Decision Theory in Expert Systems and Artificial Intelligence*

Eric J. Horvitz
Medical Computer Science Group
Knowledge Systems Laboratory
Stanford University
Stanford, California 94305

John S. Breese
Rockwell International Science Center
Palo Alto Laboratory
444 High Street
Palo Alto, CA 94301

Max Henrion
Department of Engineering and Public Policy
Carnegie Mellon University
Pittsburgh, PA 15213

July 1988

Abstract

Despite their different perspectives, artificial intelligence (AI) and the disciplines of decision science have common roots and strive for similar goals. This paper surveys the potential for addressing problems in representation, inference, knowledge engineering, and explanation within the decision-theoretic framework. Recent analyses of the restrictions of several traditional AI reasoning techniques, coupled with the development of more tractable and expressive decision-theoretic representation and inference strategies, have stimulated renewed interest in decision theory and decision analysis. We describe early experience with simple probabilistic schemes for automated reasoning, review the dominant expert-system paradigm, and survey some recent research at the crossroads of AI and decision science. In particular, we present the belief network and influence diagram representations. Finally, we discuss issues that have not been studied in detail within the expert-systems setting, yet are crucial for developing theoretical methods and computational architectures for automated reasoners.

*This is the technical report version of a paper appearing in the *International Journal of Approximate Reasoning*, 2:247–302, 1988 (Special Issue on Uncertainty in Artificial Intelligence). This work was supported by a NASA Fellowship to Eric Horvitz under Grant NCC-220-51 to Stanford University, by the National Science Foundation under Grant IRI-8703710, and by the National Library of Medicine under Grant R01LM04529. Max Henrion is partially supported by National Science Foundation Grant IST-8603493 to Carnegie-Mellon University. Computing facilities were provided by the SUMEX-AIM Resource under NIH Grant RR-00785.

Contents

List of Figures

1 Introduction

Reasoning about action under incomplete information and scarce resources is central to solving difficult problems in artificial intelligence (AI). The time is ripe for AI to apply and extend techniques developed in decision science for addressing resource allocation and decision making under uncertainty. By *decision science*, we mean Bayesian probability and decision theory, the study of the psychology of judgment, and their practical application in operations research and decision analysis. In particular, decision theory can provide a valuable framework for addressing some of the foundational problems in AI, and forms the basis for a range of practical tools.

Artificial intelligence and the decision sciences emerged from research on systematic methods for problem solving and decision making that blossomed in the 1940s. These disciplines were stimulated by new possibilities for automated reasoning unleashed by the development of the computer. Although the fields had common roots, AI soon distinguished itself from the others in its concern with autonomous problem solving, its emphasis on symbolic rather than numeric information, its use of declarative representations, and its interest in analogies between computer programs and human thinking.

Some of the earliest AI research centered on an analysis of the sufficiency of alternative approximation strategies and heuristic methods to accomplish the task of more complex decision-theoretic representation and inference [137]. However, many AI researchers soon lost interest in decision theory. This disenchantment seems to have arisen, in part, from a perception that decision-theoretic approaches were hopelessly intractable and were inadequate for expressing the rich structure of human knowledge [51, 144]. This view is reflected in a statement by Szolovits, a researcher who had investigated the application of decision theory in early medical reasoning systems: "The typical language of probability and utility theory is not rich enough to discuss such complex medical issues, and its extension within the original spirit leads to untenably large decision problems" (Szolovits [144], p. 7).

Although similar views are still widespread among AI researchers, there has been a recent resurgence of interest in the application of probability theory, decision theory, and decision analysis to AI. In this paper, we examine some of the reasons for this renewed interest, including an increasing recognition of the shortcomings of some traditional AI methods for inference and decision making under uncertainty, and the recent development of more expressive decision-theoretic representations and more practical knowledge-engineering techniques.

The potential contributions of decision science for tackling AI problems derive from decision science's explicit theoretical framework and practical methodologies for reasoning about decisions under uncertainty. Decisions underlie any action that a problem solver may take in structuring problems, in reasoning, in allocating computational resources, in displaying information, or in controlling some physical activity. As AI has moved beyond toy problems to grapple with complex, real-world decisions, adequate treatment of uncertainty has become increasingly important. Attempts to build systems in such areas as medicine, investment, aerospace, and military planning have uncovered the ubiquity of uncertainty associated with incomplete models. The discovery of useful heuristic procedures can be quite difficult in these complex domains.

Work in real-world domains has also uncovered the importance of reasoning about actions and human values. Traditional AI research has paid little attention to the modeling of complex preferences and attitudes toward risk. These issues are central in decision science. We believe that the correspondence of concerns with AI and decision theory will become increasingly obvious as investigators begin to enter complex real-world domains that are dominated by uncertainty and high stakes.

We discuss research that applies concepts and techniques from probability, decision theory, and decision analysis to problems in AI. After outlining key ideas in decision science, we explore advances in the use of decision-theoretic ideas in diagnostic expert systems. We examine initial probabilistic approaches to expert systems and the heuristic approaches that achieved prominence in the mid-1970*s*. We then review current research on the use of decision-theoretic concepts in expert systems, including representation, knowledge engineering, tractable inference, and explanation. Finally, we move beyond expert systems to survey applications of decision theory to a variety of topics in AI research.

Although some researchers have made noteworthy progress, much of this work is still in its initial stages. It remains to be seen how effective these approaches can be in addressing long-standing problems in AI. It would be premature to attempt a comprehensive assessment of the eventual influence of this area of research. Our purpose here is simply to present a perspective on central issues and to provide pointers to promising avenues of research.

2 Foundations

> [T]he theory of probability is no more than a calculus of good sense. By this theory, we learn to appreciate precisely what a sound mind feels through a kind of intuition often without realizing it. The theory leaves nothing arbitrary in choosing opinions or in making decisions, and we can always select, with the help of this theory, the most advantageous choice on our own. It is a refreshing supplement to the ignorance and feebleness of the human mind. (Laplace,[94], p. 196).

The foundations of probability theory extend at least as far back as the seventeenth century in the works of Pascal, Bernoulli, and Fermat. Probability provides a language for making statements about uncertainty and thus makes explicit the notion of partial belief and incomplete information. Decision theory extends this language to allow us to make statements about what alternative actions are and how alternative outcomes (the results of actions) are valued relative to one another. Probability theory and the more encompassing decision theory provide principles for rational inference and decision making under uncertainty. These theoretical ideas, however, tell us little about how to apply these principles to real problems in a tractable manner; this is the realm of decision analysis. In this section, we review central concepts of Bayesian probability theory, decision theory, and decision analysis. Our intent is not to provide a comprehensive review of these topics. Rather, we wish to highlight several key ideas.

2.1 Probability Is a Measure of Personal Belief

A Bayesian or *subjectivist* views the probability of an event as a measure of a person's degree of belief in the event, given the information available to that person. A probability of 1 corresponds to belief in the absolute truth of a proposition, a probability of zero to belief in the proposition's negation, and intervening values to partial belief or knowledge. According to this perspective, probabilities are properties of the state of knowledge of an individual rather than properties of a sequence of events (e.g., tosses of a "fair" coin). This approach generalizes the classical notion of a probability as a long-run frequency of a "repeatable" event. A subjectivist also is willing to assign belief to unique events that are not members of any obvious repeatable sequence of events (e.g., the probability that we will finish the manuscript this week). The assignment of a subjective probability should be based on all information available to an individual, including those items that are known to be true or deducible in a logical sense, as well as empirical frequency information.

A single real number is used to represent the belief that an agent has in the truth of a proposition. Subjectivists often draw attention to the state of information on which a probability is based, or *conditioned*, by specifying it explicitly. The notation for the probability of a proposition or event X conditioned on a state of information ξ may be specified as $p(X|\xi)$. To make it clear that any belief assignment is based on background knowledge, we explicitly include ξ in the conditioning statement. Thus, the revised probability of X given a new piece of evidence E is written $p(X|E,\xi)$, where the comma denotes the conjunction of E and ξ.

Subjective probabilities abide by the same set of axioms as do classical probabilities or frequencies. The axioms are rules for the consistent combination of probabilities for related events. A classic axiomatization of probability contains the following definitions:[1]

$$0 <= p(X|\xi) <= 1$$

$$p(X|\xi) + p(not\ X|\xi) = 1$$

$$p(X\, or\, Y|\xi) = p(X|\xi) + p(Y|\xi) - p(X\, and\, Y|\xi)$$

$$p(X\, and\, Y|\xi) = p(X|Y,\xi)p(Y|\xi)$$

Sets of belief assignments that are consistent with the axioms of probability theory are said to be *coherent*. In this sense, the theory provides a consistency test for uncertain beliefs. Persuasive examples suggest that a rational person would wish to avoid making decisions based on incoherent beliefs. For example, someone willing to bet according to incoherent beliefs would be willing to accept a "Dutch book"–that is, a combination of bets leading to guaranteed loss under any outcome [96, 134].

2.2 Probability Is Sufficient for Representing Uncertainty

A number of researchers have provided lists of fundamental properties that they consider intuitively desirable for continuous measures of belief in the truth of a proposition [25, 149, 100]. A recent reformulation of desirable properties of belief is [75]:

[1] Several axiomatizations of probability theory have been proposed.

1. *Clarity*: Propositions should be well defined.
2. *Scalar continuity*: A single real number is both necessary and sufficient for representing a degree of belief in a proposition.
3. *Completeness*: A degree of belief can be assigned to any well-defined proposition.
4. *Context dependency*: The belief assigned to a proposition can depend on the belief in other propositions.
5. *Hypothetical conditioning* : There exists some function that allows the belief in a conjunction of propositions, B(X and Y), to be calculated from the belief in one proposition and the belief in the other proposition given that the first proposition is true. That is,

$$B(X\, and\, Y) = f[B(X|Y), B(Y)]$$

6. *Complementarity* : The belief in the negation of a proposition is a monotonically decreasing function of the belief in the proposition itself.
7. *Consistency* : There will be equal belief in propositions that are logically equivalent.

Cox and other researchers have demonstrated that, taken together, these properties logically imply that the measure of belief must satisfy the axioms of probability theory [25]. The proof of the necessary relationship between the intuitive properties and the axioms of probability theory is based on an analysis of solutions to the functional forms implied by the intuitive properties. Thus, according to Cox's proof, if one accepts these intuitive properties as desirable, one must accept probabilities as a desirable measure of belief.

These principles provide a useful framework for comparing alternative formalisms for representing uncertainty, in terms of which of the principles the formalisms reject [75]. For example, fuzzy-set theory [160] rejects the property of *clarity*, allowing linguistic imprecision in the definition of propositions. Some AI researchers have also rejected *scalar continuity*, arguing that a single number is insufficiently rich to represent belief [19]. Dempster-Shafer theory [133] rejects *completeness*, denying that it is possible to assign a belief to every well-defined proposition. Most heuristic quantitative approaches to representing uncertainty used in AI, even when they use the term *probability* as in Prospector [36], implicitly violate *consistency.*

2.3 The Direction of Probabilistic Inference Can Be Reversed

Probability theory, and in particular Bayes' theorem, allows us to reverse the direction of inference. That is, given the influence of hypothesis H on observable evidence E, expressed as $p(E|H,\xi)$, Bayes' theorem allows us to compute the influence of E on H, expressed as $p(H|E,\xi)$. Commonly, the hypothesis H is perceived as *causing* the evidence E. If H is a disease and E is an observable symptom, we can express the evidential relationship in the *causal* direction (i.e., $p(E|H,\xi)$), and then use Bayes' theorem to *reverse* the inference and reason in the *diagnostic* direction (i.e., $p(H|E,\xi)$) [132]. This bidirectionality is a consequence of Bayes' theorem.

Bayes' theorem follows from the last axiom of probability, relating the probability of a joint event (i.e., a conjunction) to conditional probabilities [4]. The theorem written in its standard form for relating a hypothesis H to evidence E is

$$p(H|E,\xi) = \frac{p(E|H,\xi)p(H|\xi)}{p(E|\xi)}$$

We can expand this equation, in terms of the negation of H, $\neg H$, to

$$= \frac{p(E|H,\xi)p(H|\xi)}{p(E|H,\xi)p(H|\xi) + p(E|\neg H,\xi)p(\neg H\xi)}$$

The theorem simply states that the belief in the hypothesis in light of the evidence, $p(H|E,\xi)$ (the *posterior* probability), depends on how likely it is that we observe a particular piece of evidence given the hypothesis and its negation, $p(E|H,\xi)$ and $p(E|\neg H,\xi)$, and the *prior* probability of the hypothesis, $p(H|\xi)$.

The inferential symmetry of probabilistic reasoning can be useful when probabilities are available in one direction but are required in the reverse direction. For example, domain experts may be more comfortable with specifying probabilities in the *causal* direction, through assessing $p(E|H,\xi)$, but may wish to calculate beliefs in the *diagnostic* direction, reasoning about the belief in hypotheses given evidence, $p(H|E,\xi)$. Representing belief in the causal direction frequently is a more parsimonious and invariant representation of the uncertain relationships than is the diagnostic relationship, which will vary with prior probabilities (e.g., for different populations). Moreover, in causal form, complex relationships among multiple hypotheses and multiple effects can be frequently decomposed into simpler relationships from each hypothesis to its individual effects, which can be assessed separately.

2.4 Decision Theory Provides a Framework for Reasoning About Preferences

Decision theory is based on the axioms of probability and utility. Where probability theory provides a framework for coherent assignment of beliefs with incomplete information, utility theory introduces a set of principles for consistency among preferences and decisions. A *decision* is an irrevocable allocation of resources under control of the decision maker. *Preferences* describe a decision maker's relative valuations for possible states of the world, or outcomes. The valuation of an outcome may be based on the traditional attributes of money and time, as well as on other dimensions of value including pleasure, pain, life-years, and computational effort.

Utility theory is based on a set of simple axioms or rules concerning choices under uncertainty. Like the axioms of probability theory, these rules are fairly intuitive. The reader is referred elsewhere for a detailed presentation of different versions of the axioms, their rationale, and implications [151, 129, 32, 41]. Here, we only try to give the axioms' flavor.

The first set of axioms concerns preferences for outcomes under certainty. The axiom of *orderability* asserts that all outcomes are comparable, even if described by many attributes. Thus, for any two possible outcomes x and y, either one prefers x to y or one prefers y to x, or one is indifferent between them. The axiom of *transitivity* asserts that these orderings are consistent; that is, if one prefers x to y and y to z, then one prefers x to z. These axioms, together with two auxiliary axioms, ensure a weak preference ordering of all outcomes. This result implies the existence of a scalar value function *V(x)*, which maps from all outcomes x into a scalar "value" such that one will always prefer outcomes with a higher "value."

The second set of axioms describes preferences under uncertainty. They involve the notion of a *lottery*, an uncertain situation with more than one possible outcome. Each outcome has an assignable probability of occurrence. The *monotonicity* axiom says that, when comparing two lotteries, each with the same two alternative outcomes but different probabilities, a decision maker should prefer the lottery that has the higher probability of the preferred outcome. The *decomposability* axiom says that a decision maker should be indifferent between lotteries that have the same set of eventual outcomes and probabilities, even if they are reached by different means. For example, a lottery whose outcomes are other lotteries can be decomposed into an equivalent one-stage lottery using the standard rules of probability. The *substitutability* axiom asserts that, if a decision maker is indifferent between a lottery and some certain outcome (the *certainty equivalent* of the lottery), then substituting one for the other as a possible outcome in some more complex lottery should not affect her preference for that lottery. Finally, the *continuity* axiom says that, if one prefers outcome x to y, and y to z, then there is some probability p such that one is indifferent between getting the intermediate outcome y for sure and a lottery with a p chance of x (the best outcome) and *(1 - p)* chance of z (the worst outcome).

It follows from accepting the axioms of utility that there exists a scalar *utility function U(x,d)*, which assigns a number on a cardinal scale to each outcome x and decision d, indicating its relative desirability. Further, it follows that when there is uncertainty about x, preferred decisions d are those that maximize the expected utility $E[U(x,d)|\xi]$ over the probability distribution for x.

The consistency criteria embodied in classical decision theory can be stated as follows: Given a set of preferences expressed as a utility function, beliefs expressed as probability distributions, and a set of decision alternatives, a decision maker should choose that course of action that maximizes expected utility. The power of this result is that it allows preferences for complex and uncertain combinations of outcomes with multiple attributes to be computed from preferences expressed for simple components. Thus, it may be used as a tool to help people think about complex choices by decomposing them into simpler choices.

A utility function for uncertain situations also allows one to express attitudes toward risk, such as *risk aversion*, when contemplating lotteries involving quantitative attributes such as money. Risk aversion is exhibited by many people, when they prefer to receive a monetary prize for certain over a lottery with an identical expected value. Decision theory provides various techniques for eliciting and encoding different attitudes toward risk for supporting decision making under uncertainty [80].

Although the valuation of alternative states and choices about the allocation of resources often is central for computational agents, the crucial notions of decision and preference have not been addressed in a well-defined manner in AI.

2.5 Decision Theory Is Normative

If one finds the axioms of decision theory compelling as principles for rational choice, then the theory is *normative*. In other words, the axioms provide a set of criteria for consistency among beliefs, preferences and choices that "should" be adhered to by a rational decision maker. Alternatively, given a set of beliefs and preferences, the theory prescribes as rational only those decisions that maximize expected utility. A system that makes decisions or recommendations consistent with the axioms may also be termed *normative*.[2]

It is important to understand that decision theory is not generally proposed as a *descriptive* theory; it does not purport to provide a description of how people actually behave when reasoning under uncertainty. Indeed, studies have demonstrated that people frequently do not behave in accordance with decision theory [85]. In fact, characteristic (and often costly) biases exhibited in intuitive judgment are part of the justification for applying decision sciences to assist people with decision making.

2.6 Good Decisions Must Be Distinguished from Good Outcomes

A decision-theoretic perspective distinguishes between a good decision (a choice made consistent with preferences and beliefs) and a good outcome (the result of a choice that turns out to be desirable). We can labor mightily to elicit probabilities, to structure values, and to assess alternatives and still make a choice that leads to a bad outcome. Alternatively, a random or poor selection may well turn out fortuitously. Such is the nature of acting under incomplete information. Decision theory strives for good decisions that lead to better outcomes on average.

2.7 Incompleteness and Uncertainty Are Unavoidable

Systems that reason about real-world problems can represent only a portion of reality. It is clear that any computational representation must be a dramatic simplification of the objects and relations in the universe that may have relevance to a decision problem. The inescapable incompleteness in representation leads to unavoidable uncertainties about the state of the world and about the consequences of our actions. In practice, uncertainty is particularly acute when dealing with multiple actors, complex preferences, high stakes, and long-term consequences.

2.8 Decision Analysis Is Applied Decision Theory

Decision analysis is an engineering discipline that addresses the pragmatics of applying decision theory to real-world problems. Decision theory only dictates a set of formal consistency constraints; it says nothing about how we elicit or represent a utility function or probability distribution, or about the manner in which we represent or reason about a decision problem (e.g., the level of granularity and detail). It also does not address search procedures for a utility-maximizing

[2]Logic-based reasoning methods also may be considered normative in that they prescribe a set of rules for correct inference under certainty; that is, a system that reasons or makes recommendations using these rules may be viewed as normative with respect to deterministic knowledge.

decision. Decision analysis, in contrast, addresses these issues directly in terms of decision making and tractability.

The discipline of decision analysis emerged in the 1960s; it grew out of a recognition that probability and decision theory, hitherto applied primarily to problems of statistical estimation [129, 122], also could be applied to real-world decision problems [78, 120]. Since its inception, decision analysis has grown into an established academic and professional discipline [83, 152, 88]. There are a number of commercial consulting and research firms that perform decision analyses for government and private clients. Some large corporations routinely apply decision analysis to scheduling, capital expansion, and research and development decisions. The emphasis has been on assisting people and organizations faced with high stakes and complex resource-allocation problems.

Decision analysis can be thought of as a set of techniques for focusing attention. It provides methods that help a decision maker to clarify the problem by explicating decision alternatives, values, and information. It provides a variety of techniques for *sensitivity analysis*, to help a person identify those uncertainties and assumptions that could have a significant effect on the decision recommendations. The decision maker can then focus attention on those factors that make a difference in decisions, and can ignore aspects of the problem that turn out to have relatively minor relevance. Resources therefore can be directed to the most important or *sensitive* aspects of the problem.

3 Early Research on Expert Systems

The area of AI in which decision theory has had the most obvious influence is that of diagnostic expert systems. This emphasis has occurred, in large part, because expert systems are often concerned with inference and decision making under uncertainty. In this section, we review the early application of probabilistic methods in expert systems. We then discuss the more popular heuristic approaches that were developed later, partly as a reaction to the perceived limitations of the early schemes.

By *expert system*, we mean a reasoning system that performs at a level comparable to or better than a human expert does within a specified domain. We have found it useful to divide tasks for which expert systems have been constructed into *analytic* and *synthetic* categories. In systems dedicated to analytic tasks, a set of alternatives such as possible diagnoses or decisions either are explicitly enumerated or are relatively easy to enumerate; the central task is the valuation of the alternatives. With synthetic tasks, the space of alternatives (e.g., the set of possible configurations or plans) may be extremely large, and the main problem is constructing one or more feasible options. Analytic tasks include prediction, classification, diagnosis, and decision making about a limited set of options. Synthetic tasks include the generation of alternatives, design, configuration, and planning. Many of the best-known expert systems address analytic tasks, such as medical diagnosis. However, some of the most successful systems are applied to synthetic problems, such as R1 for computer-hardware configuration [103].

Decision theory provides an appealing approach to analytic tasks, particularly to those involving inference and decision making under uncertainty. Consequently, we focus on expert systems for analytic tasks. Decision theory also can be relevant to synthetic tasks, because useful alternatives often must be selected from large numbers of options.

Much of the pioneering work in analytic expert systems has been done on medical applications, although, more recently, fault diagnosis in electronic components and mechanical devices has been examined [33, 45]. In general, three kinds of task are involved. The first task is *diagnosis* : How can we infer the most probable causes of observed problems (e.g., diseases or machine-component failures) given a set of evidence (e.g., symptoms, patient characteristics, operating conditions, or test results)? The second task is making *information-acquisition decisions* : Which additional information should we request or which additional tests should we conduct? This choice involves weighing the costs of obtaining the information verses its benefits in more accurate diagnosis. The third task is making *isions*: What can we do to ameliorate or cure the problem?

3.1 The Problem of Diagnosis

First we formulate the problem of diagnostic inference. Suppose we are considering a set $\mathbf{H}$ of n possible hypotheses,

$$\mathbf{H} = \{H_1, H_2, \ldots, H_n\}$$

and a set $\mathbf{E}$ of m pieces of evidence,

$$\mathbf{E} = \{E_1, E_2, \ldots, E_m\}$$

Assume that all hypotheses and pieces of evidence are two-valued, logical variables, each either true or false. In a deterministic world, we could assume a relation $C(\mathbf{H}, \mathbf{E})$ between hypotheses and evidence, such that $c(H_i, E_j)$ means that hypothesis H_i implies (or causes) evidence E_j. A diagnosis or explanation is a set of hypotheses believed to be present (with all others absent). Given a set of evidence $\mathbf{E}'$, the deterministic diagnostic problem is to discover one or more diagnoses $\mathbf{D} \subseteq \mathbf{H}$, that can explain the observed evidence. In particular, $\mathbf{D}$ should contain, for all E_j in $\mathbf{E}'$, a hypothesis H_i such that $c(H_i, E_j)$. Reggia [123] proposed this formulation of the problem of diagnosis and developed set-covering algorithms for finding the minimum set of causes that could explain a set of observations.

In the real world, the relationships among hypotheses and evidence generally are uncertain. The probabilistic approach is to represent these relationships as the conditional probability distribution $p(\mathbf{E}'|\mathbf{D}, \xi)$ for the evidence, given each possible diagnosis $\mathbf{D}$ in $\mathbf{H}$. If, in addition, we have the prior probability $p(\mathbf{D}|\xi)$ for each subset $\mathbf{D}$ in $\mathbf{H}$ representing the believed prevalence rates of combinations of hypotheses, we can apply Bayes' theorem to compute the posterior probability of each diagnosis, after observing evidence, $\mathbf{E}'$:

$$p(\mathbf{D}|\mathbf{E}', \xi) = \frac{p(\mathbf{E}'|\mathbf{D}, \xi)p(\mathbf{D}|\xi)}{p(\mathbf{E}'|\xi)}$$

The problem of diagnosis is computationally complex. Because a patient may have more than one disease out of n possible diseases, the number of possible diagnoses (i.e., disease combinations) is 2^n. So the number of independent parameters necessary to specify the complete prior distribution is $2^n - 1$. For m pieces of evidence, the general conditional distribution has $2^m - 1$ independent parameters given each hypothesis, requiring the specification of $2^n(2^m - 1)$ independent parameters in total for all diagnoses. Clearly, this approach is quite impractical for more than two or three hypotheses and pieces of evidence without some kind of simplification.

3.2 Early Probabilistic Approaches

A set of research projects on automated probabilistic reasoning for diagnosis was undertaken during the 1960s. Two simplifying assumptions often were made. First (A1), that the hypotheses in $\mathbf{H}$ are mutually exclusive and collectively exhaustive. Second (A2), that there is *conditional independence* of evidence given any hypothesis. That is, given any hypothesis H, the occurrence of any piece of evidence E_i of the component hypotheses is independent of the occurrence of any other piece of evidence E_j:

$$p(E_i|H,\xi) = p(E_i|E_j, H, \xi).$$

With assumption $A1$, the only diagnoses we need to consider are the n singleton hypotheses H_i. With assumption $A2$, the conditional probability distribution of the evidence $\mathbf{E}'$ given a disease H_i, (as required for Bayes' theorem) can be decomposed into the product of the conditionals for individual pieces of evidence as follows:

$$p(\mathbf{E}'|H_i,\xi) = p(E_1, E_2, \ldots, E_j|H_i,\xi) = p(E_1|H_i,\xi)p(E_2|H_i,\xi)\ldots p(E_j|H_i,\xi)$$

Under the assumptions $A1$ and $A2$, only mn conditional probabilities and $n - 1$ prior probabilities are required. The simplicity of probabilistic systems based on these two assumptions made the approach popular. Several medical diagnostic systems have been constructed based on the simplified probabilistic scheme [145], including systems for the diagnosis of heart disease [154, 52], and of acute abdominal pain [31]. The popularity of the simplified probabilistic inference has led some people to believe that the assumptions are absolute requirements of probabilistic inference. It is a misconception, however, to regard this simplified Bayesian scheme as defining practical probabilistic inference. In the section on current research we describe the development of more expressive representations of probabilistic dependencies.

3.3 Performance of the Early Probabilistic Systems

How well did these early systems perform in terms of diagnostic accuracy? We note that the early probabilistic systems performed within their limited domains at a level comparable to experts, and sometimes at a considerably higher level [51, 30, 29]. For example, the system of de Dombal and his colleagues averaged over 90% correct diagnoses of acute abdominal pain, where expert physicians were averaging 65% – 80% correct [30]. Patrick's diagnostic aid for chest pain reportedly averaged 80% accuracy, whereas clinicians averaged 51% [113]. These systems certainly qualify as expert systems according to our definition.

It is interesting to ask why these systems performed better than experts given that they made simplifying assumptions ($A1$ and $A2$) and frequently considered only a fraction of the information available to physicians. One answer is that some of the computer programs were based on statistical analysis of empirical data rather than purely on expert judgment. However, the use of more reliable data does not explain the performance of several of the systems in which probabilities were based partly or entirely on expert judgment.

In fact, such good performance of simple models based on subjective parameters relative to unaided expert judgment is consistent with well-established experimental results from numerous studies [29]. Studies in a wide variety of domains of clinical and other expert judgment have found that simple linear models with subjectively assessed weights do as well as or better than experts. One reason for these results seems to be that the simple formal models are more consistent and reliable than human experts, being less subject to whims, carelessness, or misguided inspiration. There also are fundamental mathematical reasons why simple linear models can be robust approximations to more complex, nonlinear relationships [29].

The relevance of these surprising results to research in expert systems and artificial intelligence has only recently been pointed out [54, 16]. Several preconditions for the applicability of the findings have been elucidated. For the results to apply, tasks must fulfill at least two conditions: (1) the behavioral criterion must be some monotonic function of each input, and (2) there must be some noise in the inputs or the model, so that even optimal performance is limited. These conditions appear to apply to many diagnostic tasks in complex areas like medicine. Nevertheless, it is unclear just how well simple linear models can do relative to human experts and expert systems for different kinds of diagnostic tasks. Further theoretical and empirical research is needed on the usefulness of simple models. Of particular interest is identifying and characterizing task attributes that would be useful in predicting the relative performance of different approaches.

3.4 Problems Attributed to the Early Expert Systems

Enthusiasm for probabilistic and decision-theoretic methods faded in the early 1970s. Given the encouraging performance of these systems, why have they not seen wider application? The answer seems to involve a complex tangle of factors, including both technical and sociological ones. One often-cited reason is the restricted problem domains to which the probabilistic approach has been applied. A second reason is the unwarranted simplifying assumptions of mutual exclusivity and conditional independence–and the immediate intractability associated with attempts to move beyond the assumptions. More generally, critics of the approach have pointed out the limited expressiveness of the simplified Bayesian formulation, citing the apparent mismatch between the rigorous, formal, quantitative approach of probabilistic inference and the informal, qualitative approach characteristic of human reasoning. They suggest that the mismatch leads to problems both in encoding expertise and in explaining the results of probabilistic inference, so that users could understand and trust them [144, 27, 51].

One interesting lesson from the early research on probabilistic reasoning is the distinction between the performance and acceptability of diagnostic systems. In principle, it might seem that none of the objections we have listed should be insuperable in the face of superior diagnostic performance. However, it is clear that factors other than performance perform a key role in determining acceptance. Such factors may include the poor user interface of many early systems [140] and the general lack of attention paid to how the use of such systems might be integrated with the habits and environment of the diagnostic practitioner.

3.5 AI Approaches to Expert Systems

Concern about the restrictive assumptions of the simplified probabilistic scheme coupled with the perception that a combinatoric explosion would threaten any attempt to move beyond these assumptions or to larger domains led to disenchantment with the approach. At the same time, the new AI techniques being developed in the early 1970s appeared to provide a promising alternative to the design of expert systems. With the development of heuristic inference methods came reduced concern with normative optimality and methods for decision and inference under uncertainty. The attention of mainstream AI researchers became focused on the crucial role of representing and reasoning with large amounts of expert knowledge. Of particular interest was the potential of applying these AI reasoning techniques for building larger systems that could make use of richer and more varied expert knowledge than seemed possible for Bayesian schemes. Many of the researchers who became involved in the development of this new generation of systems came from other backgrounds and had little exposure to or interest in the earlier decision-theoretic schemes, which fell into relative neglect.

A key feature of the new expert-system paradigm was the application of the production-rule architecture to real-world diagnosis. Production rules had appeal as providing a general and flexible scheme for representing expert knowledge in a declarative and modular form [12]. The production rule has the form of logical implication. To apply production rules in real-world diagnosis, investigators found it desirable to extend the representation to represent uncertainty, both about the truth of propositions and about the applicability of each production rule. The two best-known attempts to develop representations of uncertainty as an extension of deterministic rule-based expert systems were the Mycin [12] and Prospector [36] projects.

Mycin was designed to assist physicians in the diagnosis and treatment of bacterial infections. Mycin introduced the *certainty-factor* model. The certainty factor (CF) is a number representing the degree of confirmation (between 0 and 1) or disconfirmation (between 0 and -1) of each proposition or rule. Mycin's basic knowledge representation and uncertain inference scheme have been made available for other applications as Emycin and are employed in several commercially available expert-system shells. Prospector was constructed as an aid in the identification of mineral deposits of commercial interest. Prospector uses probabilistic quantities to represent degrees of belief in propositions, although its updating rules are not exactly consistent with a coherent probabilistic interpretation. The developers of both systems have implied in their writings that they intended the systems' behaviors as approximations to the probabilistic ideal, which they saw as unattainable for the reasons we have discussed.

3.6 Problems with the Representation of Prior Belief

A common objection to probabilistic approaches is the difficulty of assessing prior probabilities–the initial measures of belief assigned to hypotheses before new evidence is considered. Empirical data often are hard to obtain and subjective estimates are deemed to be unreliable. Many heuristic schemes–including Prospector, Casnet and PIP–also require prior beliefs, and so do not evade this problem either. But some, including the Mycin certainty factor model and Internist-1 (and its descendant QMR), appear to reason without prior belief.

The Mycin certainty factor (CF) model represents, combines, and propagates the effects of multiple sources of evidence in terms of their joint degree of confirmation or disconfirmation of each hypothesis of interest. Thus, contrary to most popular interpretations, the CF originally was intended to represent an *update* or change in belief induced by the evidence, not an *absolute* degree of belief (such as a probability) [74]. It therefore does not explicitly represent the prior or posterior degree of belief in each hypothesis. By representing only updates rather than absolute degrees of belief, it appears to avoid the need for priors.

When a CF-based system recommends a decision (for example, when Mycin suggests treatment for a suspected infection), it makes use of the CFs assigned to the competing diseases to assess the amount of evidence for each. Because it makes decisions without any explicit reference to priors or prevalence rates, it is, in effect, treating all infections as having equal prior probabilities.[3] The Internist-1 and QMR systems make similar assumptions [58]. The equal-priors assumption is valid in contexts where diagnoses are believed to be equally likely and in contexts where no information is available about the prior probabilities.

Prior beliefs, at some level of precision, frequently are available. For example, experienced physicians have significant knowledge about the prevalence rates of different diseases, even though they may find these rates difficult to quantify precisely. In fact, diseases often differ in prevalence rates by many orders of magnitude. Assuming equal priors could lead to a serious error in a treatment recommendation in a case where two diseases with widely differing prevalence rates were assigned comparable CFs. For example, the fairly prevalent mononucleosis and relatively rare Hodgkin's disease can present with a similar set of evidence (microscopic features within a lymph node biopsy); the differences in the prior probabilities can be essential in diagnosis and treatment.

The errors that accrue from assuming equal prior probabilities may be less serious in domains where the quantity and quality of evidence typically overwhelms the priors. A knowledge engineer might be warranted in making simplifying assumptions about priors, given a demonstrated insensitivity of system performance to the assumption coupled with an analysis of the costs of representing prior information. However, in general, even approximate information about prior probabilities may be valuable knowledge that is important to represent explicitly in a knowledge-based system, and discarding this information can lead to significant errors.

Other heuristic systems that explicitly incorporate prior probabilities have difficulties due to *incoherence* among the probabilities. For example, Prospector uses probabilities to represent prior

[3] The handling of priors in the CF model is consistent with studies of how people reason under uncertainty [85] that show people tend to ignore priors.

degrees of belief in its hypotheses and in its evidential variables. The system makes use of two probabilistic quantities that *overspecify* the joint probability distribution.[4]

Since these measures are assessed independently by the domain expert, and their relationship is not intuitively obvious, they generally will be inconsistent with one another. To cope with this problem, Prospector employs a heuristic scheme for diagnostic inference that employs an interpolation between the conflicting quantities. However, the underlying incoherence of the probability distributions limits the scope of the inference. For example, the incoherence destroys the bidirectionality of probabilistic inference, obstructing the graceful integration of causal and diagnostic inference.

3.7 Problems with Modularity

An often-cited advantage of the rule-based representation scheme is the ability to add or remove rules from a knowledge base without modifying other rules [28]. This property has been referred to as *modularity*. The modularity of rules in a logical production system is a consequence of the monotonicity of logic: Once asserted, the truth of a proposition cannot be changed by other facts. This notion of rules as a modular representation of knowledge in deterministic production systems was carried over to rule-based methods for uncertain reasoning. However, analysis of modularity has demonstrated that uncertain beliefs are intrinsically less modular than beliefs held with certainty, frequently making the rule-based calculi inefficient for reasoning with uncertainty [59]. It has become apparent that the traditional assumption of modularity in rule-based approaches for reasoning under uncertainty has restrictive implications that had not been previously appreciated.

To explain this, we must define modularity more precisely in terms of procedures for updating belief in hypotheses [74, 57]. First we define the notion of a belief update. Suppose $B(H, \xi)$ denotes a scalar degree of belief in hypothesis H given some specific background evidence ξ. Let the scalar function $U(H, E_1, \xi)$ denote a *belief update* or a change in some measure of belief for a single hypothesis H, given some new evidence E_1, in the presence of previously acquired background evidence. These arguments should be sufficient to determine how we combine new evidence with previous evidence to establish the posterior degree of belief in H given E_1 in the context of the background state of information. Thus, there should be some scalar function f such that

$$B(H, E_1\ and\ \xi) = f[U(H, E_1, \xi), B(H, \xi)]$$

In general, the update $U(H, E_1, \xi)$ may depend on other evidence in the background information ξ. Thus, a belief update would have to be specified for every H and relevant E_i for every possible ξ. Clearly, this poses an intractable representation and inference problem. The simplifying assumption of modularity in rule-based systems is that the update associated with a piece of evidence E_1 is

[4]The strength of each rule linking evidence E with hypothesis H is represented by two numbers, representing the two likelihood ratios:

$$LS = p(E|H, \xi)/p(E|\neg H, \xi)\ and\ LN = p(\neg E|H, \xi)/p(\neg E|\neg H, \xi)$$

independent of all other evidence. We define the *modular update* property as

$$U(H, E_1, \xi) = U(H, E_1, E_i \, and \, \xi) \forall E_i, i \neq 1$$

Now, suppose there are two pieces of evidence, E_1 and E_2, that bear directly on the hypothesis H. We say that the *combination* of updates is modular if a single belief update encompassing the effects of E_1 and E_2 on H is a simple function of the two separate updates. Thus, we define the *modular combination* property as

$$U(H, E_1 and \, E_2, \xi) = g[U(H, E_1, \xi), U(H, E_2, \xi)]$$

where g is a scalar function that is continuous and monotonically increasing in each argument, given that the other argument is held constant. The modular combination property follows directly from the modular update property for each of the two component updates.

Now let us apply these ideas in the probabilistic framework. A well-known form of probabilistic update is the *likelihood ratio*. Heckerman has shown that *any* probabilistic belief update must be some monotonic transformation of the likelihood ratio [55].

If we divide Bayes' theorem for hypothesis H, evidence E, and background evidence ξ by Bayes' theorem for the negation of the hypothesis, $\neg H$, we get

$$\frac{p(H|E,\xi)}{p(\neg H|E,\xi)} = \frac{p(E|H,\xi)p(H|\xi)}{p(E|\neg H,\xi)p(\neg H|\xi)}$$

This is called the *odds-likelihood* form of Bayes' theorem. From left to right, the first and last ratios are respectively the posterior and prior ratios of probability, or the *odds*. We write these ratios as $O(H|E,\xi)$ and $O(H|\xi)$. The second ratio is the likelihood ratio, for which we use the term λ (H, E, ξ). We can rewrite the odds-likelihood equation as

$$O(H|E,\xi) = \lambda(H, E, \xi)O(H|\xi)$$

This shows that the posterior odds of H given E is the product of the likelihood ratio and the prior odds of H. In this case, the update function f is simply the product of its arguments.

Suppose we want to combine the updates on H from two pieces of evidence,E_1 and E_2. For the combination function to have the modular combination property, we require that E_1 and E_2 be conditionally independent given H and $\neg H$. From the definition of λ, it then follows that

$$\lambda(H, E_1 \, and \, E_2, \xi) = \lambda(H, E_2, \xi)\lambda(H, E_1, \xi)$$

That is, the modular combination function g for likelihoods is also the product of its arguments, the two component likelihoods. Thus, for updating schemes based on the multiplication of likelihood ratios (such as probabilistic interpretations of the popular rule-based schemes), modularity requires a stronger form of conditional independence than that required for the simplified probabilistic scheme described earlier.

All modular updating schemes assume that all pieces of evidence bearing on the belief in a hypothesis H can be combined to determine an overall effect on H, through examining the belief in each piece of evidence. If the beliefs in the pieces of evidence are each represented by a scalar, they cannot explicitly express the possible dependence between them. The representation simply is not rich enough. Capturing the effects of arbitrary dependencies in a modular scheme generally requires information that is unavailable to a local combination function. Attempting to generate behavior consistent with complex dependency within a modular updating scheme is an unreasonable pursuit of "something for nothing" behavior [74]. Thus, we cannot capture information about arbitrary dependencies with simple scalar functions.

Modular evidence combination and belief updating schemes must make some default assumptions about dependency among pieces of evidence updating the same hypothesis. Henrion has demonstrated that any rule-based scheme for uncertain inference with local updating must make some general assumption about dependence among these convergent lines of evidence [62]. The assumption of conditional independence by modular schemes based on the likelihood ratio defines only one set of such assumptions. It is possible to have modular combination functions that dictate more complex default assumptions about *patterns* of dependency. This is a current area of investigation [56].

In summary, like the early probabilistic systems, the popular rule-based method imposes strong restrictions on the kinds of dependence that can be represented effectively. Unlike the explicit assumptions of the simplified probabilistic systems, the restrictive assumptions in the heuristic approaches have been less apparent. One might argue that the implicit nature of the assumptions in rule-based systems has tended to promote a dangerous "myth of modularity" among uncertain propositions: Rule-based approaches, like the simple probabilistic approaches, do not have the *expressiveness* necessary to represent coherently the relationships among uncertain beliefs.

3.8 Robustness of the Heuristic Approaches

A common response to criticisms of the assumptions embodied in the heuristic approaches to reasoning under uncertainty is that the choice of uncertainty calculus and assumptions is not important in the performance of real systems. Indeed, Mycin and Internist-1 perform at the expert level despite the identified inconsistencies. A formal study by Cooper and Clancey demonstrated that Mycin's performance was fairly insensitive to the precision of the numbers used for certainty factors [13]. This view is buttressed by the findings we mentioned, that the early probabilistic expert systems performed well (often better than human experts) despite their simplifying assumptions.

It is dangerous, however, to generalize from these results. The Mycin domain is relatively forgiving; for example, the use of wide-spectrum antibiotics to "cover" several leading hypotheses means that misdiagnosis of the infecting organism does not necessarily lead to inadequate treatment. Careful examination of the results of a comparison of CFs with probabilistic inference presented in the original paper on CFs [135] shows that, on average, the CF-based system underresponded to the diagnosticity of the data by a factor of two [159, 62]. In 25% of the cases, it actually responded to the wrong direction, interpreting evidence that overall supported a conclusion to be disconfirming, or vice versa.

Investigators have found that inappropriate assumptions of conditional independence in simplified Bayesian systems can lead to noticeable degradation of performance. Norusis found significant improvement in the performance of a medical diagnostic system as the number of dependencies explicitly represented was increased [109]. Fryback discovered that problems with assuming conditional independence can grow as the number of variables represented in a diagnostic model are increased; that is, the potential benefits of considering a larger number of variables can be overwhelmed by the proportional increases in the missing dependencies [43].

Problems with the use of an updating scheme making the strong conditional independence assumptions of the CF model were noted in early research on the Pathfinder expert system [76] for diagnosing tissue pathology. Moving to a simplified probabilistic combination scheme, assuming conditional independence of evidence given hypotheses, yielded significant increases in diagnostic performance. The increased performance has been quantified with an evaluation scheme incorporating decision-theoretic and ad hoc measures [58].

Wise experimentally compared the performance of six common uncertain inference schemes for small rule sets and found that differences in performance between heuristic and probabilistic schemes depend heavily on the situation [159]. As we might expect, when there was strong evidence in one direction, most schemes performed well. But when the evidence was weak or conflicting, heuristic schemes tended to perform poorly, and in some cases did no better than random.

Thus, we should not conclude from the apparent robustness of the performance in a forgiving domain that handling of uncertainty makes little difference in other applications. For example, in models where expensive information-acquisition decisions typically are required to reduce uncertainty, the inappropriate handling of uncertainty may lead to costly decisions. The sensitivity of a system's performance to inconsistency, to assumptions of modularity, or to the use of inaccurate measures of belief will depend to a great extent on the task. In many cases, the inconsistencies and assumptions can lead to costly error [74, 62].

Buchanan and Shortliffe, the creators of the certainty-factor model, have warned against uncritical application of the certainty-factor calculus to other domains [12]. However, there has been relatively little discussion about the applicability of these warnings in the popularity of expert systems employing certainty factors (i.e., the widely used derivatives of the Emycin shell) and similar heuristic schemes. Further theoretical and experimental studies of the sensitivity of inference schemes to different assumptions and errors in situations of weak or conflicting evidence in real-world problem solving are required to understand the costs associated with the use of different heuristic methods.

3.9 Toward More Expressive Representations

In summary, early schemes using simplified probabilistic representations and inference often have been successful in terms of performance relative to that of human experts in small domain areas. The systems, however, have not been widely adopted for a variety of reasons, including their apparently unrealistic assumptions and their inability to represent the range of qualitative knowledge available to the expert. Originally, AI techniques were applied to the development of expert systems with the hope that they might avoid such arbitrary assumptions and incorporate a richer range of qualitative knowledge with smaller engineering costs. However, recent work has shown that many well-known AI approaches for representing and reasoning about uncertain knowledge also have made strong assumptions about prior probabilities and modularity.

4 Current Research on Decision Theory in Expert Systems

As we have seen, there has been justified criticism of the restrictive assumptions of both the simplified probabilistic schemes and several heuristic approaches to uncertain inference. In recent work, researchers have attempted to develop richer knowledge representations that are based in a principled way on probability and decision theory and are capable of expressing, in a flexible and tractable manner, a wider range of both qualitative and quantitative knowledge. Much of this work has centered on the use of graphs or networks to represent uncertain relationships, including *belief networks* and *influence diagrams*. These representations can facilitate assessment of coherent prior distributions, make assumptions explicit, and allow assumptions to be manipulated easily by knowledge engineers and experts.

In this section, we review the basic ideas on which these knowledge representation schemes are based and survey current methods for using them in decision-theoretic expert systems. We examine knowledge engineering, the process of encoding expert knowledge, using these schemes. We present various classes of inference techniques that use these representations for propagating evidence and finding optimal decisions. Finally, we review research on explaining the results of decision-theoretic inference.

4.1 Knowledge Representation for Decision-Theoretic Problems

Howard has called the complete model of a decision problem the *decision basis* [83]. A comprehensive decision basis consists of components that represent the *alternatives*, *states*, *preferences*, and *relationships* in a decision situation. *Decisions* are the alternative courses of action available to the decision maker. The alternative *states* of the world are those factors or variables that will be explicitly represented in the model, and the range of values that are considered reasonable or possible. The *preferences* of the decision maker are represented as a ranking in terms of the various possible outcomes. The preference information captures factors in a decision situation that are important in judging the desirability of alternative outcomes, as well as the manner in which tradeoffs among dimensions of outcomes are to be made. As we mentioned earlier, AI systems have not directly addressed explicit representation of knowledge about preferences. The final component of a decision basis is the set of *relationships* among states of the world, decisions and preferences. In general, these relationships can be expressed logically, probabilistically, or qualitatively.

A variety of representations for a decision basis have been developed in the decision sciences. These representations include joint probability distributions over variables coupled with a loss function (as used in probability and statistics), and decision trees, which evolved with the development of decision analysis [120]. Although these representations are useful and general, they do not

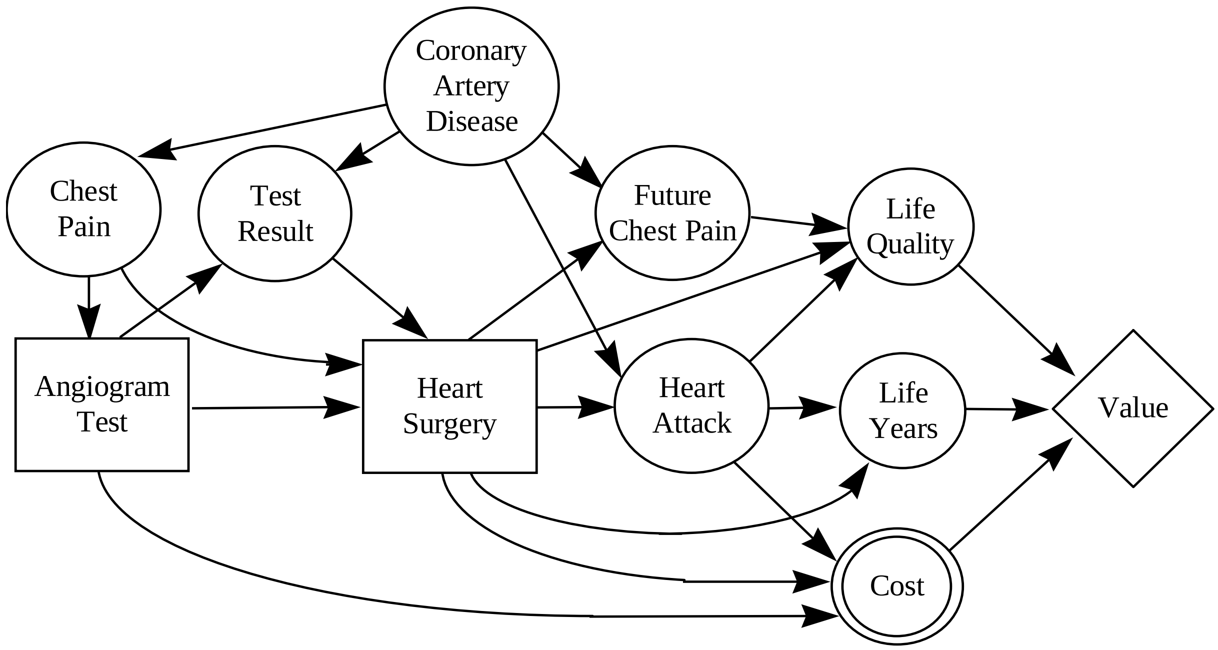

Figure 1: An Influence Diagram Describing the Decisions That a Patient with Heart Disease Might Face. Dependencies (arcs) link propositions (nodes) that represent states of information (circles), decisions (squares), and patient values (diamond).

provide a perspicuous means of representing *independence* in a manner accessible to both human and machine reasoners. Influence diagrams and belief networks were designed with precisely these objectives in mind.

Influence Diagrams and Belief Networks The influence diagram is a graphical knowledge-representation language that represents the decision basis [82, 111, 110]. The influence diagram is an acyclic directed graph containing nodes representing propositions or quantities of interest (i.e., alternatives, states) and arcs representing interactions between the nodes. Nodes representing propositions are associated with a set of mutually exclusive and exhaustive values that represent alternative possible states. The arcs represent deterministic, probabilistic, or informational relationships between nodes.

Influence diagrams formally describe a decision basis, yet have a human-oriented qualitative structure that facilitates knowledge acquisition and communication. An influence diagram for a medical decision problem is shown in Figure 1. The diagram encodes a decision problem about whether to undergo coronary artery bypass graft (CABG) surgery. The danger in this situation is the risk of myocardial infarction (MI) (i.e., heart attack).

The example demonstrates the four different kinds of nodes in an influence diagram. *Decision* nodes represent the possible actions available to a decision maker. They are the variables in an influence diagram under the direct control of a decision-making agent. These nodes are portrayed as rectangles in influence diagrams. Two decisions are depicted in the example: The Angiogram Test node refers to an artery-imaging procedure that provides information about the extent of Coronary-Artery Disease in the patient. Heart Surgery refers to a decision to undergo a CABG surgical procedure. The decisions are whether to undertake none, one, or both of the procedures.

The arcs into a decision node indicate what information is available (i.e., values of uncertain variables or previous decisions that have been resolved) at the time the choice is made. The diagram indicates that, when he makes the surgery decision, the decision maker will know the outcome of the angiogram test if it was performed.

Chance nodes represent states of the world that are uncertain. We depict chance nodes as circles or ovals. There are two kinds of chance nodes: *stochastic* and *deterministic* (the latter are portrayed as double-lined circles). The belief associated with a stochastic chance node is a probabilistic function of the outcomes of its predecessor nodes. For example, the probability distribution over the values of Life Years (i.e., years of life remaining) depends on whether heart surgery was performed (because there is a risk of death from the surgery itself) and the reduced risk of a future fatal heart attack if the operation is successful. The value of a *deterministic* node is a deterministic function of the outcomes of its predecessor nodes. In this example, we are assuming there is a deterministic function yielding costs based on the monetary expense of the angiogram test, the surgical procedure, and the hospitalization following a heart attack. A deterministic chance node is a special case of a stochastic chance node: The probability distribution is an impulse on a particular value, because the values of the predecessors determine the node's value with certainty.

Finally, *value* nodes capture the preferences of a decision maker. These nodes are depicted as diamonds. The predecessors to the value node indicate those outcomes or attributes that are included in the evaluation of a choice or plan. For the heart disease example, the attributes are life quality, life years, and cost. The graph shows that the quality of life is influenced by the chest pain at a particular level of exertion and the morbidity of surgery. The value function (a real-valued scalar function) encapsulates tradeoffs among these attributes for an individual patient, as well as individual preferences about risk and time.

Much of the research on representation and inference with these graphical representations has focused on specializations of influence diagrams that contain only chance nodes [126, 82, 97, 21, 116, 89]. These express probabilistic relationships among states of the world exclusively, without explicit consideration of decisions and values. Several different terms are used for these representations, including *causal networks*, *Bayesian nets*, and *belief networks* [114]. We use *belief networks*, as this term is the most popular.

Three Levels of Representation The expressiveness and sufficiency of influence diagrams is based in the representation's three levels of specification: *relation*, *function*, and *number* [82]. We can express relations at one level without explicitly referring to more specific levels.

The *relation* level captures the qualitative structure of the problem as expressed in the topology of the network. At this level, the arcs and nodes describe dependencies between the values of propositions or variables (nodes). Influence diagrams at the level of relation are similar to several common representations in modeling and in AI research, such as semantic nets. Each variable in an influence diagram is associated with a set of mutually exclusive and collectively exhaustive values (values for each node are not pictured in Figure 1.). For example, the node Chest Pain in our example is characterized by values of *none*, *mild discomfort*, and *crushing sensation* in response to a particular level of exertion. Coronary Artery Disease is characterized by *none*, *single-vessel*, *two-vessel*, and *three-vessel*, describing the number of arteries in the heart that are diseased. It is important that the outcomes of each node in the diagram be defined unambiguously. In the example, the arc between the Coronary Artery Disease and Chest Pain nodes expresses knowledge about the existence of a dependency between the values that coronary artery disease and chest pain may assume.

At the level of *function*, the functional form of the relationships among nodes is specified. For instance, the form of the conditional probability relating the outcome (value) of Coronary Artery Disease to the probability distribution over the values of Chest Pain is specified.

Finally, at the level of *number*, we specify numerical values that are operated on by the functional forms. This level represents the quantitative details of the dependence of each variable on its *parents* (the nodes that influence the variable).

An uncertain influence is represented by the conditional probability distribution for a variable given the values of its parents. As an example, at the level of number, we might specify that p(Chest Pain = *mild discomfort*|Coronary-Artery Disease = *one vessel*) = 0.25. Chance nodes without predecessors are specified at the level of number with unconditional or prior probability distributions.

Conditional Independence Independence usually is defined as a quantitative relation among probabilities–for example, as satisfaction of the product rule–expressed here in terms of marginal distributions.

$$p(a, b|\xi) = p(a|\xi)p(b|\xi)$$

However, we also may express independence with the following, more qualitative relationship, expressing that belief in proposition a is not affected by knowledge of the truth of proposition b, given background information ξ:

$$p(a|b, \xi) = p(a|\xi)$$

A belief network expresses independence graphically. The arcs in a belief network, and more precisely the lack of arcs among variables, are qualitative expressions of probabilistic independence of various kinds. In particular, source variables (i.e., those variables with no predecessors or directed pathway between them) are marginally independent. Variables u and w in Figure 2 are marginally independent of each other. Where two variables have one or more common parents but no arc between them, they are conditionally independent of each other given their common parent(s). In Figure 2, variables v and x are conditionally independent of each other given w. Finally, a node is conditionally independent of its indirect predecessors (i.e., nodes at a minimal directed path of distance greater than 1) given all of the variable's immediate predecessors (i.e., those nodes from

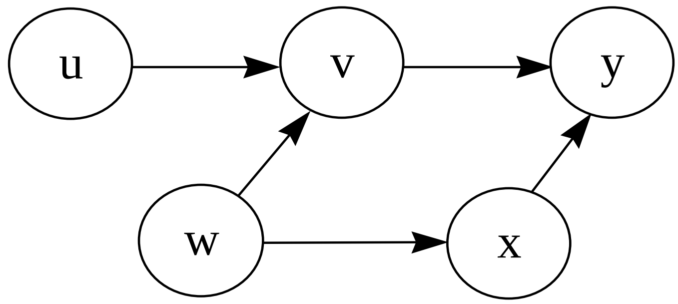

Figure 2: A simple belief network demonstrating conditional independence among propositions.

which it receives an arc directly). For example, in Figure 2, variable y is conditionally independent of u and w, given v and x.

At the numerical level, we assign marginal probability distributions to source variables (w and u in the example) and conditional probability distributions to all other variables given their immediate predecessors. We can compute the joint distribution over all the uncertain variables in a belief network simply as the product of all these marginal and conditional distributions:

$$p(u, v, w, x, y|\xi) = p(y|v, x, \xi)p(v|u, w, \xi)p(x|w, \xi)p(w|\xi)p(u|\xi)$$

Provided that the influence diagram has no directed cycles, the probability distributions assigned in this way are guaranteed to be *complete* (that is, to have no unspecified parameters) and *consistent* (that is, to contain no conflicts). In this way the belief network provides a simple solution to the problem that was unsolved in Prospector and related systems–namely, how to assign probabilities to variables and links without creating incoherence.

The influence diagram representation grants knowledge engineers the freedom to define and manipulate dependencies–or more important, *independencies*. The representation allows engineers to explicitly control modularity assumptions. The independencies in an influence diagram are a formal expression of the locality of effect among variables. The effects of one variable on a distant one can propagate only along the influence arcs. More precisely, a variable is screened from the effects of distant variables (is conditionally independent of them) given its *Markov blanket*; that is, given its direct predecessors, direct successors, and the direct predecessors of these successors (i.e., parents, children, and spouses). The presence of arcs explicitly defines *possible* dependence relations among the nodes in the graph.[5]

We can now interpret more carefully the heart surgery influence diagram (Figure 1) in terms of conditional independence. In the diagram, the primary expression of conditional independence involves Coronary Artery Disease and its effects. The diagram asserts that the probabilities over the values of Chest Pain (both current and future), the values of Angiogram test, and the values of

[5]An arc at the level of relation indicates only the possibility of dependence; at the detailed *number* level, it may turn out that the probability distribution over the values of a node is actually independent of a predecessor.

MI are all dependent on the value of Coronary-Artery Disease. Furthermore, given knowledge of Coronary-Artery Disease, these effects of the disease are conditionally independent of one another. Once we know the precise extent of a patient's heart disease, then presence of chest pain does not change our belief that he might have a heart attack at some time in the future. The knowledge of coronary artery disease as the causal agent tells us all the information available about the interaction of its effects.

For diagnostic reasoning and decision making, however, we might wish to reason backward; that is, we may wish to infer the probability of MI, given a specified degree of chest pain. As we mentioned earlier, the primary mechanism for this type of inference is Bayes' theorem. Later in this section we describe inference techniques that work with influence diagrams and belief networks to provide this type of reasoning.

4.2 Knowledge Engineering for Decision-Theoretic Systems

Knowledge engineering is the process by which expert knowledge is obtained, represented, refined, codified, and installed in computer-based diagnostic and decision-making systems. Although the term *knowledge engineering* has not been used traditionally in the field of decision analysis, the fundamental activities of a decision analyst and a knowledge engineer are similar. Both work with a decision maker or a domain expert to construct a formal representation of knowledge. The knowledge engineer typically uses rule-based or object-based representations, typically coupling them with some type of deductive inference method, whereas the decision analyst constructs influence diagrams or decision trees for use with decision-theoretic inference methods.[6]

The core of decision-analytic knowledge engineering is the construction of an informative, credible, and computable decision basis [68, 10]. As we have seen, influence diagrams reduce the complexity of assessing influences by allowing explicit graphical representation of dependencies and independencies. The diagram itself, therefore, is an important tool in knowledge engineering, as well as in computation.

In most current applications of influence diagrams and belief networks in expert systems, the model is constructed as part of the knowledge engineering process and is then is used during consultation [65, 5, 76]. We expect that, over time, additional components of knowledge engineering decision-theoretic systems will be automated.

In the remainder of this section, we briefly review some of the fundamental issues in engineering decision-theoretic systems. Our primary objective is to provide pointers to literature that addresses these issues in more detail.

Identifying Decisions and Generating Alternatives The set of decision alternatives has a tremendous effect on the overall value of an expert consultation. A new alternative often is worth more than extensive reasoning and analysis. The generation of new alternatives is a synthetic activity focusing on constructing actions or sequences of actions that achieve certain goals. Little research has been done on the knowledge-based generation of decision alternatives.

[6] See [65] for an experimental comparison of rule-based and decision-analytic paradigms for knowledge engineering.

There has been work on the problem of dealing with the explosion of decision sequences that occurs when a series of decisions is possible and each decision has several alternatives. The technique of *strategy tables*, from decision analysis, involves selecting several representative strategies from the full combinatoric set of possible sequences [105]. Each strategy consists of a sequence of decisions that are synergistic, or internally consistent. The strategies then are treated as the alternatives in the decision analysis. Langlotz and colleagues [91] propose a similar method and use heuristic search to generate a reduced set of possible medical therapy plans for detailed decision analysis.

Value Structuring and Preference Encoding AI investigators, to date, have placed little emphasis on the preferences or desires of decision makers or reasoning agents. Decision theorists have been studying preference in a subfield of decision analysis that emphasizes multiattribute decision problems [152, 88]. A decision-theoretic analysis is driven in large part by the attributes that are important to the decision maker (life duration, life quality, and monetary cost in the heart surgery example) and by the manner in which these attributes are combined in assigning value to alternative outcomes. Von Winterfeldt [152] and Keeney [88] present numerous theoretical results on multiattribute value issues and discuss elicitation procedures for assessing the complex preference structures in terms of individual attributes.

An additional important component of preference is the encoding of the decision maker's attitude toward risk. Utility theory deals directly with attitudes toward alternatives that have uncertain outcomes. A substantial literature addresses the theoretical and practical aspects of risk-preference encoding [152, 88, 80, 128, 87, 14]. Most techniques involve asking the decision maker about her or his preferences for various hypothetical gambles and then combining the results and checking for consistency. Several researchers have examined computer aids for value structuring and preference modeling [155, 67].

Encoding Probabilities One of the central tasks in engineering a decision-theoretic system is that of assessing probabilities. A frequent concern is the availability of probabilities as well as the numbers of probabilities that may be required. From a subjectivist viewpoint, it is possible in principle for anyone to assess a probability for any event, no matter how little he knows about it. After all, assessing a probability distribution is a way to express how little or much a person knows about something.

Nonetheless, expressing human knowledge in terms of probabilities is a demanding task. Researchers of human judgment under uncertainty have identified a set of biases and heuristics that tend to distort human decision making and judgments about uncertain events [85]. Such biases tend to narrow or skew assessed probability distributions, and interviewing methods emphasize making implicit assumptions explicit and encouraging the subject to consider a full range of information and possibilities. Decision analysts have drawn on this research, as well as on professional practice, to develop methods to mitigate the effects of these biases [139].

Once we have specified a general dependency structure for a set of probabilistic relationships, we can quantify the influences as conditional and marginal probability distributions. The conditional distribution $p(X_i|Y_1, Y_2, ..., Y_n, \xi)$ in general requires 2^n parameters for binary X and Y_i. Figure 3

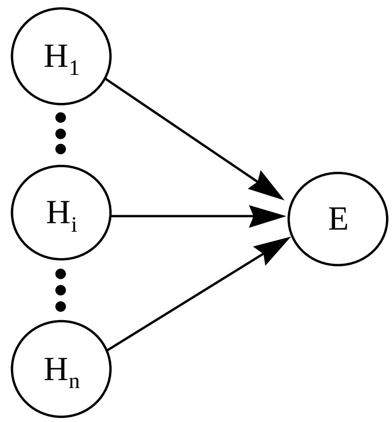

Figure 3: The Noisy-OR Prototypical Dependency Structure.

shows the conditional assessment task posed in terms of E, some observable output or evidence, and a set of inputs or hypotheses, $H_1, H_2, ..., H_n$.

We mentioned earlier that there has been recent work on the use of functions that specify patterns of independence. Recently, investigators have suggested methods for streamlining the probability assessment task, by specifying such *prototypical functions* for the probability distributions [56, 64, 114]. One example of a prototypical independence structure is termed the *noisy OR-gate*. We review this structure as an example of the assessment savings that may be gained through identifying and representing analogous patterns of independence.

The noisy-OR structure is a probabilistic generalization of a standard Boolean OR. With a Boolean OR, any single one of a number of input signals being true is sufficient to induce a true value for the output. In a noisy-OR, each input signal has some probability of being sufficient to cause the output; the processes that prevent the signal from being sufficient are independent. This structure has been found to be representative of many real-world probabilistic relationships as well as an efficient representation of probability information.

The noisy-OR relationship allows the full conditional distribution to be derived from the individual probabilities of the evidence given each of the hypotheses and so requires only n parameters [47, 114]. The basis for this savings is straightforward. Suppose that p_i is $p(E|\,only\ H_i, \xi)$–that is, the probability of E given that only H_i occurs. Then the probability of E given that all the H_i occur is

$$p(E|H_1, H_2, .., H_n, \xi) = 1 - \prod_{i=1}^{n} 1 - p_i$$

Researchers are seeking techniques for explicitly acquiring and representing several forms of independence. One means of identifying and assessing conditional probabilities in a perspicuous fashion is an attention-focusing representation called *similarity networks* [56]. A similarity network helps a knowledge engineer to identify sets of evidence that can disambiguate between pairs of hypotheses. The graphical display of these relationships indicates constraints on the conditional probability relationships between hypotheses and evidence.

Decision analysts have developed various techniques for eliciting numerical probabilities [139, 99]. Some of these assessment techniques ask directly for probabilities, whereas others seek specific values of a variable while holding the probability fixed. A popular method uses a probability wheel, a simple graphical tool consisting of a disk with two adjustable complementary sectors of different colors. The size of one colored sector can be adjusted to correspond to the probability to be assessed, as a way of expressing a probability without explicitly mentioning a number. For extremely low or high probabilities, techniques that use odds or log-odds have been shown to be useful [152].

This discussion of probabilities brings up the issue of the completeness of probability assessment. There is always a tradeoff between assigning a probability based on a current state of understanding and expending additional effort in modeling and introspection to come up with a better estimate [49, 101]. And, in practice, the assessor of a probability often is uncertain and uncomfortable about the distribution he is providing. Of course, the probability distribution that we would assess given additional time or effort is an uncertain quantity, and there is no fundamental barrier to using a probabilistic representation to represent this uncertainty (i.e., a second-order probability) [49, 60]. However, uncertainty about probabilities often masks the existence of other conditioning events for which the distribution is considered stable. It is the task of the knowledge engineer to draw out those conditioning events and thereby to expand the model (also referred to as *extending the conversation* [150]) to account for the uncertainty in assessment.

Model Refinement and Sensitivity Analysis The scope and detail of a decision model are central to determining the ultimate usefulness of an analysis. It may seem best to include all relevant factors in the analysis, in order to make the model as accurate and realistic as possible. In medicine, for example, variables related to family history, physical activity, diet, and life style may all be related to a decision regarding whether or not to proceed with a surgical procedure. However, as we noted, no model is complete, and inclusion of additional relevant variables can rapidly overwhelm our ability to solve, interpret, and explain the model.

Therefore, there is a tradeoff between the benefits of complete, detailed models and those of simplified, more computable models. In decision-theoretic knowledge engineering, we refer to the process of alternately expanding or contracting the model as *completeness modulation* [72, 67]. An attempt is made to produce a model that includes those variables that are most important to a particular set of decisions, in terms of having major influence on the recommendations developed by the model.

Sensitivity analysis is used to determine which parameters, uncertainties, and assumptions have the most influence on the behavior and recommendations of a model. It involves exploring the space of possible models in order to build a model that is simultaneously informative and tractable. Typically, this information is used to limit effort on construction and refinement of the model. There are several classes of sensitivity analysis that are appropriate at various junctures in the process of building a decision model [79, 81]. Sensitivity to risk tolerance, discretization, and uncertainty are routinely performed as part of a professional decision analysis. Since the probabilistic representation of a variable exacts costs in elicitation, representation, and inference, it is desirable to include only those uncertainties that matter. Henrion introduces the ı(expected value of including uncertainty) (EVIU) as a sensitivity measure of the importance of uncertainty [61].

To date, few people have investigated the automation of sensitivity analysis for probabilistic reasoning. A promising area in this regard is *error analysis*, the notion of determining the extent to which errors in inputs and assumptions, such as assessed probabilities, affect ultimate conclusions in decision-theoretic models [62]. Note that sensitivity is defined with respect to values and decisions and provides insight about the important components of a model. As AI applications seek to deal with ever larger knowledge bases, it will become increasingly important to use the individual decision-making context to develop tractable methods for reasoning.

4.3 Inference with Belief Networks and Influence Diagrams

Once we have constructed a belief network or influence diagram at the level of relation, and have assessed the details of the probabilistic dependencies at the level of number, we can perform inference about how changes in the belief of one of more propositions will change the belief in other propositions. There are several categories of inference. We may wish to compute the marginal probability distribution for a variable. For example, we may want to determine the probability of MI (heart attack) for a specific patient (see Figure 1). We may wish to compute the updated probability distributions over a variable–for example, the probability of different values of Coronary Artery Disease given the truth of a value of a related variable (e.g., Chest Pain = *none* ; Angiogram = *negative*). Finally, we may wish to select the best decision. In the sample influence diagram, we wish to decide whether to perform an angiogram and whether to operate, given the available information.

Unlike inference with a rule-based *inference net*, belief networks and influence diagrams allow inference in a direction opposite to the direction in which the influence was assessed. Consequently, it is possible to propagate the effect of observing any set of variables on the probability distribution of any other variable or function of variables.

In the next three sections, we discuss algorithms for probabilistic inference in belief networks. We consider exact and approximate methods. Finally, we consider decision making with the influence-diagram representation.

Brute-Force Probabilistic Inference As we have shown, a belief network with probabilities assigned to all source nodes and influences specifies a complete joint probability distribution over the variables in the network. We can generate this joint distribution simply by taking the product of all of these assigned distributions. Given a joint distribution, it is straightforward to compute the marginal probability for any value of a variable or Boolean combination of values, by summing over the relevant dimensions of the joint distribution. Similarly, the conditional probability $p(x|e, \xi)$ for any value of x, given evidence e, can be calculated as the ratio $p(x, e|\xi)/p(e|\xi)$. Unfortunately, the size of the joint distribution is exponential in the number of variables. Although this approach is conceptually simple, it requires computational effort that is exponential in the number of variables and is thus impractical for problems with more than a handful of variables.

Exact Methods A key to computational efficiency for inference in belief networks is the exploitation of specified independence relations to avoid having to calculate explicitly the full joint probability distribution. A variety of methods have been developed, each focusing on particular families of belief-network topology.

Kim and Pearl have developed a distributed algorithm for solving singly connected networks, or *polytrees* [89]. The algorithm is linear in the number of variables in the network. In this scheme, each node in the network obtains messages from each of its parent and child nodes, representing all the evidence available from alternative portions of the network. The single-connectedness guarantees that the information in each message is independent and that a scalar is sufficient to represent the total information from each linked node (if nodes are logical). Each time a new observation is made, messages are propagated throughout the network to update the probabilities associated with the other variables.

Unfortunately, most real networks are multiply connected, so more complex methods are required. One approach, developed by Shachter [131], allows computation of the conditional probability distribution for any function f, of a set of variables X, given evidence E, as $p(f(X)|E, \xi)$. This algorithm focuses on a single function of variables rather than on updating the probability of all nodes given evidence. It applies a sequence of operators to the network to reverse the links, using Bayes' theorem, and sums over nodes to eliminate them. The process continues until only the node representing the original probabilistic query remains. Shachter's algorithm will work with multiply connected networks but requires detailed knowledge of the topology to operate. The Shachter algorithm can be significantly more efficient than the brute-force approach of computing the complete joint probability distribution. The extent of the efficiency gains depend on the topology of the network.

Other exact approaches rely on manipulating multiply connected networks to reduce them to singly connected networks [114]. The Kim and Pearl algorithm or similar methods can then be applied to the network. Instantiation of nodes within a loop can effectively break the loop; thus Pearl [114] has suggested focusing on determining the minimal cutsets of nodes that could be instantiated to eliminate loops [114]. These nodes must be instantiated with each possible value (or combination of values). The resulting probabilities are averaged over the results from each instantiation, weighted by the prior probabilities of the instantiated variables.

Lauritzen and Spiegelhalter suggest a different approach based on a reformulation of the belief network [95]. First they "moralize" the graph by adding arcs between all pairs of nodes that have a common successor (i.e., parents with a common child). They then triangulate it, adding arcs so that there are no undirected cycles of more than three nodes without an internal chord. They then identify all the *cliques* –that is, all maximal sets of nodes that are completely interconnected. They show that, by this transformation, any network can be converted into a corresponding singly connected network of cliques. They provide an algorithm for propagation of evidence within this tree of cliques, which is somewhat analogous to the propagation of belief in a singly connected network of variables.

The computational complexity of these algorithms has not been completely analyzed in terms of the topology of the network. However, all of the algorithms are liable to tractability problems if there are many intersecting loops in the network. For example, in the approach of Lauritzen and Spiegelhalter, the joint distribution for each clique must be represented explicitly; thus, the algorithm is exponential in the size of the largest clique. That clique can be very large in a network with many intersecting loops.

More generally, Cooper [23] has shown recently that the general problem of inference in a belief network is NP-hard, so we should not expect to find an exact method that is computationally efficient for arbitrary networks. Nevertheless, exact methods for the tractable solution of specific classes of belief networks may be possible.

Stochastic Methods Researchers have developed various approaches that employ approximation methods; Cooper's result on the NP-hardness of exact probabilistic inference suggests that approximate approaches may be more productive than exact approaches for many cases. One approach, *stochastic simulation*, is attractive because it represents the probabilistic problem as a sample of deterministic, logical cases and reduces the probabilistic representation to a simpler, and perhaps more transparent, logical representation. The accuracy of the representation depends on the sample size or on the number of simulation cycles. We can use standard statistical techniques to estimate the error in the approximation from a given sample size, and we can reduce the uncertainty to an arbitrary degree by increasing the sample size.

Bundy [15] suggested a Monte Carlo sampling approach for computing the probabilities of Boolean combinations of correlated logical variables, which he calls the *incidence calculus*. Henrion [63] developed an extension of this approach for inference in belief networks, termed *probabilistic logic sampling*. In this approach, a belief network is approximately represented by a sample of deterministic cases. For each case or simulation run, each source and entry of conditional probability arcs is randomly generated as a truth value or as a logical implication rule using the specified probabilities. Diagnostic inference is performed by estimating the probability of a hypothesis as the fraction of simulations that give rise to the observed set of evidence. This method is linear in the number of nodes in the network, regardless of the degree of interconnectedness of cycles. Unfortunately, it is exponential in the number of pieces of evidence observed.

Chin and Cooper [18] have used the logic-sampling approach to generate samples of medical cases for simulation purposes. They avoid the exponential complexity of the general problem by rearranging the direction of the links in the network using Shachter's algorithm, so that all observed variables are inputs (sources) to the network. Unfortunately, this is not a general solution to the problem, because the rearrangement is liable to exponential complexity for highly interconnected networks.

Pearl [115] has developed a stochastic-sampling scheme that involves direct propagation in both directions along each influence, rather than solely through the encoded direction, as in logic sampling. In this method, the conditional-probability distribution is computed for each node given all the neighbors in its Markov blanket. First, all the nodes are initialized with random truth values. During simulation, the truth value of a node is updated according to the values of that node's neighbors when the node fires. The node's new truth value is generated at random using the conditional probability for that node given the state of all its neighbors. The probability of each node is estimated as the fraction of simulation cycles for which it is true.

A merit of the Pearl approach is that it could be implemented as a network of parallel-distributed processors, each operating independently, receiving messages from its neighbors and sending messages to them. Unfortunately, as Chin and Cooper demonstrated, simulation approaches are liable to convergence problems when the network contains probabilities that are close to zero or 1. Unlike in the logic-sampling approach, successive cycles are not independent, and the network can get trapped in a state from which it takes many cycles to escape. It remains to be seen whether there are techniques for avoiding this problem.

Bounding Methods As we discussed under "The Problem of Diagnosis," when multiple disorders or faults are possible in a diagnostic problem, the total number of diagnoses is exponential in the number of hypotheses. To compute the exact posterior probability of any diagnosis, $p(D|\mathbf{E},\xi)$, we must compute

$$p(\mathbf{E}|\xi) = \sum_{D_i \in 2^H} p(\mathbf{E}|D_i,\xi)p(D_i|\xi)$$

in the denominator of Bayes' theorem, which involves the exponential task of summing over all of the diagnoses. However, computation of the *ratio* of the probabilities of any two diagnoses is much simpler, because the $p(\mathbf{E}|\xi)$ in the denominator of each cancels out and so does not need to be computed. That is,

$$\frac{p(D_1|\mathbf{E},\xi)}{p(D_2|\mathbf{E},\xi)} = \frac{p(D_1|\xi)p(\mathbf{E}|D_1,\xi)}{p(D_2|\xi)p(\mathbf{E}|D_2,\xi)}$$

Cooper [21] and Peng [118] describe branch-and-bound methods for searching through the space of possible diagnoses, which can identify the most probable diagnoses without examining all possible ones. These methods are able to prune the search by eliminating all extensions of a diagnosis that are provably less probable than the current best, and so can be a great deal more efficient than the exhaustive methods. Peng's method is more efficient but works only for two-level belief networks consisting of a set of faults (level 1) and a set of evidence (level 2), with the only arcs being from faults or disorders to evidence. In this method, the effects of multiple disorders are combined with noisy-OR gates, described earlier.

Bounding methods can be used to compute bounds on the absolute probability for any diagnosis. They sometimes allow us to identify the most probable n diagnoses in a set $\mathbf{D}$ without calculating over the total joint probability space. For example, the partial sum of $p(D_i|\xi)p(\mathbf{E}|D_i,\xi)$ gives a lower bound on $p(\mathbf{E}|\xi)$. Cooper [21] showed how to use this approach to compute upper bounds for absolute posterior probabilities as well. He also gave a related method for computing lower bounds.

Inference Within Influence Diagrams So far, we have focused on inference in belief networks. With influence diagrams, we also must consider the question of how to find the best decision strategy, or decisions that will maximize our expected utility given the available information. Not all influence diagrams represent a well-defined decision problem; those that do are termed *decision networks*. A decision network must have at least one value node and a well-ordered path through all its decision nodes. That is, each decision node must directly precede and directly influence successor decision nodes. Furthermore, immediate predecessors of a decision node must be immediate predecessors of all subsequent decisions. This constraint (referred to as the *no-forgetting* condition) ensures that all information available for a decision is also available for subsequent decisions.

The most popular representation for finding an optimal decision strategy is the use of a decision tree. In a decision tree, each terminal node represents a particular scenario or combination of values for all the uncertain and decision variables. The standard *roll-back* method to evaluate a decision tree is to compute the utility for each terminal node. This method computes the expected utility over the branches at each outcome variable and the maximum expected utility over the alternatives at each decision. In the worst case, this algorithm is exponential in the number of outcome and decision variables. However, there is no need to follow branches with zero probability or decisions that are unavailable, and so the effort may be much smaller in highly asymmetric trees.

A decision network can always be converted into its corresponding decision tree, but this is not necessarily the best way to analyze it. Olmsted [110] and Shachter [130] have developed techniques for operating directly on influence diagrams. The algorithms apply a sequence of operations to the diagram, successively eliminating nodes when their effects have been accounted for through expected value calculations. The operations correspond to (1) applying Bayes' theorem (equivalent to reversing an arc), (2) forming conditional expectations (equivalent to removing a chance node), and (3) maximizing the expected utility (equivalent to removing a decision node). The algorithm's results are the optimal decisions conditional on the information available when each decision is made, and the expected value of the decision strategies. The algorithm will work with multiply connected networks, although it is liable to the same NP-hardness problem that plagues exact probabilistic inference.

One of the benefits of using the influence diagram for representing and solving decision problems is the ease of estimating the value of information. In an influence diagram, value of information can be calculated by adding an arc from a selected chance node to a decision node and re-solving for the optimal decisions. The net improvement in expected value measures how much better decisions could be made given perfect knowledge about the variable. Other value-of-information calculations–for example, for imperfect or noisy information–can be performed in a similar manner.

State of the Art in Inference As just reviewed, there exist a number of exact techniques that can be used effectively for propagating beliefs and for finding optimal decisions for problems of moderate size, even when the latter are multiply connected. The Kim and Pearl algorithm will work efficiently for large belief networks as long as they are singly connected. Although the general inference problem in multiply connected belief networks and influence diagrams has been proven NP-hard, there are a number of promising approximation techniques. There is still considerable room for further research in refining these techniques and in broadening their applicability. Decision analysts generally deal with inference in complex problems by performing sensitivity analysis and eliminating unimportant variables until the model is small enough to be tractable. The development of automated algorithms employing this approach is an interesting area for research. Also, methods for decomposing multiply connected networks into a set of belief network subproblems and for reasoning about the application of combinations of alternative exact and approximate inference methods are promising areas of current investigation [72].

4.4 Explanation in Decision-Theoretic Expert Systems

A frequent criticism of decision-theoretic reasoning is that it is difficult to explain [144, 27, 119]. Teach and Shortliffe [147] identified the ability of an expert system to explain its reasoning strategies and results to users as an important factor in its acceptance. Researchers have constructed systems that give explanations of logical reasoning for applications spanning the range from blocks world [158] to medicine [136, 12, 143, 153, 112]. Unfortunately, relatively little has been done on the explanation of decision-theoretic inference. We shall review some of the ongoing research on techniques for justifying the results of decision-theoretic reasoning strategies.

Evidence Weights One approach to explaining probabilistic inference is to decompose a probabilistic result into a set of evidential subproblems, each focused on the relevance of a particular piece of evidence on the belief in alternative hypotheses. Within each subproblem, the contribution of a piece of evidence to the belief assigned to competing hypotheses is presented. The likelihood ratio and the logarithm of the likelihood ratio, $\ln \lambda(H, E, \xi)$, termed the *weight of evidence*, have been used as the quantitative measures that capture the contribution of different pieces of evidence to belief in competing hypotheses. Under conditions of conditional independence, weights of evidence have the useful property of being additive. That is, the update in belief corresponding to the combined evidence is just the *sum* of individual updates. The naturalness of weights of evidence for acquiring and making inferences with uncertainty was first pointed out by Peirce in 1878 [117]. I.J. Good popularized the measure among philosophers of science and statisticians [47]. Several other researchers, including Turing [47] and Minsky and Selfridge [106], independently found this measure to be useful.

The additive property of evidence weights is conducive to producing informative graphical displays that represent the weights as the length of graphical elements to be added or subtracted. Several expert systems have made use of likelihood ratios and weights of evidence for explaining the relevance of evidence to hypotheses under consideration. Sequences of likelihood ratios are used to explain how evidence affects belief in competing hypotheses in the Glasgow-Dyspepsia expert system for assisting in gastroenterology diagnosis [140], in the Pathfinder system for reasoning about tissue pathology [76], in the Neurex system for diagnosis of neurological findings [124], and in the Medas system for assisting physicians in emergency medicine [5]. Likelihood ratios and weights of evidence also have been optionally converted from graphical to qualitative text descriptions in the Pathfinder and Medas projects. Pathfinder developers investigated the explanation of user-specific multiattribute-utility considerations associated with test decisions. Figure 4 portrays a portion of a consultation with Pathfinder, demonstrating the explanation of probability and utility considerations examined in the generation of a recommendation.

Modulating Completeness and Abstraction Another approach to creating more natural decision-theoretic inference and explanation is to control the level of abstraction at which inference occurs. For example, rather than directly reasoning about the beliefs associated with each disease in response to a pattern of evidence observed in a tissue section under a microscope, a physician may prefer to reason about–and review explanations about–classes of disease, such as inflammatory, infectious, and malignant. At this higher level of abstraction, the uncertain reasoning problem is simplified, and thus easier to understand and explain. A component of explanation research within the Pathfinder project studied the simplification of probabilistic inference through using heuristic abstraction hierarchies to control reasoning. The hierarchies are used to classify diseases into a set of disease groups that depends on the diagnostic problem at hand [77]. Related research on the naturalness of alternative grouping schemes has been conducted by Ben-Bassat and Teeni [6].

Researchers also have investigated the application of decision theory at the metalevel to *control* the construction of explanations of decision-theoretic inference and of mathematical modeling [70, 104]. The research has studied the explicit metareasoning about the costs, benefits, and tradeoffs associated with alternative explanations and with the understandability of alternative reasoning methods. The work centers on the use of multiattribute utility models to reason *about* explanation tradeoffs, such as the tradeoff between the simplicity and the completeness of an explanation.

Qualititative Explanation Determining the most important rationale for a recommendation allows a system to present an explanation qualitatively. Researchers have suggested that, as people may primarily use qualitative mental representations, it is useful to translate explanations of quantitative reasoning–for example, of decision-analytic models–into more qualitative descriptions [65]. Several projects have used this approach.

Langlotz and associates [93] constructed a system called QXQ that explains medical decision analysis problems qualitatively. The system identifies the most important factors affecting a decision and applies a set of explanation heuristics. The quantitative probabilistic and utility models considered by the system are translated into qualitative explanations of why particular decisions have higher expected value than others. The Pathfinder team examined the usefulness of creating qualitative explanations through mapping linguistic modifiers onto probability and utility considerations. A qualitative explanation from Pathfinder justifying an evidence-gathering recommendation is displayed at the bottom of Figure 4. Elsaesser employed a related approach for the qualitative explanation of Bayesian updating [37]. Recent work by Klein has examined the qualitative explanation of multiattribute utility models [90].

```
I recommend that the following features be evaluated:
  Architecture
  Follicles density
  Subcapsular sinuses
  Medullary sinuses

> justify

Which features do you want justified?
   None
  Architecture
  Follicles density
  Subcapsular sinuses
  Medullary sinuses

> architecture

The following table elucidates the discriminating power of this
feature.  The position of the asterisk indicates which of the
two groups of diseases is favored by each value.

Malignant
          Benign
|         |
v         v
. . . . . . . *   preserved
. . . . . . . *   partially obliterated
. . . *. . . .    greatly obliterated
*. . . . . . .    completely obliterated

Pragmatic considerations in the evaluation of the feature,

. . . . . ! . .   tedium
$. . . . . . .    cost
```

```
While it is somewhat tedious to answer,  in part because
it entails a shift in attention from high-power to
low-power, the question Architecture has been recommended
because it is very good for discriminating between two
disease groups and entails no monetary cost.
```

Figure 4: Graphical and Qualitative Explanation Customized for a Nonexpert Pathologist. The explanation includes probability and utility considerations associated with additional information acquisition in the Pathfinder system.

5 Decision-Theoretic Techniques In AI

5.1 Overview

In previous sections, we discussed applications of decision theory to AI problems with obvious relationships to expert-systems research; most research on applying decision theory has had this focus. However, the capabilities of decision-theoretic techniques for reasoning under uncertainty, considering complex preferences, and reasoning about decisions are applicable to other areas of AI. A major component of automated reasoning is decision making about the value of alternative representation and inference strategies for solving a problem. For example, search problems can be recast as decision problems at each node in the search tree. This is not to suggest that decision theory should necessarily be chosen as a reasoning methodology. In many cases, it may be counterproductive to apply decision-theoretic inference and representation techniques explicitly. Even in these cases, decision theory can provide a perspective useful for developing an approach and for reasoning about the usefulness of alternative problem-solving techniques.

In recent years, decision-theory has been applied to a number of problems in AI, including planning, control of inference, perception, learning, problem formulation, temporal reasoning, and nonmonotonic reasoning. We review briefly aspects of these research topics. More comprehensive discussions are found in [86, 98].

5.2 Recent Research

The earliest and most prominent applications of decision theory in AI were in planning research, much of which centers on the *construction* of sequences of actions that will achieve a set of goals. Goal sets and operators are typically Boolean: If a plan does not successfully lead to a goal, it fails. Research on the application of decision theory to planning has focused on the *evaluation* of alternative plans, considering complex preferences based on several attributes. The decision-theoretic paradigm has been used to define the valuation function and the important uncertainties in a problem, as well as to select the best sequence of actions. Feldman and Sproull [40] showed how decision theory could be applied to control the application of planning operators in solving the monkey-and-bananas problem. Coles et al. [20] and Jacobs [84] used utility theory to evaluate alternative plans for robots immersed in an uncertain world. Langlotz et al. applied decision theory to ranking alternative cancer therapy plans within the Oncocin project [92]. Wellman [156] has applied logical theorem-proving techniques to prove the dominance of a set of plans within a qualitative influence-diagram formalism. The qualitative influence-diagram research centers on representing probabilistic dependencies qualitatively based on stochastic dominance.

Decision theory also has been used for the control of inference. Most AI work on the control of inference has been based on heuristic control techniques [26, 38, 2]. Recent research has focused on the potential for decision theory to be useful in planning a problem-solving approach, in evaluating the costs and benefits of alternative inference methods, and in combining the effort of several strategies [48, 71]. Smith [138] and Treitel and Genesereth [148] have applied decision theory to reasoning about the control of logical reasoning. Smith uses expected value notions to select among search paths in database queries. Treitel has explored the costs associated with alternative sequentialization strategies in logical theorem proving. Horvitz [73] has investigated issues surrounding the use of decision theory to control several computational tasks including decision-theoretic inference itself. In this work, alternative reasoning strategies are evaluated through weighing their expected informational benefits with inference-related costs, such as the expense associated with delay. Russell and Wefald [127] have examined the application of decision theory in search for game playing, building upon earlier work by Good [50]. Other recent research on decision-theoretic control, by Fehling and Breese, centers on the application of decision theory to a problem with robot planning [39], considering the costs and benefits of alternative reasoning strategies to a robot decision maker.

Several research projects have examined the representation of the semantics of probabilistic knowledge within predicate calculus. Nilsson's *probabilistic logic* [108] extends the idea of logical entailment to probabilistic domains. Within probabilistic logic, the probability of any sentence in first-order predicate calculus is determined. The probabilities assigned to arbitrary combinations of propositions are based on a logical analysis of alternative possible worlds. In related work, Cooper [22] developed an algorithm for calculating the probability of an arbitrary statement in propositional logic when a belief network is used as the representation of uncertainty.

There also is ongoing work on the logical analysis of belief networks. Pearl [115] has developed logical techniques for reasoning about the allowable decompositions of belief-network problems. Such decomposition techniques focus on issues of *relevance* among portions of a belief network to reformulate an unwieldy inference problem into a set of smaller independent problems that can be solved more efficiently.

Many problems remain in applying automated logical-reasoning techniques developed by AI investigators to the construction of decision models. At the foundations of any decision model are decisions about the propositions to include within the decision basis, which often are based on logical relationships. Several researchers have examined the automated construction of decision models [67, 11, 156]. There are many unanswered questions regarding the automated assembling, pruning, and reasoning about decision models. There is potential to develop tools for assisting engineers with the construction of influence diagrams in the spirit of recent work by Heckerman on the efficient representation of alternative classes of independence among propositions [56].

There have been several research projects examining probabilistic approaches to temporal reasoning. Cooper, et al. [24] developed and implemented a model of probabilistic temporal reasoning that updates belief in competing hypotheses over time as events are observed. Within this work a set of temporal-locality and conditional-independence assumptions were studied. Prototypical functions, representing knowledge about the temporal properties of propositions and about the temporal relationships among propositions considered in the model, are used for updating belief in competing hypotheses over time. Related research by de Zegher-Geets [35] has explored the use of prototypical temporal belief functions among symptoms and diseases for helping physicians to recognize changes in a patient's condition over time. Dean and Kanazawa [34] define a number of temporal predicates and show how functions representing the probability of specified states over time can be used in temporal reasoning. In other work, Tatman¡ [146] made use of the influence diagram representation to identify and to solve classes of temporal decision problems amenable to solution with dynamic programming methods.

It is clear that incompleteness of knowledge and the possibility of changing information are intrinsic to almost all complex real-world problem solving. Nonmonotonic logics deal with new information by formalizing the process of defeating beliefs within a logical framework. Probability theory assigns a continuous measure of belief and provides mechanisms for updating in light of new information. Several researchers are integrating these perspectives as a means of dealing with incomplete and changing information [102, 107, 53].

Decision theory also can be applied to problems involving modal reasoning. Analogous to extensions of first-order predicate calculus to modal reasoning, probabilities can be used to describe the uncertain beliefs that one agent holds about another agent's beliefs [44, 42]. Decision analysis can be applied to communication and cooperation issues. Recent research has examined how an autonomous agent might be endowed with the ability to apply utility theory and probability theory to reason about the knowledge and potential behaviors of another agent [125]. Related work within decision science has investigated the application of decision theory to reasoning about the actions of competing decision makers [121, 157, 142].

Decision theory can provide a framework for considering the relationships among alternative inference strategies. Langlotz et al. [91] describe an attempt to justify heuristic default strategies with decision theory. Other work has described how a "suboptimal" default or approximate strategy could be preferred to a complete decision-theoretic analysis given the cost of reasoning, and the importance of techniques for gracefully degrading the value of a system's performance from a complete analysis to an approximate one as the costs of representation or reasoning resources increase [69]. As we mentioned in the second section of this article, analyses have been carried out in attempts to understand how alternative formalisms for reasoning under uncertainty–such as nonmonotonic reasoning [46], fuzzy set theory [160], and Dempster-Shafer theory [133]–relate to probabilistic reasoning. See Kanal and Lemmer [86, 98] for some detailed analyses of these approaches.

The application of influence diagrams in new areas is facilitated by their relatively unconstrained dependency structure at the level of relation. As an example, machine-vision researchers have applied probabilistic inference to perceptual tasks. In recent research, Binford and Levitt [7] have used a belief network to assign probabilities to alternative plausible three-dimensional objects, given a two-dimensional projection.

Other researchers have examined learning within the decision-theoretic framework. Machine-learning researchers have dwelled almost exclusively on deterministic relationships. Concepts developed in learning research, such as bias and explanation-based generalization, might be extended to learning examples of greater complexity through the integration of decision-theoretic notions. Star [141] described how explanation-based generalization models for learning might be extended to reason about preferences under uncertainty through the application of decision theory.

Decision-theoretic approaches also hold promise for extending AI discovery research. Several research projects have been undertaken to apply probabilistic reasoning to discovery. Within the Radix project [8], probabilistic and logical reasoning were used to control the generation and confirmation of hypotheses about interesting relationships within a large medical database. Cheeseman and associates have studied the automatic induction of a useful set of categories from data acquired by sensors on a wandering robot [17]. Pearl and Verma [116] described logical methods for reformulating belief networks to suggest plausible causal relationships to explain a set of probabilistic relationships.

6 Conclusions

We have reviewed the application of concepts from decision science to AI research. Despite their different perspectives, decision science and AI have common roots and strive for similar goals. We have concentrated on current expert-systems research at the crossroads of AI and decision science. Historically, the development of heuristic reasoning methods arose partly in response to the complexity and poor expressiveness of the early probabilistic expert systems. Popular heuristic schemes, such as the rule-based approach to reasoning under uncertainty, were intended to be more tractable and expressive than probabilistic inference. However, recent work has uncovered inconsistencies in and limitations of these heuristic schemes, and has shown that they can lead to unreliable conclusions. From the decision-theoretic perspective, it is clear that no scheme for reasoning about complex decisions under uncertainty can avoid making assumptions about prior belief and independence, whether these assumptions are implicit or explicit.

Recognition of the difficulties of the heuristic approaches, coupled with the promise of more tractable and expressive decision-theoretic representation and inference strategies seems to be stimulating renewed interest in the decision-theoretic approaches to automated problem solving. In particular, belief networks and influence diagrams are appealing knowledge-representation schemes. They can express knowledge about uncertain beliefs and relationships in both qualitative and more quantitative forms in a flexible, yet principled, fashion. These representations are being used in several areas of expert-systems research, including the development of graphical knowledge-acquisition tools; the search for efficient algorithms for inference and decision in complex, multiply connected belief networks and influence diagrams; and the development of techniques for explaining decision-theoretic reasoning.

Although recent research on the application of decision-science ideas in expert systems seems promising [1, 3, 9, 21, 56, 65, 66, 67, 76], for the most part, only prototype systems have been demonstrated to date. There is urgent need for further research on the sensitivity of various inference schemes to seemingly unrealistic assumptions. Such research could determine the conditions under which the assumptions lead to serious errors. Continuing investigation on the successes and failures of heuristic methods also might lead to the discovery of useful and well-characterized approximation techniques for specific problem-solving contexts.

Moving beyond expert systems, we see substantial opportunities for application of ideas from decision science and AI to planning and problem solving, control of reasoning, speech recognition, vision, temporal reasoning, and learning. However, substantial barriers remain in applying these ideas in automated reasoners. Problems still unresolved include the construction and maintenance of large, coherent knowledge bases; inference in large, arbitrary belief networks and influence diagrams; automation of sensitivity analysis for knowledge engineering and computation; generation and screening of alternatives; qualitative abstraction of decision models and conclusions; and automated construction of decision models. Many of these problems have not been addressed in detail within the expert-systems research program yet are crucial for developing theoretical methods and computational architectures for automated reasoners.

In summary, we believe that the investigation of central problems in representation and inference can be facilitated within the decision-theoretic framework. Conversely, there are indications that AI can make contributions to problems and assist in developing relatively unexplored frontiers in the decision sciences. We anticipate major advances in both AI and decision science based on increased interaction between the disciplines. We hope that this paper will help to promote such collaboration. For now, we temper our enthusiasm about early developments and await the results of ongoing theoretical work as well as the accumulation of experience with the application of decision-science techniques in automated reasoning systems.

7 Acknowledgments

We thank Greg Cooper, Renwick Curry, Lyn Dupre, Michael Fehling, Kenneth Fertig, Michael Genesereth, David Heckerman, Edward Herskovits, Harold Lehmann, and Peter Szolovits for critiquing drafts of this manuscript.

References

[1] A.M. Agogino and A. Rege. IDES: Influence diagram based expert system. *Mathematical modelling*, 8:227–233, 1987.

[2] J.S. Aikens. Representation of control knowledge in expert systems. In *Proceedings of the First Annual National Conference on Artificial Intelligence, Stanford, CA*, pages 121–123. American Association for Artificial Intelligence, Palo Alto, CA, ?? 1980.

[3] S. Andreassen, M. Woldbye, B. Falck, and S.K. Andersen. 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 Kaufman, San Mateo, CA, August 1987.

[4] T. Bayes. An essay towards solving a problem in the doctrine of chances. *Biometrika*, 46:293–298, 1958. Reprint of Bayes' 1763 manuscript.

[5] M. Ben-Bassat, V.K. Carlson, V.K. Puri, M.D. Davenport, J.A. Schriver, M.M. Latif, R. Smith, E.H. Lipnick, and M.H. Weil. Pattern-based interactive diagnosis of multiple disorders: The MEDAS system. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2:148–160, 1980.

[6] M. Ben-Bassat and D. Teeni. Human-oriented information acquistion in sequential pattern classification: Part 1 - single membership classification. *IEEE Transactions on Systems, Man, and Cybernetics*, 14:131–138, 1984.

[7] T.O. Binford, T.S. Levitt, and W.B. Mann. Bayesian inference in model-based machine vision. In L.N. Kanal, J.F. Lemmer, and T.S. Levitt, editors, *Uncertainty in Artificial Intelligence 3*, pages 73–96. North Holland, New York, 1989.

[8] R.L. Blum. Discovery, confirmation and incorporation of causal relationships from a large time-oriented clinical data base: The RX project. In W.J. Clancey and E.H. Shortliffe, editors, *Readings in Medical Artificial Intelligence: The First Decade*, chapter 17, pages 399–425. Addison-Wesley, Reading, Mass., 1984.

[9] J.S. Breese. An expert system for decision analysis in securities trading. In *Proceedings of the First Annual Conference on Expert Systems in Business, New York*, pages 19–26. ??, October 1987.

[10] J.S. Breese. *Knowledge Representation and Inference in Intelligent Decision Systems*. PhD thesis, Department of Engineering-Economic Systems, Stanford University, Stanford, CA, 1987.

[11] J.S. Breese and E. Tse. Integrating logical and probabilistic reasoning for decision making. In *Proceedings of Third Workshop on Uncertainty in Artificial Intelligence, Seattle, WA*, pages 355–362. Association for Uncertainty in Artificial Intelligence, Mountain View, CA, July 1987.

[12] B.G. Buchanan and E.H. Shortliffe, editors. *Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project*. Addison-Wesley, Reading, MA, 1984.

[13] B.G. Buchanan and E.H. Shortliffe. Uncertainty and evidential support. In B.G. Buchanan and E.H. Shortliffe, editors, *Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project*, pages 209–232. Addison-Wesley, Reading, MA, 1984.

[14] D.M. Buede. Structuring value attributes. *Interfaces*, 16:52–62, 1986.

[15] A. Bundy. Incidence calculus: A mechanism for probabilistic reasoning. In *Proceedings of the Workshop on Uncertainty in Artificial Intelligence, Los Angeles, CA*, pages 177–184. Association for Uncertainty in Artificial Intelligence, Mountain View, CA, August 1985.

[16] N. Carrol. Expert systems for clinical diagnosis: Are they worth the effort? *Behavioral Science*, 32:274–292, 1987.

[17] P. Cheeseman, J. Kelly, M. Self, and J. Stutz. Automatic Bayesian induction of classes. In *Proceedings of the NASA Artificial Intelligence Forum, Palo Alto, CA*, pages 224–239. NASA-Ames Research Center, Mountain View, CA, November 1987.

[18] H.L. Chin and G.F. Cooper. Bayesian belief network inference using simulation. In L.N. Kanal, J.F. Lemmer, and T.S. Levitt, editors, *Uncertainty in Artificial Intelligence 3*, pages 129–148. North Holland, New York, 1989.

[19] P.R. Cohen. *Heuristic Reasoning About Uncertainty: An AI Approach.* Pitman, London, 1985.

[20] L.S. Coles, A.M. Robb, P.L. Sinclair, M.H. Smith, and R.R. Sobek. Decision analysis for an experimental robot with unreliable sensors. In *Proceedings of the Fourth International Joint Conference on Artificial Intelligence, Georgia, USSR*, pages 749–757. International Joint Conference on Artificial Intelligence, August 1975.

[21] G.F. Cooper. *NESTOR: A Computer-based Medical Diagnostic Aid that Integrates Causal and Probabilistic Knowledge.* PhD thesis, Computer Science Department, Stanford University, Stanford, CA, November 1984. Rep. No. STAN-CS-84-48. Also numbered HPP-84-48.

[22] G.F. Cooper. An algorithm for computing probabilistic propositions. In L.N. Kanal, J.F. Lemmer, and T.S. Levitt, editors, *Uncertainty in Artificial Intelligence 3*, pages 3–14. North Holland, New York, 1989.

[23] G.F. Cooper. Probabilistic inference using belief networks is NP-hard. *Artificial Intelligence*, 42:393–405, 1990.

[24] G.F. Cooper, E.J. Horvitz, and D.E. Heckerman. A model for temporal probabilistic reasoning. Technical Report KSL-88-30, Stanford University, Section on Medical Informatics, Stanford, California, April 1988.

[25] R. Cox. Probability, frequency and reasonable expectation. *American Journal of Physics*, 14:1–13, 1946.

[26] R. Davis. Meta-rules: Reasoning about control. *Artificial Intelligence*, 15:179–222, 1980.

[27] R. Davis. Consultation, knowledge acquisition, and instruction. In P. Szolovits, editor, *Artificial Intelligence In Medicine*, pages 57–78. Westview Press, Boulder, CO, 1982.

[28] R. Davis. Diagnosis via causal reasoning: Paths of interaction and the locality principle. In *Proceedings of the Eighth International Joint Conference on Artificial Intelligence, Karlsruhe, West Germany*, pages 88–94. National Conference on Artificial Intelligence, August 1983.

[29] R.M. Dawes and B. Corrigan. Linear models in decision making. *Psychological Bulletin*, 81:95–106, 1974.

[30] F.T. de Dombal, D.J. Leaper, J.C. Horrocks, J.R. Staniland, and A.P. McCain. Human and computer-aided diagnosis of abdominal pain: further report with emphasis on performance. *British Medical Journal*, 1:376–380, 1974.

[31] F.T. de Dombal, D.J. Leaper, J.R. Staniland, A.P. McCann, and J.C. Horrocks. Computer-aided diagnosis of acute abdominal pain. *British Medical Journal*, 2:9–13, 1972.

[32] B. de Finetti. *Theory of Probability.* Wiley and Sons, New York, 1970.

[33] J. de Kleer and B. Williams. Diagnosing multiple faults. *Artificial Intelligence*, 32:97–130, 1987.

[34] T.L. Dean and K. Kanazawa. Probabilistic temporal reasoning. Technical report, Brown University, May, 1988.

[35] I. deZegher Geets. Idefix: Intelligent summarization of a time-oriented medical database. Technical Report KSL-88-34, Section on Medical Informatics, Stanford University, Stanford, California, August 1987.

[36] R. Duda, J. Gaschnig, and P. Hart. Model design in the PROSPECTOR consultant system for mineral exploration. In D. Michie, editor, *Expert Systems in the Microelectronic Age*, pages 153–167. Edinburg University Press, Edinburgh, Scotland, 1979.

[37] C. Elsaesser. Explanation of probabilistic inference. In L.N. Kanal, J.F. Lemmer, and T.S. Levitt, editors, *Uncertainty in Artificial Intelligence 3*, pages 387–400. North Holland, New York, 1989.

[38] L. Erman, F. Hayes-Roth, V.R Lesser, and D.R. Reddy. The HEARSAY-II speech understanding system: integrating knowledge to resolve uncertainty. *ACM Computing Surveys*, 12:213–253, 1980.

[39] M.R. Fehling and J.S. Breese. A computational model for the decision-theoretic control of problem solving under uncertainty. Technical Report Rockwell Technical Report 837-88-5, Rockwell International Science Center, April 1988.

[40] J.A. Feldman and R.F. Sproull. Decision theory and artificial intelligence ii: the hungry monkey. *Cognitive Science*, 1:158–192, 1975.

[41] P.C. Fishburn. Subjective expected utility: A review of normative theories. *Theory and Decision*, 13:139–199, 1981.

[42] A.M. Frisch and P. Haddawy. Probability as a modal operator. Technical report, Computer Science Department, University of Illinois, April 1988.

[43] D.G. Fryback. Bayes' theorem and conditional nonindependence of data in medical diagnosis. *Computers and Biomedical Research*, 11:423–434, 1978.

[44] H. Gaifman. A theory of higher order probabilities. In *Proceedings of the 1986 Conference of Theoretical Aspects of Reasoning About Knowledge, ??*, page ?? Morgan Kaufmann, San Mateo, CA, ?? 1986.

[45] M. Genesereth. The use of design descriptions in automated diagnosis. *Artificial Intelligence*, 24:311–319, 1984.

[46] M.L. Ginsberg. *Readings in Nonmonotonic Logic*. Morgan Kaufman, San Mateo, CA, 1987.

[47] I.J. Good. *Probability and the Weighing of Evidence*. Hafners, New York, 1950.

[48] I.J. Good. Rational decisions. *J. R. Statist. Soc. B*, 14:107–114, 1952.

[49] I.J. Good. Subjective probability as the measure of a non-measureable set. In *Logic, Methodology, and Philosophy of Science: Proceedings of the 1960 International Congress*, pages 319–329. Stanford University Press, 1962. Also in Good Thinking: The Foundations of Probability and Its Applications, I.J. Good, University of Minnesota Press, 1983.

[50] I.J. Good. Dynamic probability, computer chess, and the measurement of knowledge. In E.W. Elcock and Michie D., editors, *Machine Intelligence*, pages 139–150. Wiley, New York, 1977.

[51] G.A. Gorry. Computer-assisted clinical decision making. *Methods of Information in Medicine*, 12:45–51, 1973.

[52] G.A. Gorry and G.O. Barnett. Experience with a model of sequential diagnosis. *Computers and Biomedical Research*, 1:490–507, 1968.

[53] B. Grosof. Nonmonotonic reasoning in probabilistic reasoning. In J.F. Lemmer and L.N. Kanal, editors, *Uncertainty in Artificial Intelligence 2*, pages 237–250. North Holland, New York, 1988.

[54] K.R. Hammond. Towards a unified approach to the study of expert judgment. In J.L. Mumpower, editor, *Expert Judgment and Expert Systems*, pages 1–16. Springer-Verlag, Berlin, Heidelberg, 1987.

[55] D.E. Heckerman. Probabilistic interpretations for MYCIN's certainty factors. In L.N. Kanal and J.F. Lemmer, editors, *Uncertainty in Artificial Intelligence*, pages 167–196. North Holland, New York, 1986.

[56] D.E. Heckerman. Formalizing heuristic methods for reasoning with uncertainty. Technical Report KSL-88-07, Medical Computer Science Group, Section on Medical Informatics, Stanford University, Stanford, CA, May 1987.

[57] D.E. Heckerman. An axiomatic framework for belief updates. In J.F. Lemmer and L.N. Kanal, editors, *Uncertainty in Artificial Intelligence 2*, pages 11–22. North Holland, New York, 1988.

[58] D.E. Heckerman. An empirical comparison of three inference methods. In R. Shachter T.S. Levitt, J. Lemmer, and L.N. Kanal, editors, *Uncertainty in Artificial Intelligence 4*. North Holland, New York, 1990.

[59] D.E. Heckerman and E.J. Horvitz. 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, July 1987.

[60] D.E. Heckerman and H. Jimison. A perspective on confidence and its use in focusing attention during knowledge acquisition. In L. Kanal, T. Levitt, and J. Lemmer, editors, *Uncertainty in Artificial Intelligence 3*, pages 123–131. North Holland, New York, 1989.

[61] M. Henrion. *The Value of Knowing How Little You Know: The advantages of a probabilistic treatment of uncertainty in policy analysis.* PhD thesis, Department of Engineering and Public Policy, Carnegie Mellon University, 1982.

[62] M. Henrion. Uncertainty in artificial intelligence: Is probability epistemologically and heuristically adequate? In J.L. Mumpower, editor, *Expert Judgment and Expert Systems*, pages 105–130. Springer-Verlag, Berlin, Heidelberg, 1987.

[63] M. Henrion. Propagation of uncertainty by probabilistic logic sampling in Bayes' networks. In J.F. Lemmer and L.N. Kanal, editors, *Uncertainty in Artificial Intelligence 2*, pages 149–164. North Holland, New York, 1988.

[64] M. Henrion. Some practical issues in constructing belief networks. In L.N. Kanal, J.F. Lemmer, and T.S. Levitt, editors, *Uncertainty in Artificial Intelligence 3*, pages 161–174. North Holland, New York, 1989.

[65] M. Henrion and D.R. Cooley. An experimental comparison of knowedge engineering for expert systems and for decision analysis. In *Proceedings AAAI-87 Sixth National Conference on Artificial Intelligence, Seattle, WA*, pages 471–476. Morgan Kaufmann, San Mateo, CA, July 1987.

[66] M.C. Higgins. NAIVE: A method for representing uncertainty and temporal relationships in an automated reasoner. In *Proceedings of Third Workshop on Uncertainty in Artificial Intelligence, Seattle, WA*, pages 140–147. Association for Uncertainty in Artificial Intelligence, Mountain View, CA, 1987.

[67] S. Holtzman. *Intelligent Decision Systems.* PhD thesis, Department of Engineering–Economic Systems, Stanford University, Stanford, CA, March 1985.

[68] S. Holtzman and J.S. Breese. Exact reasoning about uncertainty: On the design of expert systems for decision support. In L.N. Kanal and J.F. Lemmer, editors, *Uncertainty in Artificial Intelligence*, pages 339–345. North Holland, New York, 1986.

[69] E.J. Horvitz. Toward a science of expert systems. *Proceedings of the 18th Symposium on the Interface of Computer Science and Statistics*, pages 45–52, March 1986.

[70] E.J. Horvitz. A multiattribute utility approach to inference understandability and explanation. Technical Report KSL-28-87, Medical Computer Science Group, Section on Medical Informatics, Stanford University, Stanford, CA, March 1987.

[71] E.J. Horvitz. Problem-solving design: Reasoning about computational value, tradeoffs, and resources. In *Proceedings of the NASA Artificial Intelligence Forum, Palo Alto, CA*, pages 26–43. NASA-Ames Research Center, Mountain View, CA, November 1987.

[72] E.J. Horvitz. Reasoning about beliefs and actions under computational resource constraints. In *Proceedings of Third Workshop on Uncertainty in Artificial Intelligence*, pages 429–444, Seattle, Washington, July 1987. American Association for Artificial Intelligence. To appear in L. Kanal, T. Levitt, and J. Lemmer, ed., *Uncertainty in Artificial Intelligence 3,* Elsevier, 1988 (in press).

[73] E.J. Horvitz. Reasoning under varying and uncertain resource constraints. In *Proceedings of AAAI-88, Seventh National Conference on Artificial Intelligence, Minneapolis, MN*, pages 111–116. Morgan Kaufmann, San Mateo, CA, August 1988.

[74] E.J. Horvitz and D.E. Heckerman. The inconsistent use of measures of certainty in artificial intelligence research. In L.N. Kanal and J.F. Lemmer, editors, *Uncertainty in Artificial Intelligence*, pages 137–151. North Holland, New York, 1986.

[75] E.J. Horvitz, D.E. Heckerman, and C.P. Langlotz. 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, August 1986.

[76] E.J. Horvitz, D.E. Heckerman, B.N. Nathwani, and L.M. Fagan. Diagnostic strategies in the hypothesis-directed Pathfinder system. In *Proceedings of the First Conference on Artificial Intelligence Applications, Denver, CO*, pages 630–636. IEEE, Silver Springs, MD, December 1984.

[77] E.J. Horvitz, D.E. Heckerman, B.N. Nathwani, and L.M. Fagan. The use of a heuristic problem-solving hierarchy to facilitate the explanation of hypothesis-directed reasoning. In *Proceedings of Medinfo, Washington, DC*, pages 27–31. North Holland, New York, October 1986.

[78] R.A. Howard. Decision analysis: Applied decision theory. In D.B. Hertz and J. Melese, editors, *Proceedings of the Fourth International Conference on Operational Research*, pages 55–71. Wiley-Interscience, 1966.

[79] R.A. Howard. The foundations of decision analysis. *IEEE Transactions on Systems Science, and Cybernetics*, 4:211–219, 1968.

[80] R.A. Howard. Risk preference. In R.A. Howard and J.E. Matheson, editors, *Readings on the Principles and Applications of Decision Analysis*, volume II, chapter 34,, pages 629–663. Strategic Decisions Group, Menlo Park, Ca., 1970.

[81] R.A. Howard. Proximal decision analysis. *Management Science*, 17, 1971.

[82] R.A. Howard and J.E. Matheson. Influence diagrams. In R.A. Howard and J.E. Matheson, editors, *Readings on the Principles and Applications of Decision Analysis*, volume II, pages 721–762. Strategic Decisions Group, Menlo Park, CA, 1981.

[83] R.A. Howard and J.E. Matheson, editors. *Readings on the Principles and Applications of Decision Analysis*. Strategic Decisions Group, Menlo Park, Ca., 1984.

[84] W. Jacobs and M. Keifer. Robot decisions based on maximizing utility. In *Proceedings of the Third International Joint Conference on Artificial Intelligence, ??*, pages 402–411. International Joint Conference on Artificial Intelligence, August 1973.

[85] D. Kahneman, P. Slovic, and A. Tversky, editors. *Judgment Under Uncertainty: Heuristics and Biases*. Cambridge University Press, New York, 1982.

[86] L.N. Kanal and J.F. Lemmer, editors. *Uncertainty in Artificial Intelligence*. North Holland, New York, 1986.

[87] R.L. Keeney. The art of assessing multi-attribute utility functions. *Organizational Behavior and Human Performance*, 19:267–310, 1977.

[88] R.L. Keeney and H. Raiffa. *Decisions with Multiple Objectives: Preferences and Value Trade-offs*. Wiley and Sons, New York, 1976.

[89] J.H. Kim and J. Pearl. A computational model for causal and diagnostic reasoning in inference engines. In *Proceedings of the 8th International Joint Conference on Artificial Intelligence, Karlsruhe, West Germany*, pages 190–193. International Joint Conference on Artificial Intelligence, August 1983.

[90] D. Klein. Explaining and refining decision-theoretic choices. Technical Report MS-CIS-87-57 (LINC 74), Dept. of Computer and Information Science, University of Pennsylvania, 1987.

[91] C.P. Langlotz, L.M. Fagan, S.W. Tu, B.I. Sikic, and E.H. Shortliffe. Combining artificial intelligence and decision analysis for automated therapy planning assistance. In *Proceedings of Medinfo, Washington, DC*. North Holland, New York, October 1986.

[92] C.P. Langlotz, L.M. Fagan, S.W. Tu, J. Williams, and B.I. Sikic. ONYX: An architecture for planning in uncertain environments. In *Proceedings of the Ninth International Joint Conference on Artificial Intelligence, Los Angeles, CA*, pages 447–449. International Joint Conference on Artificial Intelligence??, August 1985.

[93] C.P. Langlotz, E.H. Shortliffe, and L.M. Fagan. A methodology for computer-based explanation of decision analysis. Technical Report KSL-86-57, Stanford University, November 1986.

[94] P.S. Laplace. *A Philosphical Essay on Probabilities.* Wiley and Sons, New York, 1902. Translated from the 6th French Edition, 1812.

[95] S.L. Lauritzen and D.J. Spiegelhalter. Fast manipulation of probabilities with local representations with applications to expert systems. Technical Report R-87-7, Institute of Electronic Systems, Aalborg University, Aalborg, Denmark, March 1987.

[96] R.S. Lehman. On confirmation and rational betting. *Journal of Symbolic Logic*, 20:251–262, 1955.

[97] J.F. Lemmer. Generalized Bayesian updating of incompletely specified distributions. *Large Scale Systems*, 5, 1983.

[98] J.F Lemmer and L.N. Kanal, editors. *Uncertainty in Artificial Intelligence 2.* North Holland, New York, 1988.

[99] S. Lichenstein, B. Fishoff, and L.D. Philipps. Calibration of probabilities: The state of the art to 1980. In D. Kahneman, P. Slovic, and A. Tversky, editors, *Judgement under uncertainty: Heuristics and biases*, pages 275–324. Cambridge University Press, New York, 1982.

[100] D.V. Lindley. Scoring rules and the inevitability of probability. *International Statistical Review*, 50:1–26, 1982.

[101] Logan. *The Value of Probability Assessment.* PhD thesis, Department of Engineering-Economic Systems, Stanford University, March 1985.

[102] R. Loui. *Theory and Computation of Uncertain Inference.* PhD thesis, Department of Computer Science, University of Rochester, 1987. Also available as TR-228, University of Rochester, Department ofComputer Science.

[103] J. McDermott. R1: A rule-based configurer of computer systems. *Artificial Intelligence*, 19:39–88, 1982.

[104] J. Mclaughlin. The utility-directed presentation of graphical simulation. Technical Report TR-87-59, Stanford University, December 1987.

[105] P. McNamee and J. Celona. *Decision Analysis for the Professional with Supertree.* The Scientific Press, Redwood City, 1987.

[106] M. Minsky and O.G. Selfridge. Learning in random nets. In C. Cherry, editor, *Information Theory*, pages 335–347. Butterworths, London, 1961.

[107] E. Neufeld and D. Poole. Towards solving the multiple extension problem: Combining defaults and probabilities. In L.N. Kanal, J.F. Lemmer, and T.S. Levitt, editors, *Uncertainty in Artificial Intelligence 3*, pages 35–44. North Holland, New York, 1989.

[108] N. Nilsson. Probabilistic logic. *Artificial Intelligence*, 28:71–87, 1986.

[109] M.J. Norusis and J.A. Jacquez. Diagnosis 1: Symptom non-independence in mathematical models for diagnosis. *Computers and Biomedical Research*, 8:156–172, 1975.

[110] S.M. Olmsted. *On Representing and Solving Decision Problems.* PhD thesis, Department of Engineering-Economic Systems, Stanford University, December 1983.

[111] D.L. Owen. The use of influence diagrams in structuring complex decision problems. In R.A. Howard and J.E. Matheson, editors, *Readings on the Principles and Applications of Decision Analysis*, volume II, chapter 38, pages 763–771. Strategic Decisions Group, Menlo Park, Ca., 1978.

[112] R.S. Patil. *Causal representation of patient illness for electrolyte and acid-base diagnosis.* PhD thesis, Computer Science Department, Massachusetts Institute of Technology, October 1981.

[113] E.A. Patrick. Review of pattern recognition in medicine. *IEEE Transactions on Systems, Man and Cybernetics*, 6, 1977.

[114] J. Pearl. Fusion, propagation, and structuring in belief networks. *Artificial Intelligence*, 29:241–288, 1986.

[115] J. Pearl. Evidential reasoning using stochastic simulation of causal models. *Artificial Intelligence*, 32:245–257, 1987.

[116] J. Pearl and T.S. Verma. The logic of representing dependencies by directed graphs. In *Proceedings AAAI-87 Sixth National Conference on Artificial Intelligence, Seattle, WA.* American Association for Artificial Intelligence, Palo Alto, CA, July 1987.

[117] C.S. Peirce. The probability of induction. In *The World of Mathematics, V.2*, pages 1341–1354. Simon and Shuster, New York, 1956.

[118] Y. Peng. *A Formalization of Parsimonious Covering and Probabilistic Reasoning in Abductive Diagnostic Inference.* PhD thesis, Department of Computer Science, University of Maryland, 1986. TR-1615.

[119] P.E. Politser. Explanations of statistical concepts: Can they penetrate the haze of Bayes? *Methods of Information in Medicine*, 23:99–108, 1984.

[120] H. Raiffa. *Decision Analysis: Introductory Lectures on Choice Under Uncertainty.* Addison-Wesley, Reading, Ma., 1968.

[121] H. Raiffa. *The Art and Science of Negotiation.* Belknap Press, Cambridge, Ma., 1982.

[122] H. Raiffa and R. Schlaifer. *Applied Statistical Decision Theory.* Harvard University Press, Cambridge, Ma., 1961.

[123] J.A. Reggia. Diagnostic expert systems based on a set covering model. *International Journal of Man-Machine Studies*, 19:437–460, 1983.

[124] J.A. Reggia and B.T. Perricone. Answer justification in medical decision support systems based on Bayesian classification. *Computers in Biology and Medicine*, 15:161–167, 1985.

[125] J.S. Rosenschein and J.S. Breese. Communication-free interactions among rational agents: A probabilistic approach. In *Proceedings of the 1988 AAAI Workshop on Distributed Artificial Intelligence, ??* Morgan Kaufmann, San Mateo, CA, May 1988.

[126] W.F. Rousseau. A method for computing probabilities in complex situations. Technical Report 6252-2, Center for Systems Research, Stanford University, Stanford, CA, May 1968.

[127] S.J. Russell and E. Wefald. Multi-level decision-theoretic search. In *Proceedings of the AAAI Spring Symposium on Game Playing, Stanford, CA*, pages 3–7. AAAI, March 1988.

[128] R.K. Sarin. Strength of preference and risky choice. *Operations Research*, 30:982–997, 1982.

[129] L.J. Savage. *The Foundations of Statistics.* Dover, New York, 1972. 2nd edition, First edition 1954.

[130] R.D. Shachter. Evaluating influence diagrams. *Operations Research*, 34:871–882, 1986.

[131] R.D. Shachter. Probabilistic inference and influence diagrams. *Operations Research*, 36:589–604, 1988.

[132] R.D. Shachter and D.E. Heckerman. Thinking backward for knowledge acquisition. *AI Magazine*, 8:55–63, 1987.

[133] G. Shafer. Probability judgment in artificial intelligence. In L.N. Kanal and J.F. Lemmer, editors, *Uncertainty in Artificial Intelligence.* North-Holland, New York, 1986.

[134] A. Shimony. Coherence and the axioms of confirmation. *Journal of Symbolic Logic*, 20:1–28, 1955.

[135] E.H. Shortliffe. *Computer-based Medical Consultations: MYCIN.* North Holland, New York, 1976.

[136] E.H. Shortliffe. Explanation capabilities for medical consultation systems. In *Proceedings of American Association of Medical Systems and Informatics 1984, Third Spring Joint National Conference, San Francisco, CA*, pages 193–197. AAMSI, 1984.

[137] H.A. Simon. A behavioral model of rational choice. *Quarterly Journal of Economics*, 69:99–118, 1955.

[138] D.E. Smith. Controlling inference. Technical Report STAN-CS-86-1107, Computer Science Department, Stanford University, April 1986.

[139] C.S. Spetzler and C.S. Stael von Holstein. Probability encoding in decision analysis. *Management Science*, 22:340–358, 1975.

[140] D.J. Spiegelhalter and R.P. Knill-Jones. Statistical and knowledge-based approaches to clinical decision support systems, with an application in gastroenterology. *Journal of the Royal Statistical Society*, 147:35–77, 1984.

[141] S. Star. Theory-based inductive learning: An integration of symbolic and quantitative methods. In L.N. Kanal, J.F. Lemmer, and T.S. Levitt, editors, *Uncertainty in Artificial Intelligence 3*. North Holland, New York, 1989.

[142] R.S. Strait. *Decision Analysis of Strategic Interaction*. PhD thesis, Department of Engineering-Economic Systems, Stanford University, 1987. Also available as UCRL-53825, Lawrence Livermore National Laboratory.

[143] W. Swartout. XPLAIN: A system for creating and explaining expert consulting systems. *Artificial Intelligence*, 21:285–325, September 1983.

[144] P. Szolovits. Artificial intelligence in medicine. In P. Szolovits, editor, *Artificial Intelligence In Medicine*, pages 1–19. Westview Press, Boulder, CO, 1982.

[145] P. Szolovits and S.G. Pauker. Categorical and probabilistic reasoning in medical diagnosis. *Artificial Intelligence*, 11:115–144, 1978.

[146] J.A. Tatman. *Knowledge Representation and Inference in Intelligent Decision Systems*. PhD thesis, Department of Engineering-Economic Systems, Stanford University, 1985.

[147] R.L. Teach and E.H. Shortliffe. An analysis of physician attitudes regarding computer-based clinical consultation systems. *Computers and Biomedical Research*, 14:542–558, 1981.

[148] R. Treitel and M.R. Genesereth. Choosing directions for rules. In *Proceedings AAAI-86 Fifth National Conference on Artificial Intelligence, Philadelphia, PA*, pages 153–157. Morgan Kaufmann, San Mateo, CA, August 1986.

[149] M. Tribus. *Rational Descriptions, Decisions, and Designs*. Pergamon Press, New York, 1969.

[150] M. Tribus. What do we mean by rational? In *Rational Descriptions, Decisions and Designs*, pages 1–32. Pergamon Press, New York, 1969.

[151] J. von Neumann and O. Morgenstern. *Theory of Games and Economic Behavior*. Princeton University Press, Princeton, NJ, 1947.

[152] D. Von Winterfeldt and W. Edwards. *Decision Analysis and Behavioral Research.* Cambridge University Press, New York, 1986.

[153] J.W. Wallis and E.H. Shortliffe. Explanatory power for medical expert systems: Studies in the representation of causal relationships for clinical consultations. *Methods of Information in Medicine*, 21:127–136, 1982.

[154] H.R. Warner, A.F. Toronto, L.G. Veasy, and R. Stephenson. A mathematical approach to medical diagnosis: Application to congenital heart disease. *Journal of the American Medical Association*, 177:177–183, 1961.

[155] M.P. Wellman. Reasoning about preference models. Technical Report MIT/LCS/TR-340, Laboratory for Computer Science, Massachusetts Institute of Technology, May 1985.

[156] M.P. Wellman. Formulation of tradeoffs in planning under uncertainty. Technical Report MIT/LCS/TN-332, Laboratory for Computer Science, Massachusetts Institute of Technology, Cambridge, MA, 1987.

[157] J. Wilson. Subjective probability and the prisoner's dilemma. *Management Science*, 32:45–55, 1986.

[158] T. Winograd. A computer program for understanding natural language. Technical Report TR-17, Massachusetts Institute of Technology Artificial Intelligence Laboratory, 1971.

[159] B. Wise. *An Experimental Comparison of Uncertain Inference Systems.* PhD thesis, The Robotics Institute and Department of Engineering and Public Policy, Carnegie–Mellon University, Pittsburgh, PA, June 1986.

[160] L.A. Zadeh. The role of fuzzy logic in the management of uncertainty in expert systems. *Fuzzy Sets and Systems*, 11:199–227, 1983.