Nilsson 1980)
Nilsson, N., *Principles of Artificial Intelligence*. Tioga Press, Los Altos, CA.

(Olmsted 1983)
Olmsted, S., *On Representing and Solving Decision Problems*. PhD thesis, Department of Engineering-Economic Systems, Stanford University.

(Papadimitriou and Steiglitz 1982)
Papadimitriou, C. and Steiglitz, K., *Combinatorial Optimization: Algorithms and Complexity*. Prentice-Hall, Inc., Englewood Cliffs, New Jersey.

(Pauker et al. 1976)
Pauker, S., Gorry, G., Kassirer, J., and Schwartz, W., Toward the simulation of clinical cognition: Taking a present illness by computer. *American Journal of Medicine*, 60:981-995.

(Pearl 1988)
Pearl, J., *Probabilistic Reasoning in Intelligent Systems:*
Networks of Plausible Inference. Morgan Kaufmann, San Mateo, CA.

(Pople 1982)
Pople, H., Heuristic methods for imposing structure on ill-structured problems: The structuring of medical diagnostics. In Szolovits, P., editor, *Artificial Intelligence in Medicine*, pages 119-190. Westview Press, Boulder, CO.

(Raiffa and Schlaifer 1961) Raiffa, H. and Schlaifer, R., *Applied Statistical Decision Theory*. Harvard University Press, Cambridge, MA.

(Russell and Wefald 1989),
Russell, S. and Wefald, E.H. Principles of metareasoning.
In Brachman, R.J., Levesque, H.J., and Reiter, R., editors, *Proceedings of the First International Conference on Principles of Knowledge Representation and Reasoning*}, Toronto. Morgan Kaufmann.

(Samuel 1959)
Samuel, A.L., Some studies of machine learning using the game of checkers. *IBM Journal of Research and Development*, 3:211-229.

(Samuel 1967)
Samuel, A.L., Some studies of machine learning using the game of checkers II--recent progress. *IBM Journal of Research and Development*, 11(6):601-617.

(Savage 1972)
Savage, L., *The Foundations of Statistics*. Dover, New York. 2nd edition, First edition 1954.

(Shachter 1986)
Shachter, R. Evaluating influence diagrams. *Operations Research*, 34:871-882.

(Shachter 1988)
Shachter, R., Probabilistic inference and influence diagrams. *Operations Research*, 36:589--604.

(Shortliffe 1976)
Shortliffe, E., Computer-based Medical Consultations: MYCIN. North Holland, New York.

(Shortliffe and Buchanan 1975)
Shortliffe, E. and Buchanan, B., A model of inexact reasoning in medicine. *Mathematical Biosciences*, 23:351-379.

(Simon 1969)
Simon, H., *The Sciences of the Artificial*. M.I.T. Press, Cambridge, Massachusetts.

(Simon 1972)
Simon, H., The theory of problem solving. *Information Processing*, 71:261-277.

(Simon 1987)
Simon, H., Two heads are better than one: The collaboration between AI and OR. *Interfaces*, 17:8--15.

(Spetzler and Stael von Holstein 1975)
Spetzler, C. and Stael von Holstein, C., Probability encoding in decision analysis. *Management Science*, 22:340-358.

(Stefik 1981)
Stefik, M. J., Planning with constraints: Molgen, Part 1. *Artificial Intelligence*, 16(2):111-139.

(Szolovits 1982)
Szolovits, P., Artificial intelligence in medicine. In Szolovits, P., editor, *Artificial Intelligence in Medicine*, pages 1-19. Westview Press, Boulder, CO.

(Tribus 1969)
Tribus, M., *Rational Descriptions, Decisions, and Designs*. Pergamon Press, New York.

(Tversky and Kahneman 1974)
Tversky, A. and Kahneman, D., Judgment under uncertainty: Heuristics and biases. *Science*, 185:1124-1131.

(Ullman 1979)
Ullman, S., *Interpretation of visual motion.* M.I.T. Press, Cambridge, MA.

(von Neumann and Morgenstern 1947)
von Neumann, J. and Morgenstern, O., *Theory of Games and Economic Behavior*. Princeton University Press, Princeton, NJ.

(Warner et al. 1961)
Warner, H., Toronto, A., Veasy, L., and Stephenson, R., A mathematical approach to medical diagnosis: Application to congenital heart disease. *Journal of the American Medical Association*, 177:177-183.

(Weiss et al. 1978)
Weiss, J., Kulikowski, C., Amarel, S., and Safir, A., A model-based method for computer-aided medical decision-making. *Artificial Intelligence*, 11:145--172.

(Wellman 1988)
Wellman, M., *Formulation of Tradeoffs in Planning Under Uncertainty*. PhD thesis, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA.

(Winston 1975)
Winston, P., *The Psychology of Computer Vision*. McGraw-Hill, New York.