Decision Analysis Techniques for Knowledge Acquisition: Combining Information and Preferences using *Aquinas*

Jeffrey M. Bradshaw and John H. Boose
Knowledge Systems Laboratory, Advanced Technology Center, Boeing Computer Services
P.O. Box 24346, M/S 7L-64, Seattle, Washington 98124 USA (206) 865-3422

ABSTRACT

The field of decision analysis is concerned with the application of formal theories of probability and utility to the guidance of action. Decision analysis has been used for many years as a way to gain insight regarding decisions that involve significant amounts of uncertain information and complex preference issues, but it has been largely overlooked by knowledge-based system researchers. This paper illustrates the value of incorporating decision analysis insights and techniques into the knowledge acquisition and decision making process. This approach is being implemented within *Aquinas*, an automated knowledge acquisition and decision support tool based on personal construct theory that is under development at Boeing Computer Services. The need for explicit preference models in knowledge-based systems will be shown. The modeling of problems will be viewed from the perspectives of decision analysis and personal construct theory. We will outline the approach of *Aquinas* and then present an example that illustrates how preferences can be used to guide the knowledge acquisition process and the selection of alternatives in decision making. Techniques for combining supervised and unsupervised inductive learning from data with expert judgment, and integration of knowledge and inference methods at varying levels of precision will be presented. Personal construct theory and decision theory are shown to be complementary: the former provides a plausible account of the dynamics of model formulation and revision, while the latter provides a consistent framework for model evaluation. Applied personal construct theory (in the form of tools such as *Aquinas*) and applied decision theory (in the form of decision analysis) are moving along convergent paths. We see the approach in this paper as the first step toward a full integration of insights from the two disciplines and their respective repertory grid and influence diagram representations.

1 THE CASE FOR EXPLICIT MODELING OF PREFERENCES

1.1 PREFERENCE MODELING AND KNOWLEDGE-BASED SYSTEMS

Many knowledge-based systems are prescriptive in nature. They aim not only to describe situations but also to recommend specific actions. Recommendations made by such systems depend on: the *alternatives* available, *information* about consequences associated with the alternatives, and *preferences* among these consequences. Research in building knowledge-based systems has typically focused on the first two considerations — many approaches have been proposed for structuring and eliciting alternatives and for modeling potentially uncertain information. By contrast, relatively little effort has been made in the knowledge acquisition community toward understanding how to explicitly represent and quantify preferences. The neglect of this issue limits the effectiveness of knowledge-based systems for many types of problems.

Example: Preferences in MYCIN. Knowledge-based systems typically treat preferences implicitly and heuristically, making no provision for value structures differing from those built into the system. Several recent papers have discussed the need for explicit preference models to be included in the knowledge engineering process (e.g., Henrion, 1987; Henrion & Cooley, 1987; Holtzman, 1989; Horvitz *et al.,* 1988; Keeney, 1986a; Langlotz, Shortliffe, & Fagan, 1986).

In their discussion of preferences, Langlotz, *et al.* (1986) cite an example rule from MYCIN (see Figure 1). This heuristic captures a physician's knowledge that tetracycline therapy should be avoided for children because it may cause dental staining.

If:
 1) The therapy under consideration is tetracycline
 2) The age (in years) of the patient is less than 8
Then:
 There is strongly suggestive evidence (.8) that tetracycline is not an appropriate therapy for use against the organism.

Figure 1: The MYCIN tetracycline heuristic, slightly simplified for illustration purposes.

Clancey (1983) gives a possible chain of four support rules for this heuristic (Figure 2). The first three inferences have to do with how one event is related to the occurrence of the next. The fourth, however, is a compiled plan of action based on the inference chain. Langlotz *et al*. make the point that no matter how finely we break down a chain of reasoning, one rule in the chain will always recommend action based on the situation. Action recommendations always presuppose a set of preferences, either stated or implied, that cannot be derived from the logic of evidence.

tetracycline in youngster
=> chelation of the drug in growing bones
=> teeth discoloration
=> undesirable body change
=> don't administer tetracycline

Figure 2: A justification for the tetracycline heuristic from Clancey (1983).

Because of the nature of heuristics, it is difficult to explicitly and flexibly represent the unique circumstances and tradeoffs that may justify an exception to the heuristic[1]. What if, for example, one or more of the following were found to be true:

- the infecting organism were resistant to all drugs except tetracycline?

- the only undesirable bodily change that tetracycline caused was minor intestinal distress?

-the probability of staining due to tetracycline for a particular patient was only 1 in 100? 1 in 1000?

When tradeoffs are implicitly embedded within heuristics, it becomes impractical to ask, let alone answer, such questions. For example, we could modify the knowledge base by adding additional premises to the rule in Figure 1:

3) The organism can be treated by something other than tetracycline
4) There is evidence that tetracycline will cause significant intestinal distress
5) The probability of dental staining due to tetracycline for the patient is less than .01

Figure 3: Additional premises for the MYCIN tetracycline heuristic.

But since the strength of our recommendation may vary depending on the circumstances present in a given situation, we would need to add a separate rule for action for each particular combination of evidence. While the addition of such heuristics may mimic the behavior of the clinician in most circumstances, it seems unnaturally cumbersome to represent the subjective experience of continuous tradeoff parameters as a list of discrete rules of thumb. Representing the knowledge in this form makes it impossible to vary the parameters of preference tradeoffs (e.g., risk of dental staining vs. effectiveness of tetracycline vs. cost of treatment) smoothly in response to differences in situation and preferences between patients. While it is possible to muster empirical arguments for the truth or falsity of some *evidential* claim, the judgments of *utility* that guide recommendations and action (given that evidence) are inherently subjective: some patients are more willing to take risks than others; some are more concerned about treatment effectiveness; some are more able or willing than others to pay for expensive alternatives that minimize risk. The greater the stakes of the decision, the more serious the consequences of implicit, inflexible representations of preferences.

1.2 THE ADVANTAGES OF EXPLICIT PREFERENCE MODELING

Explicit modeling of preferences has several advantages:

1. Unique tradeoffs for a given situation can be addressed. If a particular child had an unusual resistance to dental staining, treatment decision tradeoffs could be changed directly rather than through the fine tuning of certainty factors.

2. The impact of a particular piece of evidence on a decision can be evaluated. By performing *sensitivity analysis,* we can examine whether a particular piece of evidence in favor of a diagnosis will have any real effect on the treatment decision.

3. The value of obtaining additional information can be ascertained. We can ask, "What is the most I should pay to conduct a test of the child's susceptibility to dental staining?"

[1] Weinberg (1975) the problem of heuristic devices as being that they "don't tell you when to stop." He calls this "The Banana Principle" after the little boy who said, "Today we learned how to spell 'banana,' but we didn't learn when to stop."

4. The value of being able to control an uncertain variable can be determined. If there were a drug that effectively eliminated the possibility of dental staining due to tetracycline, for example, we could assess the value of such a drug in a given situation.

5. Risk attitude and time preference can be explicitly taken into account. In some situations, it is worthwhile to model the patient's attitude toward risk before making a recommendation. Time-critical situations may also require explicit modeling of the risks and benefits of delaying or hastening treatment.

6. System recommendations can be expressed in value terms. It may be important for a system to not only recommend the favored treatment plan, but also estimate its total cost or benefits in some meaningful unit of measurement. The quantification and joint measurement of "intangibles" such as pain and quality of life can also be addressed.

We are implementing methods for the explicit modeling of information and preferences in *Aquinas*, an automated knowledge acquisition tool under development at Boeing Computer Services (Boose and Bradshaw, 1987a, 1987b). This is accomplished through the use of model *evaluation* techniques from decision theory and model *formulation* techniques from personal construct theory.

Section 2 describes decision analysis and personal construct theory approaches to decision making. We will see how decision analysis, as *applied* decision theory, provides an effective approach to the explicit representation and numerical analysis of information, preferences, and alternatives (sections 2.1 and 2.2). Personal construct theory complements the decision analytic perspective by giving an account of the dynamics of model formulation (section 2.3). The combination of these two methodologies in *Aquinas* could provide a powerful and practical approach for the automation of many kinds of knowledge acquisition problems (section 2.4).

Section 3 applies the approach to an example illustrating how decision analytic techniques may be applied to knowledge acquisition[2]. We will introduce the problem (section 3.1), then build a preliminary model of alternatives and preferences (section 3.2). We will demonstrate how the use of decision analytic techniques for model appraisal helps decision makers focus efforts in model refinement on components most worthy of attention (section 3.3). A model of information for sensitive model components will be created, and we will show how learning from data can be integrated with expert judgment about evidential variables (section 3.4). The linking of information and preference models will be explained (section 3.5), and the combined model will be used in a test consultation to make a recommendation for a novel situation (section 3.6). Sections 4 and 5 summarize the results and discuss issues for further research.

[2] In this paper, we will not attempt to give a detailed account of how techniques from personal construct theory are applied to knowledge acquisition. Many of these are described in Boose & Bradshaw (1987a), Shaw & Gaines (1987), and Shaw, Bradshaw, Boose, & Gaines (1988).

2 BACKGROUND AND APPROACH

2.1 DECISION ANALYSIS

Decision analysis is an effort to apply decision theoretic concepts in a practical way to approach decisions that involve large amounts of uncertainty and complex preference tradeoffs (Howard, 1966a; Howard and Matheson, 1984; Keeney & Raiffa, 1976; Raiffa, 1968; Von Winterfeldt & Edwards, 1986). *Decision theory* is rooted in the axioms of probability theory (de Finetti, 1937; Savage, 1954, 1972) and utility theory (von Neumann & Morgenstern, 1953). Probability theory defines standards for the assignment of beliefs and reasoning under uncertainty, while utility theory defines how the value of outcomes may be assigned and used to select an optimal alternative when these beliefs are used in making decisions.

The *normative* basis of decision theory can be contrasted with *descriptive* approaches. While the disciplines of behavioral decision research, cognitive science, and mainstream expert system development are primarily concerned with descriptions of how people actually go about making decisions, the axioms of decision theory aim to provide principles for rational choice which, if adhered to, guarantee consistency among beliefs and preferences. Decision analysis practitioners do not deny the importance of research in human decision making, but insist that research results be applied not merely to reproduce what people do, but to discover ways to counteract systematic sources of bias in intuitive judgment (Beyth-Marom & Dekel, 1985; Kahneman, Slovic, & Tversky, 1982).

Unfortunately, the expense and scarcity of decision analysis expertise as well as the time it typically takes to build and test formal decision models have confined its application to decisions involving large amounts of resources. Recent efforts to remedy this problem by combining traditional decision analysis tools with knowledge-based systems have met with success in specific domains such as R&D project evaluation (Bradshaw & Holtzman, 1987) and infertility treatment decisions (Holtzman, 1989; Holtzman & Breese, 1986). In this paper, we address a complementary, yet somewhat different goal: rather than applying knowledge-based system techniques to automate decision analysis, we are attempting to put some of the concepts of decision analysis to practical use in the automation of knowledge acquisition (Moore & Agogino, 1987). The difference in perspective arises from our belief that the methodology of decision analysis has potentially broader application to AI problems than is generally recognized. Decisions underlie all actions of the problem solver, including:

- the selection of proper formalisms,
- the creation of efficient and comprehensible representations,
- the formulation of relevant model variables,
- the choice of appropriate solution strategies given constraints on computational resources, and
- the consideration of alternatives for model interpretation and explanation.

2.2 ELEMENTS OF A DECISION

A complete decision model, containing relevant items of problem-solving knowledge and their interrelationships, constitutes the *decision basis* (Howard and Matheson, 1984). Three things are represented in the decision basis: *information*, *preferences*, and *alternatives*. In the MYCIN example, the information consists of the knowledge a physician possesses relating symptoms and diseases; the preferences consist of factors that determine the desirability of a treatment alternative, such as cost, effectiveness, or risk; and the alternatives consist of the various possibilities for treatment. These three types of knowledge and some important subtypes are shown in Figure 4.

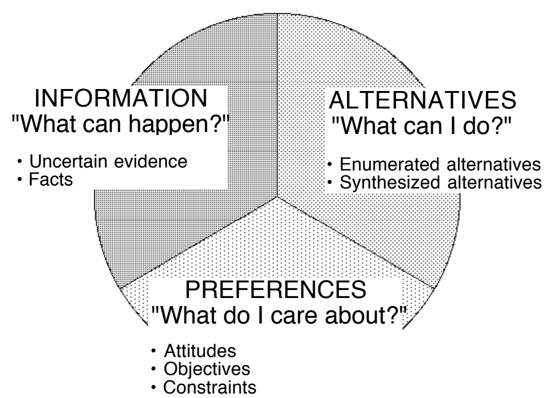

Figure 4: The decision basis is comprised of information, alternatives, preferences, and their interrelationships.

2.2.1 Information

One important part of decision knowledge is *information* about actual or possible circumstances in the world that are relevant to the decision. For example, before deciding whether to go on a picnic it would be useful to know the likelihood of rain. Finding out that rain is predicted may change our decision.

It is useful to think of information as being of two types: *uncertain evidence,* which consists of statements believed to be true with some probability,[3] and *facts,* which are statements believed with certainty. Facts, of course, are simply a special case of uncertain evidence in which the mapping onto the space of probability distributions happens to be concentrated on a single point. However, by making this distinction, we can often formulate a problem with additional clarity and computational efficiency.

2.2.2 Preferences

Preferences describe the multiple, often competing goals that are valued as outcomes of a decision. The strength of our preferences motivates us to make a careful choice among alternatives. If we really didn't prefer sunshine to rain, or going on a picnic to staying at home, any effort we put into making a decision would be wasted.

In structuring preferences, it is often useful to distinguish between *direct* and *indirect* preferences. Direct preferences relate to things we value for their own sake, while indirect preferences have no intrinsic value except as they relate to direct ones. For example, when purchasing a car, most people place a direct value on cost. However, things like fuel economy are usually only indirectly valued because of their contribution to overall cost.

We can talk about preferences as being of three different types: *attitudes, objectives,* and *constraints*.

1. *Attitudes* consist of things like *time preference* (e.g., the desire to receive good outcomes sooner rather than later) and *risk attitude* (e.g., the desire to pursue a possibly less profitable policy in order to avoid risk).

2. *Objectives* are related to the significant positive and negative consequences of the alternatives that a decision maker wishes to maximize or minimize. In decisions with multiple objectives, methods of quantification and joint measurement (i.e., commensuration) must be found so that tradeoffs can be made between them. For example, medical decision making considers not only a favorable outcome of major treatment, but also factors such as discomfort, effect on lifestyle, and length of life.

3. *Constraints* specify the conditions within which the objectives will be maximized; they define the limits of the space of acceptable outcomes. Constraints may be matters of definition (e.g., "There is $3 million budgeted for the project this year.") or of principle (e.g., a moral belief that precludes consideration of alternatives such as corporate spying to speed project development).

Many hard constraints can be relaxed. For example, if it is determined that a contemplated project cannot be finished within its original time and budget constraints, a decision maker may decide to increase the budget and hire additional personnel to meet the deadline.

Knowledge-based system researchers have used the term "preferences" to represent several different things. For example, in MOLE (Eshelman, Ehret, McDermott & Tan, 1987) experts enter "preferences" in order to rank hypotheses that might explain a particular symptom under different evidential conditions. The ranking is based on a subjective notion of evidence strength, and differs from the "preferences as desirability" concept.[4] The term has also been used in discussing tradeoffs that a particular *system* makes in reasoning under constraints of computing resources.[5] In this paper, the term "preferences" is reserved to refer to the valuations of outcomes for alternatives that individuals specify as part of building a model.

2.2.3 Alternatives

Alternatives are the various courses of action that may be recommended consistent with the decision basis. After hearing the weather report (information) and determining how important the weather will be to the success of the picnic (preferences), we will choose between going out and staying home.

There is a traditional distinction in the literature between *analysis* and *synthesis* problems. Analysis problems are those in which the alternatives can be conveniently enumerated (e.g., classification,

---

[3] This Bayesian or *subjectivist* view of probability as degree of belief is a generalization of the traditional view of probability as the frequency of a "repeatable" event. Subjective probabilities and classical probabilities share a common axiomatic foundation.

[4] This fulfills essentially the same function that numerical strength-of-evidence measures provide in *Aquinas*.

[5] In *Aquinas,* such considerations can be taken into account as part of a cost/benefit function associated with evidential variables.

diagnosis, prediction), while synthesis problems are those where the main task is *constructing* feasible alternatives in a manner that is consistent with hard constraints and optimal with respect to objectives ("soft constraints") (e.g., design, planning, configuration, scheduling). As Clancey (1984) observes, real-world problems do not always fall neatly into one of these two categories:

> "For example, if it were practical to enumerate all of the computer configurations R1 [an expert system that configures VAX computers] might select, or if the solutions were restricted to a predetermined set of designs, the program could be reconfigured to solve its problem by classification.
>
> Furthermore, as illustrated by ABEL [an expert system for medical diagnosis], it is incorrect to say that medical diagnosis is a 'classification' problem. Only routine medical diagnosis problems can be solved by classification… When there are multiple, interacting diseases, there are too many possible combinations for the problem solver to have considered them all before."

Although the focus of this paper is on analytic tasks, we note that most synthetic problems have significant analytic components. We expect that future knowledge-based systems research will yield practical approaches that integrate analytic and synthetic problem solving methodologies (Bradshaw & Boose, 1988).

2.2.4 Different requirements for different problems

Many decisions do not require the level of rigor typical of decision analysis. Sometimes the cost of such analysis is not justified by the small size of the problem[6]; sometimes the problem seems so simple that the best alternative is obvious and we can act immediately (Cyert & March, 1963; March, 1978). For personal decisions involving ethical issues, such as whether or not to consider an abortion, the methodology itself may be largely inappropriate (Levi, 1986).

Even among complex decisions, analysis requirements vary: different problems demand different amounts of emphasis on modeling the three kinds of problem-solving knowledge identified above. For example, the types of scheduling and configuration problems addressed by SALT (Marcus, 1987) are modeled in the form of constraints, objectives, and procedures for obtaining them. In such problems, uncertainty has not been a major issue. On the other hand, in a purely diagnostic or interpretive task, where no recommendations for action would be made, there is great value in modeling uncertain information but no need to model preferences. Complex decision problems require both information and preference modeling for the system to facilitate insight and effectively formulate recommendations. Figure 5 displays a set of problems as a function of the amount of uncertainty and the complexity of preference issues.

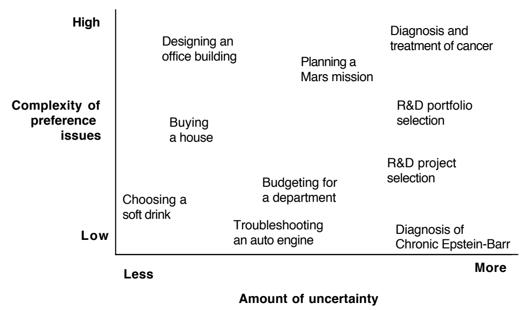

Figure 5: Approximate locations of problems as a function of amount of uncertainty and complexity of preference issues. Particular instances of these problems may prove more or less complex that suggested by the figure.

Reducing the cost of applying decision analysis techniques through automation increases the range of problems to which they can be profitably applied. However the complete automation of decision analysis techniques for use by minimally trained decision makers is probably best suited for problems involving moderate levels of resources: small problems are too trivial and very large problems too weighty to be trusted entirely to such a system. In the latter case, the high cost of professional assistance (supported by semi-automated tools) can be justified.

The amount of structure in a domain can also determine the applicability of automated decision analysis techniques (Winograd & Flores, 1987). Tasks that are highly structured and repetitive can better be hard-coded in a standard procedural program than in a decision analytic representation; on the other hand, some tasks may be so unstructured that effective computer-based procedures cannot be articulated. The ideal domain has some degree of similarity between problems, but no so much that the

6 This is true for the example problem discussed in section three. The extensive treatment we give the problem is for didactic purposes.

relevant procedures can be fully specified algorithmically.

2.3 PERSONAL CONSTRUCT THEORY AND KNOWLEDGE ACQUISITION

Researchers in knowledge acquisition have been concerned with the development, application, and validation of knowledge-based systems (Addis, Gaines, & Boose, 1988; Boose & Gaines, 1988a, 1988b; Gaines & Boose, 1988a, 1988b). They share with decision analysts a concern about how mutually acceptable formal models may be created, refined, and evaluated as a result of negotiation between the domain expert and the knowledge engineer, and efforts to solve the problem (Gaines, 1987b).

While decision theory provides a consistent standard for the assignment and use of beliefs and values, it gives little guidance for many important aspects of model structuring and representation. Current research emphasizes that the major problem in decision making is model *formulation*, not model *evaluation* (e.g., Holtzman, 1989; Wellman & Heckerman, 1987). As a consequence of incomplete theory and a limited repertoire of practical approaches to the dynamics of the modeling process, knowledge engineers have had to rely on intuition and experience as the primary means of developing and testing effective procedures[7].

One perspective that provides both a plausible theoretical foundation and an effective practical approach to the dynamics of modeling is personal construct theory (Kelly, 1955; Gaines and Shaw, 1981a)[8]. A growing number of automated knowledge acquisition tools incorporate methods derived from personal construct theory, including the Expertise Transfer System (ETS) (Boose, 1984, 1985, 1986), Planet (Gaines and Shaw, 1986a), *Aquinas* (Boose and Bradshaw, 1987a, 1987b; Kitto and Boose, 1987; Shema & Boose, 1987), FMS Aid (Garg-Janardan and Salvendy, 1987), Kitten (Shaw and Gaines, 1987), Kriton (Diederich, Ruhmann, and May, 1987; Diederich, Linster, Ruhmann, and Uthmann, 1987), KSS0 (Gaines, 1987a, 1987b), and others. We will discuss some of the ideas that have made this approach so attractive to researchers in knowledge acquisition.

2.3.1 Distinctions as the foundation of the modeling process

Personal construct theory originated in the 1950's from the research of George Kelly, a clinical psychologist who emphasized the foundational role of distinctions (*constructs*) underlying the processes of perception and reasoning:

> "In its minimum context a construct is a way in which at least two elements are similar and contrast with a third" (Kelly, 1955).

Although concepts are typically defined in terms of a property attributable to at least *two* objects, there are important reasons for regarding groups of at least *three* elements as the basis from which psychologically significant distinctions are created. While it is true that we can perceive a difference between *two* entities, that difference is confounded with simple identity. It takes three entities to create an "abstract difference" (i.e., a construct). As an example, imagine a two people, Jim and Susan, living on a small desert island where they have never encountered another human soul. Jim and Susan notice their differences but have no basis for abstracting the concepts of "male" and "female" except as they are concretely manifested in "Jim-ness" and "Susan-ness."

Continuing this line of thought, we note that a concept defined solely in terms of similarity among entities with respect to some property is of little value unless the person using that concept is immediately concerned with at least one other object in which he perceives the negation of the property. In deciding among investment alternatives, the consideration of investment risk would be pointless if all choices are seen as equally safe. However, if a new alternative with greater or less risk were introduced, the construct of risk may suddenly take on importance.

A construct, then, consists of a simultaneous abstraction of similarity and difference that arises in

---

[7] Evans (1988) reviews the contributions of research in cognitive psychology to the knowledge acquisition problem.

[8] For a number of years, Kelly's ideas were largely neglected in mainstream psychology. There were many reasons for this, including his early death, his private personality, his irreverent writing style, and his emphasis on idiosyncratic minds as opposed to nomothetic study (Allport, 1942). Accounts of the early lack of reception to personal construct theory and its renewed vitality and relevance may be found in Davisson (1978) and Jankowicz (1987). A summary of Kelly's view of decision making can be found in Currie (1985).

the context of its use[9]. The creation of a construct by a person is inherently a differentiating and integrating act[10]; it is both an affirmation and a denial of some state of affairs: "Sometimes what is implicitly denied is more focal to the person's intent than what was affirmed, as when the disgruntled first mate entered in the ship's log that 'the captain was sober tonight'" (Kelly, 1966)[11].

Kelly's formulation avoids the extremes of both solipsism and realism with respect to the origins of distinctions, and the unproductive debate between subjectivists and objectivists in the realm of probability. Both the substance on which perception feeds and the active, constructive nature of mind are taken into account: "Ideas are not in the mind, nor objects in the world, but… both are in the meeting of mind and matter" (Shaw & McIntyre, 1974).

To *construe* is both to abstract from past events and to provide a reference axis for conceptualizing future events based on that abstraction. Furthermore, the act of valuation is implicit in the creation of constructs — distinctions simply do not arise unless there is a perception that they affect something we care about[12]. The process of construal thus lays the ground for all subsequent mathematical reasoning, both for information and preferences:

> "The statistics of probability are based upon the concept of replicated events. And, of course, they are also contrived to measure the predictability of further replications of the events. The two factors from which predictions are made are the number of replications already observed and the amount of similarity which can be abstracted among the replications. The latter factor involves some complicated logical problems — for example, representative sampling — and, in practice, it is the one which usually makes predictions go awry. Since the abstractive judgment of what it is that has been replicated is the basis for measuring the amount of similarity, we find that the concept-formation task which precedes the statistical manipulation is basic to any conclusions one reaches by mathematical logic" (Kelly, 1955).

Though few would disagree with Kelly's observation, in practice designers of knowledge acquisition tools have given relatively little attention to supporting the preliminary conceptual aspects of modeling that Kelly identifies as so crucial.

2.3.2 Personal modeling

Kelly rejected the essentially passive accounts of human behavior dominant in his time. His theory provides a rich characterization of the efforts of individuals to actively anticipate and control their environment. He draws parallels between the processes that guide scientific research and those involved in every day activities:

> "There is no difference in kind between the scientist inferring the most esoteric theory of reality, on the one hand, and the simplest organism's inferring the presence of food or danger in its environment. In both cases the fundamental *activity* of the nervous system is *classification* (or abstraction) and the fundamental *function*

---

[9] "To the aesthetic eye, the form of a crab with one claw bigger than the other is not simply asymmetrical. It first proposes a rule of symmetry and then subtly denies the rule by proposing a more complex combination of rules" (Bateson, 1972).

[10] "By such an act he interposes a difference between incidents — incidents that would otherwise be imperceptible to him because they are infinitely homogeneous. But also, by such an… act, he ascribes integrity to incidents that are otherwise imperceptible because they are infinitesimally fragmented" (Kelly, 1961). This implies that to learn at all, "we must forego *some* potential discrimination of states, some possibility of learning everything" — otherwise there could be no generalizations:

> "The popular image of science envisions the scientist making the maximally precise measurements as a basis for his theories, but, in practice, scientists are lucky that measurements are not overly precise. Newton based his Law of Universal Gravitation on the elliptical orbits of Kepler, but Kepler abstracted these ellipses from the observations of Tycho Brahe. Had those observations been more precise (as precise as we now can make) the orbits would not have been seen as ellipses, and Newton's work would have been much more difficult" (Weinberg, 1975).

[11] A discussion of constructs from a philosophical point of view is given by Husain (1983).

[12] To put this another way, we may say that no learning takes place without an affective component (Piaget, 1981; Sigel & Holmgren, 1983).

of the nervous system is *modeling* (of the environment).” (Weimer, 1975)

In Kelly's view, humans model their environment and scientists model humans through the same process of *simulation*:

> “I think truth can be approached by simulation and by simulation only… Man gets at the truth of things… by erecting constructs to simulate it the best he can… [And scientists] devise machines to simulate — not man directly — but theories about man… the theories, in turn, are constructed to simulate the human processes they are supposed to explain. But the simulation does not stop there. The persons themselves are simulators. They attempt to simulate each other — too much, some say. They simulate their parents, their gods, a presumed rational way of life, and the expectations of others. In fact, a lot of people even make a big to-do about simulating themselves. This is known as ‘trying to be yourself’ and is often regarded as quite an accomplishment. Sometimes people simulate machines. This is sometimes called ‘being objective.’ [One scientist] has even programmed his people to behave like computers. Some psychologists undoubtedly will take this to mean that he has succeeded in getting people to behave psychologically.” (Kelly, 1963)

This simulation through models is not restricted to the symbolic activity of “conscious” thought. If we define symbolism in a very general way as the ability of processes to imitate each other, then it becomes apparent that virtually all neural activity is symbolic, though it may take place without words and conscious awareness:

> “The activity of neural excitation per se is totally unlike, say, the patterns of stress in a bridge, yet the *patterns* of excitation which constitute the thinking about (or calculating) that stress are isomorphic (in a structural sense) to the stress itself” (Weimer, 1975; see also Craik, 1943)[13].

Kelly's perspective on “persons-as-scientists” resulted in a unique theory of how and why personal models of the world are created and maintained. In this view, a major goal of individuals and social systems is anticipation of the future. Personal models of the world are developed in order to improve the “accuracy” of our anticipation of aspects of the future that are important to us, thus making the effects of our actions more predictable. Action is a form of active anticipation that seeks to make desirable outcomes more likely.

2.3.3 Surprise and information

Our personal models are never finished: they are constantly being revised as a result of the inadequacy of our knowledge to correctly anticipate all events and of our actions to completely satisfy our preferences. We experience our ignorance prospectively as *uncertainty* and retrospectively as *surprise* (Figure 6). The more unpredictable an event, the more likely it is that the outcome will be surprising. To the extent that an event is predictable, we will, on the average, expect to be less surprised by its outcome.

Our feelings of prospective uncertainty and retrospective surprise have been formulated quantitatively in three related concepts: *probability*, *information*, and *surprise*. These concepts can be most easily understood in terms of a fourth concept, *entropy*. Entropy is a term in *information theory* that has been borrowed from thermodynamics (Shannon & Weaver, 1964)[14]. It can be defined as the

---

[13] Of course, no model is ever fully isomorphic to the system being described. Giere (1988) speaks of the relationship between models and “real systems” as one of similarity:

> “But since anything is similar to anything else in some respects and to some degree, claims of similarity are vacuous without at least an implicit of relevant *respects* and *degrees*… To claim a hypothesis is true is to claim no more or less than that an indicated type and degree of similarity exists between a model and a real system. We can therefore forget about truth and focus on the details of the similarity.”

[14] In a physical system, entropy can be interpreted as uncertainty about which quantum state a system is in — a concept that is analogous to uncertainty about beliefs. The equations for calculating thermodynamic entropy and entropy as a measure of average information quantity are also similar. The central problem in information theory is the same one that is faced by probability theory, namely “taking information and encoding it in such a way that it enables us to act on the information without presuming more than we know, or failing to use all that we know” (Tribus, in Campbell, 1982). Campbell (1982) and Bharath

uncertainty of an event, the average amount of information yielded by an event, or the average surprise evoked by that event.

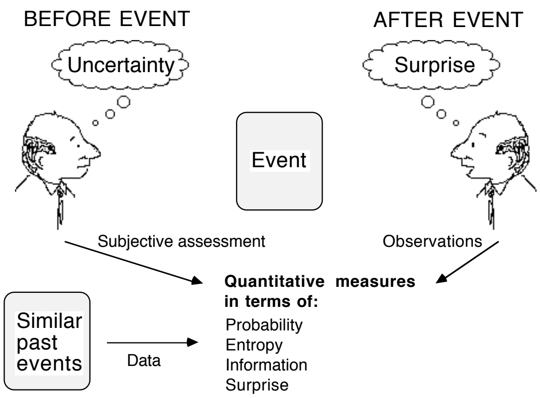

Figure 6: Relationship between uncertainty, measures of uncertainty, and surprise.

Probability (p) and entropy (H) are related mathematically by the following:

$H(x) = -\Sigma_{i=1 \text{ to } n}\ p_i \text{ x } \log_2(p_i)$.

where n is the number of possible outcomes for the event (x) and $\log_2(p_i)$ is the logarithm of p_i at base 2. Entropy is a measure of disorder; hence information (*negentropy*) is a measure of order or of organization, since the latter, compared to distribution at random, is an improbable state. Entropy is maximized when all outcomes for an event are equally probable (e.g., for a 'fair' coin, p(head) = p(tail) = .5). The assumption of "maximum entropy" (or, "minimum prejudice") has been defended by many researchers as the preferred expression of total uncertainty or ignorance (see e.g., Cheeseman, 1985). For example, in the absence of evidence relating to the fairness of a coin, a probability for the outcome of a toss could be determined by the assumption of "maximum entropy" (p(head) = p(tail) = .5)[15]. In dealing with situations of *partial* knowledge, the maximum entropy assumption means that any assignment of probability should reflect precisely the knowledge of the person making the assignment, while being "maximally vague" with respect to uncertainty (Tribus, in Campbell, 1982).

The definition of entropy as the average amount of information yielded by an event can be understood by reference to the game of "Twenty Questions" (von Bertalanffy, 1968). In this game, a group of players privately selects some object, and an individual not party to the selection attempts to guess the identity of the object by asking twenty yes-no questions. The amount of information conveyed by an answer to the first question in the game allows us to decide between one of two alternatives (e.g., animal or vegetable). The second question, allows us to conclude one of four possibilities (e.g., animal/bigger-than-a-breadbox, animal/smaller-than-a-breadbox, vegetable/bigger-than-a-breadbox, vegetable/smaller-than-a-breadbox), and so on. The logarithm at base 2 is a natural way to measure the information conveyed by the answer to each question, with a *bit* as our unit. The information in one answer is $\log_2(2) = 1$ bit, for two answers it is $\log_2(4) = 2$ bits, for three answers it is $\log_2(8) = 3$ bits, etc.[16]

Useful distinctions "create" information (or in other words, they reduce entropy) in that they provide a

(1987) provide readable introductions to the basics of information theory.

15 It could be argued that no one is ever in a state of total ignorance with respect to a situation: to make a better guess about the outcome of the coin toss, we can always infer something about the fairness of the coin based on our experience in similar situations, our knowledge of the person throwing the coin, and other contextual information. Whenever a person possesses information that suggests a more specific criterion than maximum entropy for probability assignment, the more specific criterion should be used.

The problem of what to do in these situations is most difficult in automated reasoning systems. Such systems are currently much less capable than people in devising effective fallback strategies for indeterminate situations.

16 Actually, the amount of information yielded by the answer is effectively *more* than 1 bit. At any point in time, there are several active predictive models engaged by the person that will benefit from the answer. For example, the answer not only tells the person whether the object is an animal or a vegetable for this unique situation, but also provides confirming or disconfirming evidence for hypotheses about likelihoods of one or another kind of answer for this group in general, for the game in general, etc. Furthermore, in human interaction there are several channels of communication open at any point in time: the tone of the response, the sitting position of the respondent, etc. convey information about his mood, his attitude toward the game, the questioner, and so forth. Thus, the information yielded by the answer to a single question is potentially infinite — as another example consider the field of archaeology, where whole patterns of extinct civilizations can be deduced by a skilled observer from a single piece of stone (Weinberg, 1975).

way of grouping events into categories of unequal probability. If I create a new distinction between "sunny" and "rainy" days, and observe that 60% of the time it is rainy, this reduces my uncertainty about the weather tomorrow. Before, it was at maximum entropy (50-50 chance of rain), whereas now I have reason to believe that the chance of rain will be somewhat higher. Some additional distinctions between days (e.g., winter days, summer days) may improve my confidence in making predictions and reduce the potential for surprise, while other distinctions I could define may not (e.g., even-numbered days, odd-numbered days). Useful distinctions generally provide clues to the constraints (whether physical or mental) that determine that probabilities between categories of events *will* be unequal.

Events are 'informative' only in relation to some context. A message about a probable event ("The sun rose this morning.") contains less information than a message about a less probable event ("There was an earthquake this morning."). A message contains information only to the extent it resolves uncertainty. Thus the concept of *surprise* is central to learning: events that match our expectations convey no information; learning takes place when we modify our personal models in response to 'surprises.'

2.3.4 Emotion and preferences

Learning is dependent on emotional processes associated with surprise (Shaw & Gaines, 1986). Melges (1982), for example, has argued that a major function of emotions is to:

> "attune the person to overall discrepancies between the present and the future so that he adjusts his plans of action to his future images."[17]

The experience of surprise is the primary indicator to the individual of discrepancies between the predictions of the model and incoming information. The type of emotions we experience varies in response to different surprising events: negative emotions arise when we sense a threat to aspects of the model we deem important, and positive emotions arise if events relate to aspects of the model we prefer to change (McCoy, 1981)[18].

The *intensity* of emotion is due not only to the amount of surprise we experience, but also to the relevance of the surprising event to things that we value. If losing all one's assets in the stock market and having whale steak for supper have the same probability (i.e., would be equally *surprising* in an information theoretic sense), it does *not* mean that the occurrence of either event will evoke the same degree of emotion. Conversely, a hole in a dyke will elicit a great deal of intrinsic interest regardless of its size (Ferguson, 1976). Thus the crucial measure of the effect of new information on decision making is the *value of information,* which jointly addresses the probabilistic structure of new information *and* the strength of its impact on the likelihood of preferred outcomes (Howard, 1966b)[19]. *Expected value of information* can be used as a criterion in the model building process to determine if there are aspects of the model deserving further refinement. An example of the use of expected value of information to guide model refinement is given in section 3.3 below.

2.3.5 An account of model formation

Gaines (1977, 1987a, 1987b) has extended personal construct theory to clarify how the processes we have described relate to an abstract account of model formation. He has proposed a recursive hierarchical organization of distinctions about distinctions each level of which contains predictive models of the world[20] (Figure 7):

---

[17] Pribram (1971) has shown that habituation does not indicate a loss of sensitivity by the nervous system; rather, the brain develops a neural model of the environment, a representation, an expectancy against which inputs are constantly matched. But this is only half the story: "[T]o this picture must be added the startling rider that the nervous system *creates* its own inputs":

> "Our expectations of the future are determined by our constructive memories: hence our theory of memory is *ipso facto* our theory of 'expectation' which in turn is *ipso facto* what guides our concept formation, which in turn is what guides scientific life" (Weimer, 1975).

[18] Of course, emotions are more than a passive response to events in the world. (Beier, 1966; Beier & Valens, 1975; Brown & Bradshaw, 1985; Brown, Warner, & Williams, 1987; Warner, 1982) have argued that the expression of emotions and emotion inducing behavior is a powerful means of achieving desired responses from others.

[19] This idea has profound implications for conventional signal processing. Researchers at the MIT Media Lab have begun to pursue the idea that data compression should maximize *content* and *meaning* of a message, not merely the objective signals (Brand, 1987).

[20] Gaines' hierarchy is an adaptation of Klir's (1976, 1985) epistemological hierarchy of modeling systems.

"• At level one, *constructs* are those distinctions that the particular modeling system makes, a language for describing the world[21];
• At level two, *data* are descriptions of actual case histories in terms of the constructs, an account of experiences of the world;
• At level three, *hypotheses* are the means of regenerating particular case histories from generalized accounts, rationalizations of the world (often called *models*);
• At level four, *analogies* are similarities between differently generated generalized accounts, correspondences between models;
• At level five, *abstractions* are accounts of a wide range of models, underlying foundations of analogies;
• At level six, *principles* are systemic foundations for abstraction, accounts of abstractions." (Gaines, 1987b)

Figure 7 illustrates that the fundamental processes in the model are the flow of *information* as surprise about events ("news of a difference"-Bateson, 1972) upward through the hierarchy when lower levels cannot account for events, and the flow of *preferences* downward as lower-level predictive models accounting for events are created to be consistent with higher-level ones. The flow of preference can ultimately result in action as higher levels attempt to influence the anticipated future[22].

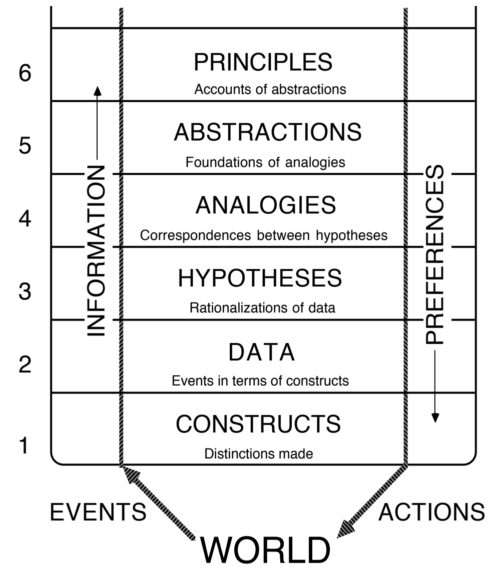

Figure 7: Gaines' (1987b) construction hierarchy of a person modeling the world.

A goal of the modeling system is to minimize the rate at which surprise (i.e., discrepancies between models and events) is passing upward through the hierarchy[23]. Priority in the resolution of discrepancies is given to events with the greatest impact on valued anticipated outcomes (i.e., those that have a high value of information). One response to surprising events is to revise the model in ways that better account for the new information. However, if model revision is perceived as too costly or too threatening to higher level distinctions, we may prefer to reject or

21 The level of constructs is primitive in the modeling hierarchy in that systems defined at this level contain no information about the relationships (constraints) between variables (Klir, 1985).

22 Gaines (1987a) cites Tucker and Williamson (1984) as possible physiological and behavioral evidence of the existence within the brain of two channels of internal communication analogous to those shown in Figure 7: "The *arousal* system passes surprise upwards to the cortex from the limbic region when unexpected events occur. The *activation* system passes preferences down from the cortex to the motor regions."

23 This is not to say that the best modeling systems are necessarily the ones that are least often surprised. Without surprise, no learning can take place. Lack of surprise, then, may be due either to an impoverished set of predictive models or a "defensiveness" to new information.

ignore the new information instead[24]. The choice of whether to revise our models or reject the information in response to surprising events is what ultimately guides the inductive inference process (Figure 8; Gaines, 1987a; Stich & Nisbett, 1984).

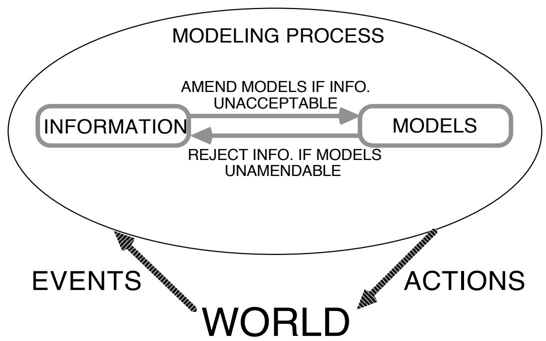

Figure 8. Dynamic equilibrium in inductive inference (adapted from Gaines, 1987a).

One can go wrong in either direction. Often we are too eager to completely substitute theory for observation, like the Viennese ladies who weigh themselves before entering Demel's Tea Room: "If they're down a kilo, they have an extra mochatorte, and if they're up a kilo they pronounce the scale 'in error' and have an extra mochatorte anyway" (Weinberg, 1975). Science works in this way — it matures by increasingly trading the "brain" for the "eye" until it becomes almost impossible to invalidate a scientific paradigm with mere empirical observations. On the other hand, we do not want to throw out a general model based on accumulated experience on the basis of a single negative case. Because models may embody a wealth of valuable information, it is usually preferable to salvage them by changing definitions or enumerating qualifying conditions and exceptions, rather than rejecting them outright[25]. An example from Kaplan (1963) illustrates the difficulty of balancing our scientific skepticism with an openness to surprising observations:

> "Describe to someone what appears to be the favorable outcome of an experiment on telepathy; he will say that the proportion of successes or the number of cases was too small to be significant. Ask him to suppose that ten times as many cases were taken, and that the results were ten times as favorable, he will suggest that trickery was involved… But at some point, what he *should* say is, '*If* the inquiry and its results *were* as you hypothetically describe them to be, I would believe in telepathy!'"

2.3.6 Criteria for model selection and revision

Given that we are unsatisfied with our previous predictive model, we are confronted by the fact that there are an infinite number of ways to change it consistent with the new data. A generally accepted criterion for model selection or revision, given equal predictive validity, is that of simplicity (or parsimony). However, Gaines (1987a) has pointed out that "simplicity/complexity ordering is arbitrary and in its most general form is just one of *preference*."

Bateson (1979) makes the same point about the arbitrary nature of preference criteria in his discussion of the problem of selecting hypotheses

---

[24] Hence Ferguson's (1976) observation that "an ounce of example is worth a pound of precept, until one learns the principle, and then it is the other way around." Taken to an extreme, the distortion of data to fit one's hypotheses is termed *hostility* (Kelly, 1957; cf. Piaget *assimilation*); its opposite, "the active elaboration of one's perceptual field" is termed *aggression* (cf. Piaget, *accommodation*). Strictly speaking, it is erroneous to say that data is acquired "objectively", then altered at a later point by our prior beliefs and intentions; rather, information is acquired in the first place in a form that reflects prior beliefs and intentions (Brown *et al.*, 1987).

[25] Distinctions near the top of the hierarchy are least susceptible to change. Kelly (1955) termed these *core constructs:*

> "Core constructs are those which govern a person's maintenance processes — that is, those by which he maintains his identity and existence."

Shaw and Gaines (1986) discuss why it is important that core constructs be relatively resistant to modification:

> "Systemically, the existence of core constructs corresponds to our not being able to reevaluate all past experience in the light of a new model. We encapsulate that experience in a model based on distinctions to which we give inertia through reluctance to change. A system that amends it core constructs becomes detached from its own history and loses what has been learned from past experience. The optimization between readiness and reluctance to change is a function of the stability of the system's environment and is part of its adaptive process."

For a related discussion, see Beach & Mitchell (1988). Core constructs appear to be related to the concept of "self-image" in their image theory of decision making.

when confronted with the problem of predicting the next in a sequence of numbers:

> "You *assume* that you can predict… But the only basis that you have is your (trained) preference for the simpler answer and your trust that my challenge meant that the sequence was incomplete and ordered.
>
> Unfortunately (or perhaps fortunately), it is so that the next fact is never available. All you have is the hope of simplicity, and the next fact may always drive you to the next level of complexity.
>
> Or let us say that for any sequence of numbers I can offer, there will always be a few ways of describing that sequence which will be simple, but there will be an *infinite* number of alternative ways not limited by the criterion of simplicity."

Thus, models cannot be fully accounted for in terms of the data and the data is never fully reflected in the model: models are constructions providing a method for regenerating idealized data in a way that satisfies some set of (arbitrary) preference criteria (Giere, 1988).

The top-down application of preferences as constraints in model selection assures that the search for an appropriate new model is efficient and consistent with the most general principles that guide our action. However, the "framing" effects of this selectiveness effectively blinds us to certain categories of hypotheses. Rubenstein (1975) reports the story of a mathematics professor who was asked by his students to give a simple solution to predict the next member in the sequence 32, 38, 44, 48, 56, 60. After much unfruitful effort, the professor gave up. The answer was "Meadowlark", the stop that followed 60th street on the subway the professor took every day to work[26].

It is our view that the interplay between inductive processes based on experience and deductive processes based on preferences constitutes the essence of the modeling process. We will now present some examples illustrating how new distinctions are formed within hierarchical modeling systems.

2.3.7 How new distinctions are created

There are an infinite number of *potential* distinctions in the world, but very few of them become *effective* differences (i.e., "differences that make a difference"[27]) for an individual (Bateson, 1979). Weinberg (1975) has pointed out that learning a card game involves recognizing features of the cards for that game and ignoring the rest. For example, suits are unimportant in War and Old Maid, whereas in bridge it is important to distinguish between *high* cards and cards below 10.

There can be no universal set of effective distinctions. *All* distinctions are potentially significant in some context. It is not hard to imagine circumstances in which even the most seemingly trivial distinction might become crucially important. The significance of a feature depends on the particular "game" we are playing. Consider the following set of distinctions purportedly taken from a certain Chinese encyclopedia:

> "Animals are divided into
> (a) those that belong to the emperor,
> (b) embalmed ones,
> (c) those that are trained,
> (d) suckling pigs,
> (e) mermaids,
> (f) fabulous ones,
> (g) stray dogs,
> (h) those that are included in this classification,
> (i) those that tremble as if they were mad,
> (j) innumerable ones,

---

26 Texts and courses claiming to enhance creativity or problem-solving skills often exploit the power of getting people to explore alternate representations and frames-of-reference when dealing with difficult problems.

27 Bateson and Bateson (1987) later refined this succinct definition of information to be a "difference which makes a difference *at a distance"* in order to emphasize the distinction between the world of the hard sciences and the world of communication and organization: "When one billiard ball strikes another, there is an energy transfer such that the motion of the second ball is energized by the impact of the first. In communicational systems, on the other hand, the energy of the response is usually provided by the respondent. If I kick a dog, his immediate sequential behavior is energized by his metabolism, not by my kick… The dog's metabolism might in the end limit his response, but… the energy supplies are large compared with the demands made upon them; and, long before the supplies are exhausted, 'economic' limitations are imposed by the finite number of available alternatives, i.e., there is an economics of probability" (Bateson, 1972).

(k) those drawn with a very fine camel's hair brush,
(l) others,
(m) those that have just broken a flower vase,
(n) those that resemble flies from a distance"
(in Borges, The Analytical Language of John Wilkins, quoted in Rucker, 1984).

We will see in section three below how the distinction *(n)* will play an important role with respect to a particular decision.

It is the present or anticipated need to take action that forces us to reorient our perspective and create new distinctions (i.e., make *potential* distinctions into *effective* ones). Potential distinctions remain imperceptible until our movement relative to them creates an *event:*

> "What happens is that a static, unchanging state of affairs, existing, supposedly, in the outside universe quite regardless of whether we sense it or not, becomes the cause of an event, a step function, a sharp change in the state of the relationship [between the individual and the entity *in time*]…
>
> The best example I can think of is the case of an automobile traveling over a bump in the road. This instance comes close, at least, to meeting our verbal definition of what happens in processes of perception by mind. External to the automobile there are two components of a difference: the level of the road and the level of the top of the bump. The car approaches these with its own energy of motion and jumps into the air under impact of the difference, using its own energy for this response…" (Bateson, 1979)[28].

As we observed previously, the significance of the information yielded by an event is determined by two things: the surprise evoked by its occurrence and its relevance to the likelihood of important anticipated outcomes. Most bumps in the road go unnoticed. However, to a group of persons riding a roller coaster, the event of a bump causing the cars to jump the track would be highly significant!

Having described the the nature of events, we can see how their occurrence can provide a context for the emergence of new distinctions. In their analysis of how language arises, Winograd and Flores (1987) give the example of a writer whose focus of attention is on the words as they appear on the screen of a word processor. A complex array of equipment mediates between the writer and the screen, however only the typing is part of his world; the components of the equipment are "invisible" to him except as they reveal themselves when there is a problem. If a letter fails to appear on the screen, the writer becomes aware of a keyboard with properties such as 'stuck keys.' Or if there is no apparent mechanical problem, he may discover that there is a part of the program called the 'keyboard handler' that can be blamed for certain kinds of malfunctions called 'bugs.' The important point is that these new distinctions do not arise spontaneously, but as part of the writer's efforts to cope with the event of a breakdown. As another example:

> "It is often remarked that Eskimos have a large number of distinctions for forms of snow. This is not just because they see a lot of snow (we see many things we don't bother talking about), but precisely because there are recurrent activities with spaces of potential breakdown for which the distinctions are relevant" (Winograd & Flores, 1987).

In summary, we have seen that in response to a surprising event, new distinctions may emerge, creating possibilities for action: "Our knowledge of the phenomenal world raises problems which can be answered only by altering the picture which our

---

[28] *Change* is the word we use to describe differences that occur across time. New constructs may arise not only from the perception of similarity and contrast among three entities but also from observations of a single entity viewed from at least three points in time. The invariance of the entity at two points in time and a difference at a third point provide the minimal criteria for the creation of a distinction. More generally, we say that constructs and elements are understood by performing transformations on them (e.g., of identity, over time, by combining with other elements) and noticing which properties are preserved and which are not. Invariance under particular kinds of transformations reveals the constraints that govern that invariance: "We understand change only by observing what remains invariant, and permanence only by what is transformed… Science may be thought of as the process of learning which ways of looking at things yield invariant laws. The laws of science may thus be *descriptions* of how the world looks ('Eureka' — I have found), or *prescriptions* for how to look at the world ('heuristic' — how to find). We really have no way of knowing which" (Weinberg, 1975).

senses give us of that world" (Hayek, 1952). The incorporation of these new distinctions into a model may lead to better anticipation of future events. Better anticipation in turn allows choices to be made that are more likely to satisfy our preferences.

Personal construct theory provides a rich account of how personal models are created and maintained. Furthermore, although it is fundamentally a theory about *persons*, it provides a foundation for an approach to the *automation* of the model building process.

2.4 MODELING INFORMATION AND PREFERENCES IN *AQUINAS*

2.4.1 Description of *Aquinas*

Aquinas grows out of work on the Expertise Transfer System (ETS), an automated knowledge acquisition tool that has been in use in Boeing for more than four years (Boose, 1986). Many prototypical knowledge-based systems have been generated by ETS in order to explore project approaches or feasibility. Several of these prototypes have grown into knowledge-based systems used at Boeing for diagnosis, structured selection, and group decision-making problems.

ETS uses techniques from personal construct theory (Kelly, 1955) to interview experts and uncover key aspects of their problem-solving knowledge. It helps build very rapid prototypes (typically in less than two hours), assists the expert in analyzing the adequacy of the knowledge for solving the problem, and can create knowledge bases for several expert system shells (S.1, M.1, OPS5, KEE, and so on) from its own internal representation.

The tools in ETS have become part of *Aquinas*, a much larger system that addresses previous limitations of ETS in representation and reasoning when acquiring knowledge for complex problems (Boose & Bradshaw, 1987a, 1987b). *Aquinas* is organized around a collection of integrated tool sets that share a common user interface, underlying knowledge representation, and database.

Aquinas is named for the twelfth century theologian and scholar, St. Thomas Aquinas:

> "Aquinas christianized Aristotle, thus reconciling faith and reason… [Before St. Thomas,] when the text of an ancient author was studied, the commentator or copyist, when he came upon something that clashed with revealed religion, either scratched out the 'erroneous' sentences, or else shifted the words to the margin. What did Thomas do instead? He aligned the divergent opinions, clarified the meaning of each, questioned everything, even the revealed datum, enumerated the possible objections, and essayed a final mediation." (Eco, 1986)

It is in this spirit that we have tried to approach the task of building a knowledge acquisition tool; our objective is to create a dialogue between the intuition and experience of the expert and the logic of the methodology that will facilitate convergence on a mutually acceptable model.

Aquinas is written in Interlisp and runs on the Xerox family of Lisp machines. Subsets of *Aquinas* also run in Interlisp on the DEC VAX and in a portable version written in C. The *Aquinas* screen is divided into a typescript window, map windows showing hierarchies, repertory grid windows, and analysis windows (Figure 9). Experts interact with *Aquinas* by text entry or by mouse through pop-up menus.

Aquinas continues to evolve as we receive feedback from individuals who have encountered difficulties in representation or reasoning for their particular applications. The approach described in this paper was devised as a first step toward resolving limitations discovered in trying to combine complex evidential reasoning with reasoning about preferences. Once this initial approach has been fully implemented, we plan to extend and generalize it to apply to a wider range of problems.

Most importantly, we plan to integrate features of *Aquinas* with those of *Axotl*, a second automated tool under development at Boeing Computer Services that uses a knowledge-based system approach to automate the formulation, analysis, and appraisal of decision analysis models (Bradshaw & Holtzman, 1987). *Axotl* is written in Smalltalk-80 and runs on the Apple Macintosh II and other platforms supporting Smalltalk (Sun, IBM, HP, Apollo).

2.4.2 Knowledge representation

Elements (alternatives or outcomes) and *constructs* (dimensions of similarity and difference between elements) are central to knowledge representation in *Aquinas*.

Constructs are defined through being applied to a series of elements. One way of representing constructs and elements is by constructing a *repertory grid* (Figure 9). A repertory grid is a matrix with elements ranged along the top and a construct, defined by extension, as a horizontal row of values within the matrix. Each construct is represented as a somewhat different pattern of values in the row. If two constructs were to match, cell by cell, throughout the length of their rows to infinity, they would be identical. When there is a high degree of match between two constructs, we say that they are 'functionally similar' within a specific context.

Most constructs are made up of a pair of opposites. For example, a construct named *temperature* may have *hot* and *cold* as labels for the two ends of the underlying conceptual dimension it defines. Constructs such as temperature define *ordered* dimensions, where it is meaningful to think of one of the opposites as representing *less* of some quality and the other as representing *more* (e.g., cold is represented by smaller numbers on the temperature scale than hot). Kelly argued that all distinctions are bi-polar, but sometimes it is more convenient to have a single *unordered* dimension subsume several ordered ones. For example, the constructs *red/not-red*, *blue/not-blue,* and *green/not-green* could be subsumed under a single construct, *color,* that could take on values of red, blue, or green.

Constructs define the traits, aspects, attributes, or characteristics by which elements can be differentiated. When applied to problem solving, they are typically used either to model preferences or information. In *preference grids*, the elements represent alternatives for a decision. For these, the constructs define specific dimensions of preference that are used to select the best alternatives. *Information grids* contain beliefs about events, qualities, or states of the world and their relationships to other events, qualities, or states. Figure 11 is an example of a preference grid and figure 21 is an example of an information grid.

Values in the grid, called *ratings,* are assigned by individuals to represent the location of an element on a particular construct dimension. People may assign point values as ratings (e.g., cold, with a certainty of 1.0), or they may distribute their belief among more than one rating if they are uncertain (e.g., cold, with a certainty of .6; hot, with a certainty of .4). Individuals define the range of values that apply to each construct, and specify whether values are subjectively defined or represent precise categorical or numeric data[29].

[29] Currently, a trait's values may be declared as being either of *nominal, ordinal, interval,* or *ratio* type.

Consistent with Kelly's theory, constructs are usually generated by presenting elements to the person three at a time and asking about a quality which makes any two of them similar and at the same time different from a third. There are practical as well as theoretical considerations that motivate the use of triads of elements in interviewing experts. Because people are so good at listing relevant distinctions, it is tempting to let them add components to the model at will rather than using a structured interview process. While it is true that most experts can readily generate a set of terms to describe their domain, our experience is that unstructured methods typically produce terms that are less interesting[30] and less general than personal construct methods. Because of the way the triadic elicitation method frames the task as one of distinguishing among elements, the expert's attention is focused on generating a minimal set of discriminating dimensions, rather than a larger set of descriptive ones that may or may not be of practical use. With respect to generality of the terms produced, distinctions based on the presentation of two entities tend to be less robust. In Kelly's terms, they are relatively *impermeable,* that is, they are more specific to the two elements being considered and thus less likely to be applicable to new elements that may be introduced later. More *permeable* constructs are produced when experts consider at least three things at a time.

The grid is a representation of a person's unique psychological space. Many analysis tools in *Aquinas* are used to directly explore interesting relationships within such a matrix produced by an individual or between matrices created by several persons (e.g., cluster analysis, similarity analysis, difference grids). The results of analysis can be used to guide refinement of the model.

Repertory grids have many advantages as a form of knowledge representation. First, they are a very general-purpose representation. Grids may be viewed as a component of a database in entity-attribute form

[30] The "interestingness" of a distinction can only be defined with respect to a specific decision context. We distinguish between distinctions that are (in order of increasing "interestingness") *relevant* (i.e., identified as being related in some way to the decision), *pertinent* (not only relevant, but also have some effect on the *value lottery* of an alternative (see section 3.2 below)), and *material* (not only pertinent, but also sufficient to swing the decision in favor of a particular alternative).

(Chen 1980): a repertory grid has elements as *entities,* constructs as *attributes,* and allocations of elements to locations of construct dimensions as *values*. In their most simple form, they can be thought of as a kind of decision table, with probabilistic rather than deterministic relationships. The representation of knowledge in grids facilitates the creation of interfaces to databases and spreadsheets (Bradshaw & Boose, 1988).

There are other advantages as well. The organization and logic of expert knowledge in a grid can be easily inspected and analyzed. Recognition and completion of patterns in the data are facilitated by the structure and relative compactness of the matrix representation as compared to rules. Furthermore, representation in grids facilitates testing for conditions of ambiguity, redundancy, and completeness (Cragun & Steudel, 1987).

Although the repertory grid representation has become strongly associated with personal-construct-based approaches, there are other equivalent forms that could be used for convenience, efficiency, portability to other knowledge-based systems, or the preference of the the person using the system[31]. Internally, *Aquinas* represents knowledge as a network from which grids are dynamically constructed. Distinctions captured in grids can be converted to many other numerical and graphical representations, including rules (Boose, 1986; Boose, Bradshaw, & Shema, 1988; Gaines & Shaw, 1986b). However, a translation from a higher level representation such as grids or networks to a less structured representation such as rules can often compromise performance of the knowledge base. Heckerman and Horvitz (1987) and Pearl (1987) present convincing arguments against the myth of modularity in rule-based systems: specifically, they show that classes of dependencies among beliefs held with uncertainty cannot be represented in rule-based systems in a natural or efficient manner regardless of the inference scheme used[32].

Although grids are ideal for making patterns of values in the data salient, there are more convenient ways of displaying relational information to an individual, such as conceptual dependencies between model variables. For example, the relationships between elements at various levels of abstraction cannot be shown within a grid. Dependencies are most naturally displayed in networks (Pearl & Verma, 1987; Pearl, Geiger, & Verma, 1988). Heckerman has recently developed a representation called *similarity networks,* which relies in part on personal construct methods and allows the identification and graphical display of relationships indicating constraints on conditional independence relationships (Horvitz *et al*., 1988). In previous work in *Aquinas*, we have allowed individuals to have both grid and structured network views of the same information. The network views are a convenient way of representing different kinds of abstraction relationships (cases, experts, elements) and preference hierarchies (constructs).

In this paper, we extend our previous approach to allow more complex preference and evidential structures in networks and grids. Recently there has been much research demonstrating the value of influence diagrams or belief networks (Howard & Matheson, 1984; Pearl, 1986; Bradshaw & Holtzman,

---

[31] Kelly himself came to regret the narrow-mindedness of researchers who had equated Personal Construct Theory and repertory grids. Ten years after his initial two-volume work was published, he told Hinkle (1977) that if he were to revise the work he would probably delete the section on the repertory grid, because it seemed to him that "methodologically-oriented researchers had let it obscure the contribution of the theory."

[32] Heckerman and Horvitz (1987) have shown that while deterministic or logical rules are modular (in consequence of the assumption of monotonicity), rules reflecting an uncertain relationship are inherently non-modular (i.e., the degree of their truth or falsity depends on the state of belief in other propositions in the network). Pearl (1987) gives the following example of how plausible reasoning violates the conditions of locality and detachment necessary for modularity to exist:

VIOLATION OF LOCALITY

```
wet —> rain                      wet —> rain
wet               sprinkler & wet
------------------  ---------------------------------
      rain                                rain?
```

Unlike deductive reasoning, truth value of wet —> rain is dependent on state of belief in sprinkler ("explaining away" not possible).

VIOLATION OF DETACHMENT

```
   wet —> rain
sprinkler —> wet
-----------------------
sprinkler
===============
          rain?
```

The abilities to perform abductive reasoning and to "explain away" come as a natural consequence of the bi-directionality of probabilistic inference.

1987; Henrion & Cooley, 1987) as a way of representing decision knowledge. We see the approach in this paper as the first step toward a full integration of repertory grid and influence diagram representations within the same framework. The display of information in grids is used to emphasize patterns of similarities and differences, while network displays of the same or different information will highlight relationships between model variables. Both graphical representations could rely on the same sets of underlying solution algorithms and analysis tools.

2.4.3 Inference

Although *Aquinas* contains several systematic procedures for creating information and preference models, none of the steps presume an exclusive commitment to a particular mathematical approach. Our intention is not so much to propose an single ideal set of numerical procedures as to outline a perspective on how a personal construct approach, combined with the techniques of decision analysis, can provide help in the initial structuring and subsequent refinement of a problem. The choice of a crude but realistic set of distinctions may serve the purposes of modeling far better than a refined model that fails to capture important problem determinants.

In some cases, simple inference procedures such as linear models with subjectively assessed weights easily outperform experts (Dawes & Corrigan, 1974; Carroll, 1987). The robustness of linear models in particular situations is due to mathematical reasons as well as the fact that the models are more consistent and reliable than human experts. Wise and Henrion (1986) compared the performance of six common approaches for reasoning under uncertainty and found that most performed well when there was strong evidence in a single direction. However under conditions of weak or conflicting evidence, heuristic approaches performed poorly, sometimes performing no better than random. We conclude that the robustness of a heuristic or simple linear approach is situation specific and that the assumption that different inference schemes always produce similar results can lead to costly error.

As we have extended the capacity of *Aquinas* for dealing with problems where greater precision is necessary, we have tried wherever possible to maintain the simplicity of the original approach used in ETS. Mechanisms for dealing with more complex structures or precise values are not invoked for persons not requiring that level of analysis. Decision theory can be effectively used to select heuristic default or approximation strategies for model refinement and reasoning where precision is not needed or where resource constraints make exact strategies too costly (Horvitz, 1987a, 1987b. 1987c). The ideal would be to provide ways of performing sensitivity analysis dynamically as the model is being built so that model components are added or more precise numerical quantities are assessed only when it will materially affect the results. The robustness of the model to simplifying mathematical assumptions could be verified as well, through similar procedures. In this way, the transition from an incomplete, crude model to a fully specified and refined one can be made incrementally. This approach is consistent with the spirit of recent research in order of magnitude reasoning (e.g., Mavrovouniotis & Stephanopoulos, 1987; Raiman, 1986).

ETS represents knowledge in a single grid whose ratings are assigned directly. Since *Aquinas* knowledge base representation has become more complex (e.g., multiple linked grids or nodes) and the size of the problems to which it can be effectively applied has grown, the implementation of efficient rating *derivation* mechanisms has become important. Various derivation mechanisms allow missing ratings to be inferred from information elsewhere in the knowledge base. For this reason, inference is something which takes place throughout all phases of knowledge base construction and refinement, not merely during consultations.

Rating derivation depends largely on two complementary processes: *inheritance* and *recognition*. Inheritance allows the system to infer missing values by looking up properties of related concepts in a conceptual hierarchy[33]. Recognition, the dual of the inheritance problem, enables the system to find concepts that best match a partial or complete description. A plausible computational account of inheritance and recognition is important not only because these processes are ubiquitous, but also because humans perform such inferences effortlessly and extremely efficiently. Shastri (1987) has suggested that inheritance and recognition may be the basic and unitary components of symbolic reasoning[34].

---

33 Although not fully implemented in *Aquinas* at the present time, the approach allows for situations of multiple as well as single inheritance in a network.

34 Shastri (1987) has implemented a connectionist network (similar in some respects to our own approach) that demonstrates how a limited class of inheritance and recognition problems can be efficiently solved. With

Within *Aquinas,* the principle of maximum entropy is applied to deal with uncertain and incomplete information. Theories of uncertain inference can be thought of as containing both *static* components (i.e., how to assure consistency of beliefs at a given point in time) and *dynamic* components (i.e., how to update beliefs in light of new evidence). Most advocates of maximum entropy inference regard it as the dynamic component of a package whose static component is standard probability theory (Hunter, 1986). The default maximum entropy evidence combination rule used by *Aquinas* is incremental, commutative, and associative, thus sharing most of the attractive features of the Dempster-Shafer evidence combination rule while being consistent with Bayesian probabilistic inference (Shastri and Feldman, 1985). Probabilities are assigned so as to maximize the entropy consistent with whatever information is available. That information acts as a constraint upon the maximum entropy calculation. Although the the preferred approach is the probabilistic one, additional approaches to uncertainty can be accommodated within *Aquinas*. For example, a certainty factor calculus is available in the system, if there is a need for compatibility with other knowledge-based tools relying on that approach.

For modeling preferences, our current implementation features an efficient method that works well in moderate-risk situations for which an additive, linear utility function is appropriate. The system is designed in such a way that more complex mechanisms for constructing and reasoning with preference structures (e.g., Howard & Matheson, 1984; Keeney & Raiffa, 1976) could be added straightforwardly.

One significant feature of the *Aquinas* architecture is that it can be implemented as a highly parallel network made up of active nodes. There is no need for a central controller or an interpreter distinct from the network itself.

2.4.4 Modeling process of *Aquinas*

There are two ways that preferences could be used in *Aquinas*. First of all, they are needed to make appropriate recommendations about alternatives, based on available evidence. Secondly, preferences can guide the modeling process itself, by focusing attention on the most important discrepancies between models and evidence and suggesting model improvements.

Figure 10 is an idealized view of the modeling process of the system in terms of Gaines' construction hierarchy. At each level of the hierarchy, there is a channel for expressing a certain type of surprise and a corresponding channel for imposing preferences at that level through model refinement (cf. Figure 8):

> "Surprise at the lowest level of the hierarchy corresponds to distinctions being inadequate to capture events; surprise at the next level to inadequate variety to experience events; at the next level to inadequate approximation to predict events; at the next level to inadequate simplicity to explain events; at the next level to inadequate comprehensiveness to account for events." (Gaines, 1987a)

---

respect to the biological plausibility of such a connectionist system, he observes:

> "The proposed encoding is certainly not intended to be a blueprint for building 'wetware.' Yet it does satisfy nearly all the constraints proposed in (Feldman & Ballard, 1982). The only serious violation of biological plausibility is the requirement that nodes perform high precision multiplication. One may interpret the connectionist system described here as an ideal realization of a formal model of evidential reasoning. One can try and identify more plausible 'approximations' of the ideal system and study the manner in which their response deviates from the prescribed behavior."

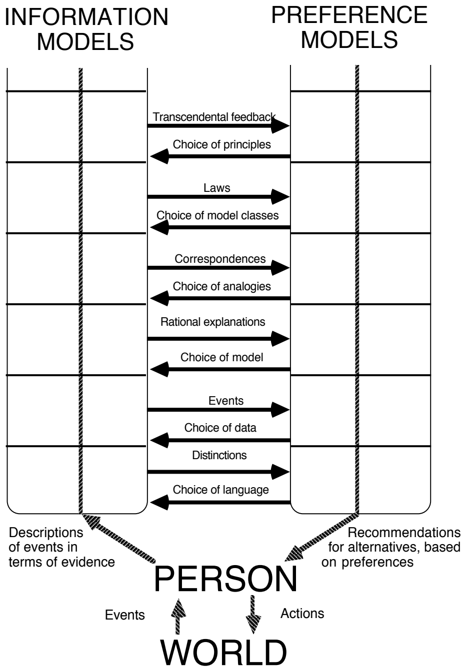

Figure 10: Idealized view of the modeling process of *Aquinas*.

At each level of the hierarchy, the goal of the modeling process is to help experts minimize surprise by analyzing the knowledge base and suggesting appropriate model refinements. Currently, the lower three levels are addressed in *Aquinas*. For example, at level one *Aquinas* identifies elements that are highly similar and helps experts discriminate between them by eliciting new constructs; at level two, the system performs analyses that reveal incompleteness and lack of representativeness of the set of elements and helps experts add new data appropriately; and at level three, the system verifies that generalized accounts square with predictions. A more complete description of the analysis tools in *Aquinas* may be found in Boose and Bradshaw (1987a) and Bradshaw and Boose (1988).

Good model building tools help persons manage the tradeoffs between model accuracy and preferences such as small model size and efficiency. Rather than encouraging experts to add model variables indiscriminately, the combination of techniques from decision analysis and personal construct theory focus the attention of domain experts on the aspects of the model most worthy of attention. This *value-driven* approach to model refinement can be applied to all phases of knowledge acquisition.

2.4.5 Steps in model building

The previous discussion has attempted to clarify the crucial role of preferences at two levels of the model building enterprise: first, as a means of representing the outcomes of interest for a particular person and decision, and secondly as a means of focusing the individual's attention during the refinement of the decision model. The specifics of the modeling process of *Aquinas* can best be understood through applying it to a problem exhibiting key features of the approach. We will follow four steps in building and testing our model:

1. Build an initial model of preferences and alternative actions.

2. Perform sensitivity analysis and look at *value of information* and *value of control.*

3. Build models of information relevant to sensitive variables.

4. Test the combined model, and continue appraisal and refinement until performance is acceptable.

3 MODELING A SIMPLE DECISION: THE TWO DOGS PROBLEM

3.1 THE PROBLEM

We will begin the discussion of our approach by introducing an example that illustrates how consideration of preference issues is essential to any discussion of learning or decision making. For the sake of brevity and comprehensibility, our problem is a simple one. However, it will become evident in our discussion how the same approach could be used for larger, more complex problems.

The *two dogs problem* begins in this way:

> "I had two dogs. They had to go into quarantine for six months. It was a hot summer. At the beginning of that time, they thought that all wasps were flies and at the end of that time, they thought that all flies were wasps" (King, 1979)[35].

This story raises some questions: Did the dogs learn anything during their quarantine? If so, what is the nature of the knowledge that was acquired? How can

[35] King continues: "This is my answer to all people who are shallow enough to believe in optimism." [But see Scheier and Carver (1987).]

we measure whether or not the dogs' situation has improved as a result of their experience? On the one hand, it could be argued that the dogs had learned nothing, since there seemed to be no increase in their ability to discriminate between the two types of insects. It is as if they had just substituted one overgeneralization ("All wasps are flies") for another ("All flies are wasps").

However, the dogs' *behavior* changed as a result of their experience. The quote implies that at the end of the summer the dogs avoided all flying insects whereas before they did not (they had no fear that "flies" would sting them). This change indicates that the dogs learned *something*, even though there was no gain in the dogs' ability to discriminate between the insects that could sting and those that could not. We can suppose that this learning consisted, in part, in the creation of a new *alternative* (i.e., avoiding the insects) that better satisfied the dogs' preferences.

While learning is typically measured solely in terms of *information* (e.g., as an increase in discrimination with a decrease in information required), we would argue that learning is ultimately a matter of increasing the expected *value* of the best course of action available to an individual for a given context. A metric for evaluating improvement in the dogs' decisions as a result of their experience takes their knowledge about the stinging potential of insects into account only to the extent that it has an effect on their ability to avoid being stung[36]. To a dog that is unable to avoid being stung or ignorant of his ability to do so, information about whether or not an insect will sting will be of little consequence. In other words, probabilistic assessment and inference are meaningless exercises unless adequate attention has been given to alternative generation and preference modeling.

We will now build a representation of the dogs' preferences for alternative actions and their information about the stinging potential of various kinds of insects in terms of repertory grids and directed graphs. This representation should not merely reflect what the dogs currently do (i.e., *descriptively*), but, more importantly, give them insight into how they might improve their models.

3.2 BUILDING A PRELIMARY MODEL OF ALTERNATIVES AND PREFERENCES

Many methods for helping experts discover and refine effective problem-solving knowledge are found in *Aquinas*. Because the focus of this paper is on decision analysis techniques rather than personal construct techniques, these will not be described here. More detailed accounts can be found in Boose and Bradshaw (1987a) and Bradshaw and Boose (1988).

---

[36] This metric should, of course, also take into account any costs associated with the learning.

3.2.1 Repertory grid representation

In modeling the dogs' preferences we hope to discover and represent key tradeoff variables relating to the alternatives. When preferences are modeled explicitly and separately from evidence, changes in preference parameters can be made directly, and without concern for unforeseen interactions with evidential variables.

We can represent preferences for avoiding and not avoiding flying insects for one of the dogs in the simple repertory grid shown in Figure 11. The alternatives (*elements*) **Avoid** and **Don't-avoid** form the grid columns, and the preference variables (*constructs*) **Level-of-effort** and **Get-stung?** form the rows. The *ratings* for **Level-of-effort** reflect the fact that avoiding the insects required more effort than not avoiding them and the ratings for **Get-stung?** reflect the dog's judgment that avoiding the insects was 100% effective in preventing stings, while not avoiding them resulted in a 50% chance of getting stung.

Once *Aquinas* has helped the expert come up with a set of elements, constructs, and ratings, the information in the grid can be used to make decisions. In making the decision, the dog has two objectives: 1. to minimize effort (the construct **Level-of-effort**), and 2. to minimize chances of getting stung (the construct **Get-stung?**). In this simple case, we will assign units of value for particular outcomes arbitrarily: a value of 1 for the preferred outcomes (Low, No) and a value of 0 to the non-preferred ones (High, Yes). This assignment of value to each outcome will be used to compute the overall *expected value* for each of the two alternatives (see below).

The two objectives are not of equal importance to the dog: the "cost" of getting stung is seen as greater than the "cost" of putting out effort to avoid the insect. The simplest way to represent the relative importance of objectives in *Aquinas* is through *preference weights*. The dog assigns .3 to **Level-of-effort** and .7 to **Get-stung?** which reflects the dog's judgment that the first objective is a little more than twice as important as the second one[37]. A pairwise comparison procedure based on a procedure similar to Saaty (1980) is available if a more careful assessment of preference weights is desired.

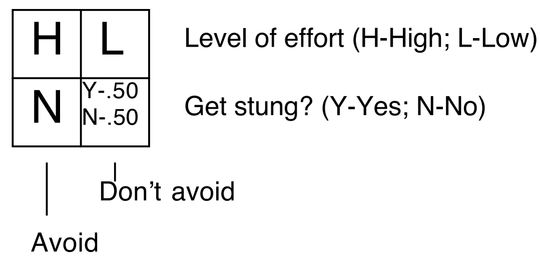

Figure 11: Repertory grid representation showing alternatives and preferences for one of the dogs. A distributed rating is shown in one of the grid cells.

3.2.2 Computing expected value for the alternatives

Sometimes it is appropriate to make assumptions that simplify calculations of overall preferences for alternatives. Assuming that the overall strength of preference for an alternative is a function of the sum of the individual preferences (*additivity*) and there is no significant risk involved (*linearity*), we will compute the expected value of an alternative in this grid by summing the value of each possible combination of outcomes weighted by the probability of that set of outcomes. First, we sum the products of the preference weight by the value of each objective, for each combination of possible outcomes (getting stung, not getting stung). The values for each possible combination of outcomes are in turn multiplied by the probability of that outcome:

Avoid:
$(((.3x0)+(.7x0)) \text{ x } 0) + (((.3x0)+(.7x1)) \text{ x } 1) = .7$
Don't-Avoid:
$(((.3x1)+(.7x0)) \text{ x } .5) + (((.3x1)+(.7x1)) \text{ x } .5) = .65$

The calculation of expected value[38] gives us a single number that can be used as a criterion for the

---

[37] While the use of preference weights provides a quick and intuitive way to get started, users may sometimes want to bypass this step to go directly to "willingness to pay" estimates or other preference elicitation methods. Keeney and Raiffa (1976) criticize the use of preference weights in multiattribute decision problems on the grounds that such weights are insensitive to the range of alternatives — weights should increase when the range increases and decrease when the range decreases. *Aquinas* avoids this problem by asking for ratings in absolute rather than relative terms. In addition, preference weights are not assigned until after the ratings are made, so that a context can first be established. Von Winterfeldt and Edwards (1986) have proposed another solution to the range problem called "swing weighting."

[38] For many decisions it is important to know not only the expected value but also the *value lottery*. The value lottery is the distribution of values across all uncertain outcomes. For example, in the Don't-Avoid alternative described above, there is a .5 chance of a value of .3 and a .5 chance of a value of 1.0. The expected value is just the mean of these values, weighted by their respective probabilities.

selection of an alternative[39]. The value is an *expectation* rather than a *certainty* because we cannot control the outcomes of uncertain events. This approach allows us to minimize our the chances of undesirable outcomes, but we can never completely eliminate risk. It is also important to remember that poor decision processes can sometimes lead to good results:

> "An Israeli visited the United States. When he came out of the airport, he hailed a cab. The driver looked at him and said, 'Are you an Israeli?' The man was surprised, saying, 'Yes, I'm an Israeli; how did you know?' The driver answered, 'That's easy. You wear a patch on your eye. Moshe Dayan also wore a patch on his eye. Dayan was an Israeli. So I figured you are, too" (Beyth-Marom & Dekel, 1985).

Although a good decision does not *guarantee* a good outcome, in the long run good decision processes make good results more *likely*.

3.2.3 Converting preferences to a meaningful unit

It is often useful to express the expected values of alternatives in a meaningful unit such as dollars. One way to do this is to ask experts to tell us how much they would pay to transform an undesirable set of outcomes into a better set. In this case, the dog indicates that he would pay $50 to go from a situation where he had been stung despite his best efforts to avoid the insect to a situation where he would not be stung and would have to make no effort. Other dogs might pay more or less for this outcome; each dog can be modeled separately. Consistent with our assumptions of additivity and linearity, we infer default values for the two traits:

Level-of-effort value = .3 x $50.00 = $15.00
Get-stung? value = .7 x $50.00 = $35.00

If the individual disagrees with these values, he can assign other ones. Otherwise, *Aquinas* assigns a default value of $35 to the situation where the dog is not stung and the level of effort is high, and $15 to the situation where he is stung and makes no effort (Figure 12):

Not-stung/Effort:
(0 x $15.00) + (1 x $35.00) = $35.00

Stung/No-effort:
(1 x $15.00) + (0 x $35.00) = $15.00

Stung/Effort:
(0 x $15.00) + (0 x $35.00) = $0.00

These values for each situation can be modified as well, if they are not consistent with the values of the person. Our approach is for the system to make reasonable default assumptions wherever possible while allowing the person to modify any default value, thus reducing the burden of effort in the knowledge acquisition task for the typical problem in a given domain.

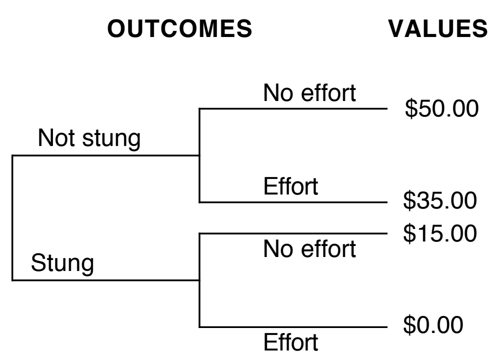

Figure 12: Subjective units of value can be converted to units such as dollars by asking experts how much they would pay to transform less desirable outcomes into more desirable ones.

The use of the "willingness to pay" approach to convert the value of non-monetary, intangible outcomes into a single monetary equivalent is more difficult for some types of decisions than others (e.g.,

---

[39] Given the preferences and alternatives expressed in the grid, the current state of information about the probability of getting stung, and the acceptance of the axioms of decision theory, the dog should consistently choose to avoid any flying insect that comes his way. Perhaps our assumption that the dog thought all flies are wasps was wrong; perhaps he was only doing what would be expected of an expected value decision maker rather than a naive dog:

> "…a naive dog, offered repeated situations in which some X sometimes means that he is to exhibit behavior A and at other times means that he should exhibit behavior B, will settle down to *guessing*… Such a dog will settle down to reflecting the approximate frequencies of appropriate response. That is, if the stimulus object in 30 percent of cases means A and in 70 percent means B, then the dog will settle down to exhibiting A in 30 percent of the cases and B in 70 percent (He will not do what a good gambler [or a dog who made his decisions by expected value] would do, namely, exhibit B in all cases.)" (Bateson, 1979).

medical decision making, personal problems, societal decisions affecting the welfare and safety of groups). Tversky, *et al.* (1987) suggest that the salience of intangible preference dimensions may be diminished by this step, rendering monetary concerns more important and vivid than they might be otherwise. Other approaches have been developed to assist decision makers where this becomes a problem (Fischhoff, Slovic, & Lichtenstein, 1980; Howard 1980a, 1984; Keeney & Raiffa, 1976).

Nothing in this approach requires that expected value be expressed in terms of dollars, however we will see later on how the appropriate use of this technique can facilitate some important aspects of model refinement.

3.2.4 Risk attitude and time preference

When the least preferred outcome of an objective is not "too bad" and the spread between the most and least preferred is not "too great," expected value is usually a good approximation of a person's optimal decision criterion. Otherwise, an individual's *risk attitude* should be taken into account. As an illustration, consider the fact that many people are *risk preferring* or *risk neutral* for small amounts (e.g., a double-or-nothing gamble on a few dollars), whereas they would be *risk aversive* for larger amounts (e.g., a double-or-nothing gamble involving the value of their home)[40]. Where risk attitude is an important consideration, the worth of alternatives should be expressed in terms of *expected utility* rather than expected value. The expected utility of an alternative has also been called its *certain equivalent*. The certain equivalent is defined as a value such that a decision maker would be indifferent to either receiving that value for certain or facing the uncertain outcomes of the decision. A risk aversive individual might be indifferent between $400 cash and a 50-50 chance on $1000, while a risk neutral individual (i.e., expected value decision maker) would be indifferent to $500 and the same bet.

In problems where the value from a decision is not fully realized until some time in the future, decision makers may wish to specify a discount rate or an interest rate that reflects the fact that money received in the future is usually worth less than money received today. A figure called net present value (NPV) can be calculated from the discount rate over the life of the investment and used as the decision criterion in such cases.

For this problem, we will not assess risk attitude or time preference[41].

3.2.5 Hierarchical decomposition of preferences

It is often useful to decompose preferences hierarchically. For example, if the dog found it difficult to give ratings for **Level-of-effort**, we could break that construct down to form a sub-grid where the constructs would be various components of effort (e.g., effort to twitch tail, effort to rise, effort to move) and the elements would be **Avoid/Don't-avoid**[42]. The unrated grid cells at the higher level could then be filled in (i.e., *derived*) automatically by *Aquinas* from information in the sub-grid. Additional examples of rating derivation are given in section 3.4.

3.3 APPRAISING THE PREFERENCE MODEL

Model appraisal is a way of focusing attention on the most important or *sensitive* components of a decision. Model comprehensibility is diminished when variables without operational significance are included in the decision basis. On the other hand, the refinement of sensitive variables through the addition of new model components can often be of great value. Consistent with these observations, model appraisal has two goals: 1. to discover whether the model can be simplified by eliminating variables or alternatives that do not affect the final decision, and 2. to discover how changes in values of one or more variables affect the selection of alternatives so that areas needing more careful modeling can be identified. Although both are important, we will illustrate only a single example of the second goal here.

3.3.1 Sensitivity analysis

There are several kinds of sensitivity analysis (e.g., deterministic, probabilistic, risk) that can be performed as a way of refining the decision basis. Here, we will give an example of *probabilistic*

---

[40] Different approaches have been taken to the problem of measuring multiple objectives together (*commensuration*) and the measurement of risk attitude. Keeney & Raiffa (1976) derive a multiattribute utility function by modeling attribute commensuration and risk attitude simultaneously, whereas Howard & Matheson (1984) favor the separation of the commensuration process from risk-attitude modeling.

[41] Wellman (1986) presents an interesting approach to the automated verification of assumptions about utility models.

[42] Howard (1987) calls such networks of values "value diagrams."

sensitivity analysis, showing the model's sensitivity to variations in uncertainty for particular variables.

The one uncertainty in our preference model is whether or not a particular insect will sting when no effort is made to avoid it (the lower right grid cell in Figure 11). In probabilistic sensitivity analysis, we repeatedly calculate the expected value for each alternative while varying the probability of each outcome. Figure 13 shows how the expected values of the two alternatives vary with the probability of getting stung. The expected value of **Don't-avoid** decreases as the probability of getting stung becomes greater (not avoiding is only a good strategy when it is unlikely you will get stung), while the expected value of **Avoid** remains constant no matter how likely it is that you will get stung (assuming that if you avoid all insects regardless of the situation, you will never get stung although you will always incur the same cost of making efforts to avoid them).

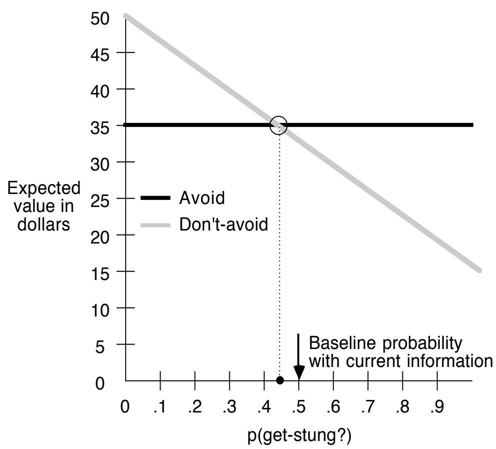

Figure 13: Graph showing how the probability of getting stung affects the expected value of the alternatives.

The two lines representing the expected value of the alternatives cross where the probability of getting stung is approximately .43. At .43,

Avoid:
(((.3 x \$0) + (.7 x \$0)) x 0) + (((.3 x \$0) + (.7 x \$50.00)) x 1)

is approximately equal to

Don't-Avoid:
(((.3 x \$50.00) + (.7 x \$0)) x .43) + (((.3 x \$50) + (.7 x \$50.00)) x .57)

This means that if the probability of getting stung were anything greater than .43, the best alternative would always be to avoid any flying insect. This is consistent with the expected value we computed earlier showing that, with our current assessment of a 50% chance of getting stung, it was better for the dogs to avoid the insects. However, if the dog could reliably predict that the chance of getting stung were less than .43, it would be better for it to relax and ignore the intrusion than to take evasive action.

Since our sensitivity analysis showed that reducing the dog's uncertainty about getting stung could result in our choosing a more highly valued alternative (Don't-avoid rather than Avoid), we will examine this variable more carefully by determining the *value of control* and the *value of information* (Matheson, 1988).

3.3.2 Value of control

The lines labeled "Don't-avoid with perfect control" in Figure 14 represent the expected value of the decision if the dog could somehow eliminate the possibility of getting stung by taking preventive action. He could, for example, build an enclosed shelter where no insects could enter. In such a case, he would never have to choose the **Avoid** alternative. The difference in expected value between the best alternative given control and the best alternative without control for a given probability is called the *value of control*. The value of *perfect* control (i.e., complete elimination of risk) can be used as an upper bound on what the decision maker should spend to generate and implement alternatives that will increase control of uncertain outcomes. Usually, the value of a new alternative that increases control over sensitive uncertainties is much greater than the value of extensive analysis of existing ones. In our example, the dog should spend no more than \$50.00-35.00=15.00 (per insect) to pursue this alternative assuming the probability of getting stung is greater than .43 and the shelter is free. The incremental value of control over no control diminishes as p(get-stung?) gets smaller than .43.

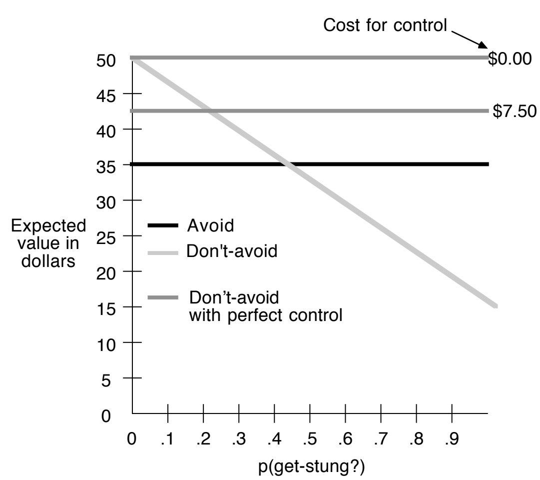

Figure 14: Graph showing the expected value of alternatives and the value of control at different costs varying with the probability of getting stung.

The figure also shows how the cost for achieving control directly affects its value. If the cost were $7.50, the range of situations where the shelter would make sense is diminished. For example, in that case, if the probability of getting stung were to be less than about .2, it would be better better to do nothing at all. A shelter that cost $15.00 (per insect) would be worthless, since it would have the same expected value as the **Avoid** alternative for all values of p(get-stung?).

3.3.3 Value of information

The lines labeled "Don't-avoid with information" (Figure 15) represent the expected value if the dog could know with greater certainty whether the insect would sting or not before he made his decision.

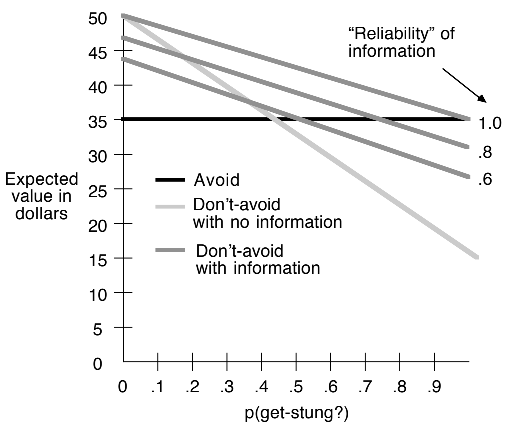

Figure 15: Graph showing the expected value of alternatives and the value of information at different levels of "reliability" varying with the probability of getting stung.

The line with a "reliability" of 1.0 represents the expected value of choosing the best alternative, given perfect information about whether or not the insect will sting. The *value of perfect information* is the upper limit on what a decision maker should spend on information gathering activities. It is computed by taking the difference between the expected value of the best alternative given perfect information and the expected value of the best alternative without information. Since in this case, the dog has only perfect anticipation rather than perfect control, we expect the value of information to be less than the value of control. The best the dog can expect is that he will have to choose the **Avoid** alternative with a relative frequency corresponding (in the long run) to the probability of getting stung. Suppose, for example, that the dog makes a number of observations and finds that he gets stung 50% of the time when an insect flies by. If the dog knew perfectly in advance whether he was going to get stung, he would end up choosing **Avoid** about 50% of the time on the average and **Don't-avoid** about 50% of the time. Figure 16 shows that if the dog could buy a stinging-insect detector that performed with 100% accuracy, he should pay no more than $42.50-$35.00=$7.50 for it, given our current state of information (p(get-stung?) = .5).

Imperfect information is less valuable than perfect information. The graph shows how much the dog ought to pay for a wasp detector that was 80% or 60% reliable (e.g., the insect would sting when it detected "sting" 80% (60%) of the time; the insect wouldn't sting when it said it wouldn't 80% (60%) of the time)[43]. Of course, a detector that is too unreliable would be worthless.

Generally, information is only available at some cost. Figure 16 shows how the value of perfect information is affected by the cost of that information.

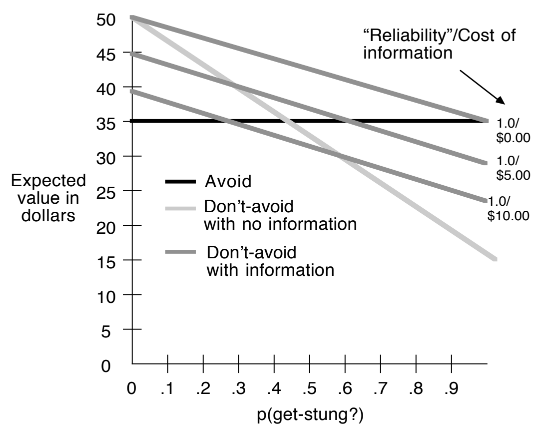

Figure 16: Graph showing the expected value of alternatives and the value of perfect information at different costs varying with the probability of getting stung.

3.4 BUILDING A MODEL OF INFORMATION

As the result of our appraisal, we could either design a new alternative (to increase control) or we could model what the dog knows about the chances of getting stung in particular situations more carefully

[43] The "reliability" of an information source is not always symmetrical. For example, a detector may be able to better identify positive instances than negative ones.

(to increase information)[44]. For this problem, we will assume that the dog is not in a position to to avail himself of a shelter. The best he can do is to increase what he knows before he commits to the avoid or not-avoid strategy, in order to make a better decision.

Aquinas has several methods to assist experts in identifying items of information that bear on uncertain outcomes of interest. From his interactions with the system, the dog determines that two major factors increase the likelihood of an insect stinging:

1. Likelihood that the insect is of a type that *can* sting.

2. Length of time the insect stays close by, given that it is a "stinger."

In considering the first factor, *Aquinas* helps the dog to think about differences he has noticed between insects that have stung him and those that have not. Figure 17 is a graphical representation of the results[45]. "Stingers" and "Non-stingers" have been assigned relative positions based on the dog's ratings for three dimensions: *size*, *origin*, and *mood*. The amount of uncertainty in the rating assignment is represented by the width of the stinger and non-stinger blocks along a particular dimension, indicating that the dog is unsure of where the exact value lies.

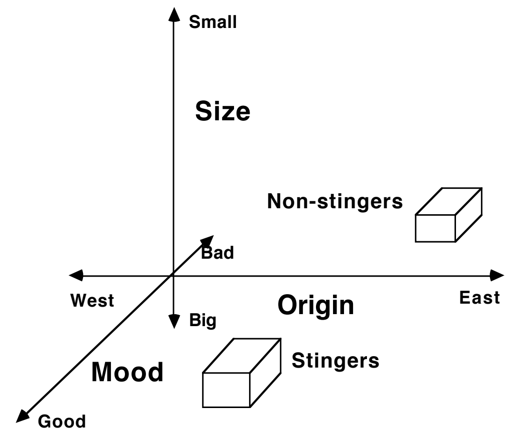

Figure 17: Dog's view of uncertain evidence relating to the insects' appearance (after Rucker, 1984). The axes are orthogonal to one another, reflecting the dog's initial belief that the dimensions are independent.

From the figure, we can see some differences: Stingers are more likely to come from the west than Non-stingers and, although bigger than Non-stingers, seem generally to be in a better mood. The fact that the depth of the blocks is greater than the width shows that the uncertainty about the mood of the insects is greater than the uncertainty about their origin — it is easier for the dog to detect origin than mood.

There are three ways we can quantify and validate the beliefs the dog holds about the characteristics of the insects:

1. Derive probabilities from data[46];
2. Ask experts to assign probabilities;
3. Combine probabilities from data and experts.

The three methods for generating probabilities will be illustrated in sections 3.4.1, 3.4.2, and 3.4.3. From the data, we will show how a classification model reflecting regularities in the example set can be induced using Bayesian and information theoretic techniques. At a later point, we will use evidence from the dog's observations about new insects to make probabilistic predictions over the set of possible

[44] Heckerman and Jimison (1987) provide an extensive discussion of of how decision theory can be used at the metalevel to balance the benefits of extending the conversation in a given region of a model against the costs associated with model extension (see also Matheson, 1968). Although we do not consider the costs of extending the model explicitly here, we will discuss related issues of cost/benefit tradeoffs in determining questions to ask at consultation time below.

[45] Kelly's original conception was of a non-Euclidean geometry and a nonparametric arithmetic for personal construct space (Kelly, 1961). This approach and the approach presented in the paper can be reconciled (see Bradshaw & Boose, 1988).

[46] Decision analysts tend to be uncomfortable with deriving probabilities directly from data. They hold that subjective interpretation is necessary in order to confirm acceptance of assumptions that may underlie the data. Extrapolation from frequency counts for example, presumes the belief that conditions in the future will be essentially similar to those in the past. Combining expert judgment and data is an even more controversial area that deserves further study.

classes (i.e., stinger, non-stinger)[47]. These predictions will be used in conjunction with the preference model to help the dog make decisions about what to do in a particular situation.

3.4.1 Deriving probabilities from data

Given a hypothesized set of differences between stinging and non-stinging insects, we can perform an experiment. The dog predicted that any new stingers would be, on the average, relatively bigger, in a better mood, and further west than the non-stingers. We will use the distinctions generated by the dogs (size, mood, and origin) and then watch a number of insects to determine the accuracy of the dog's predictions about the differences between stingers and non-stingers. Our results after observing eight insects, four of which stung, are shown in Figure 18.

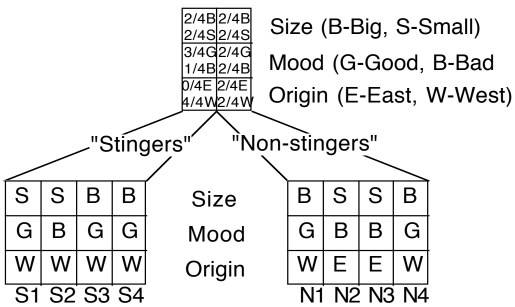

Figure 18: Results of observing eight insects, four of which stung and four of which did not. The rating distributions shown in the cells of the upper grid are the relative frequencies of each characteristic, based on observations shown in the lower grids.

Hierarchical decomposition of example set. Just as preferences can be broken down hierarchically within *Aquinas,* so elements can be reasoned about with grids at different levels of abstraction. The top grid in Figure 18 derives its rating value distributions for each column from the grids of examples shown. This grid could be used in turn to derive distributions based on observations of both stingers and non-stingers for the element "flying insects" in a possible higher level grid.

The approach of defining more abstract concepts in terms of probabilistic descriptions of less abstract ones rather than as deterministic descriptions of pure prototypes (e.g., Bruner, Goodnow, & Austin, 1956) is both intuitively appealing and consistent with recent psychological research (Cohen & Murphy, 1984; Medin & Smith, 1984). The intermediate view of Rosch (e.g., Rosch, 1975; Rosch & Mervis, 1975) that concepts are based around prototypes of most typical members is hard to distinguish from the theory that concepts are defined in terms of features with associated probabilities.

Aquinas can also use information in higher level grids to derive missing values in lower ones via an inheritance mechanism (Collins & Quillian, 1969; Johnson-Laird *et al.,* 1984). Holland *et al.* (1986) conclude on the basis of their research that:

> "To know that an instance is a member of a natural category is to have an entry point into an elaborate default hierarchy that provides a wealth of expectations about the instance."

Limitations of common methods for learning from data. Many approaches have been proposed for inducing generalizations from data. Most empirical learning systems have at least one of the following limitations:

1. Inability to find or use relationships that are only statistically true. Many empirical learning systems depend on procedures that look for relationships that are always or nearly always true (e.g., Michalski & Chilausky, 1980). Unfortunately, many knowledge-based systems are targeted for domains where causal mechanisms are incomplete or poorly understood. Such domains are dominated by probabilistic rather than deterministic relationships. For many years, the use of probabilistic inference in knowledge-based systems was thought to be impractical. However, recent advances in probabilistic modeling that demonstrate its practicality have been well documented in the literature (Cheeseman, 1985, 1986; Henrion, 1986, 1988; Pearl, 1985; Spiegelhalter, 1986a, 1986b). We can see by inspection of Figure 18 that the data set is very noisy. A non-probabilistic classification algorithm would be unable to discriminate between the classes because there is no single criterion or set of criteria that infallibly discriminates between the two classes of insects. Insects S3 and S4 are indistinguishable from

[47] Cheeseman *et al.* (1987, 1988) use class probabilities calculated for a particular case to compute the probabilities of other attributes of interest. They call the process of mapping from past data to future data via a model space "transduction" (cf. Shastri (1987) "pattern completion"): "Prediction is performed indirectly, by mapping the available information into a probabilistic class membership distribution (for a particular case), then making predictions based on this class membership." The concept of *weighted* membership of entities to classes is shown to be superior to abductive prediction where the *best* model is selected and the prediction is made as if it were a fact (Self & Cheeseman, 1987).

N1 and N4 with respect to the three constructs we have defined[48].

2. Fixed independence assumptions. Relationships between items of evidence are often complex. Complexity in probability assessment and reasoning can be greatly reduced if a researcher assumes a fixed relationship between items of evidence: i.e., either that each item of evidence occurs independently of the occurrence of any other piece of evidence (*conditional independence* assumption) or that the items of evidence are totally dependent on one another (*maximum dependence* assumption). Early attempts to model uncertainty probabilistically, such as PROSPECTOR (Duda & Reboh, 1984) and the original singly-connected Bayes network formulation (Kim & Pearl, 1983) relied on the independence assumption for reasons of computational tractability and to reduce the number of probability assessments required. Fuzzy set schemes, on the other hand, assume the maximum possible correlation between items of evidence. In some cases, these fixed assumptions can have a significant negative impact on reasoning (Wise & Henrion, 1986; Horvitz *et al.,* 1988)[49]. Ideally, a system should be able to flexibly apply and suspend independence assumptions as the particular characteristics of the problem and efficiency requirements dictate (see below).

3. Overfitting the model to the data. Even when inductive learning systems take probabilistic relationships into account, they may lack appropriate significance testing to assure that they are not "fitting the noise". Various tree-pruning heuristics or "stopping rules" have been discussed in the literature (e.g., Quinlan 1983, 1986) but no single agreed upon criterion exists.

Approach to learning from data. We are investigating ways of implementing a method for example-based learning that addresses each of these limitations. Cheeseman (1983, 1984, 1987; Cheeseman *et al.,* 1987, 1988) presents a general approach that seems promising.

Within this framework, we can regard an inductive generalization as a way of representing more briefly the information contained in the set of examples. The generalization we make will represent a tradeoff between our desire for accuracy and completeness on the one hand ("minimizing surprise") and our desire for parsimony on the other hand ("maximizing preference"). We would also like to avoid incorporating into our model the random characteristics of a particular data set (i.e., overfitting).

Our goal is to infer a set of probabilities that satisfies constraints in the example set, but is, of all such distributions, the least biased (contains the minimum information). Another way of saying this is that we want a distribution that contains no more and no less information than what is given in the examples. There are many distributions that satisfy the constraints, but there is a unique one, the maximum entropy distribution, that has minimum information (Hunter, 1986).

By taking into account the class to which the insect belongs and the relative frequency of each construct value for that class we can compute expected probability values for combinations of construct values under a generalized independence assumption

---

[48] For this example, we could have chosen more discriminating constructs, however we wanted to demonstrate the value of this approach for situations where even very noisy data.

[49] It has been shown, however, that "after optimization, the error due to approximate updating under exact assumptions can outweigh the error due to exact updating under approximate assumptions" (Wise, 1987; see also Wise *et al.*, 1987).

("maximum entropy"). The maximum entropy procedure and Bayes' rule are equivalent under the conditional independence assumption. Thus, to get the overall probability of an insect being a Stinger or Non-stinger, we compute:

$p(C_c \mid A_i, B_j, \ldots) = p(C_c)p(A_i \mid C_c)\ p(B_j \mid C_c)\ldots / p(A_i, B_j, \ldots)$.

where $p(C_c \mid A_i, B_j, \ldots)$ is the probability of the cth class given information about constructs $A_i, B_j, \ldots$ as items of evidence, and $p(A_i, B_j, \ldots)$ is a normalizing constant that comes from the requirement that $\sum_c p(C_c \mid A_i, B_j, \ldots) = 1$.

Where the frequencies for the combinations of construct values observed in the examples do not differ significantly from the computed maximum entropy probabilities, we can avoid having to represent the dependencies between constructs explicitly in a joint distribution table. Our best estimate in such instances can be efficiently computed because we can assume conditional independence. Where such differences are significant,[50] we can represent the observed values as explicit constraints on the maximum entropy calculation.[51]

This procedure allows us to estimate joint probabilities to precisely the accuracy justified by the noise in the data. The noise is due to the fact that the data is only a sample. If new data provides additional support for the significance of a value not previously found significant, the generalization can be easily and automatically revised.

Although this approach saves us from unnecessarily specifying a full joint distribution and effectively solves the problem of overfitting, the cost of computing expected probabilities increases exponentially with the number of known constraints. If this becomes a problem, the most effective use of information about significant dependencies between construct values is in the creation of new concepts (Shastri, 1985). For example, an effective use of the information "Most big non-stingers are in a good mood," would be to create the new categories of "Big-non-stingers" and "Small-non-stingers" and attach the appropriate information on mood and origin (Figure 19).

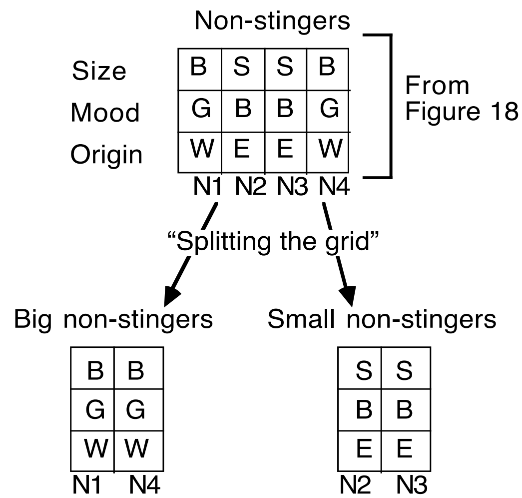

Figure 19: Using information about construct dependencies to create new concepts.

A probabilistic approach to classification and prediction, where a weighted degree of membership may be distributed across several categories, is consistent with psychological research and in contrast to machine learning approaches that require sharp definitions and clear class boundaries. Because the class information is given, it is unnecessary to search for the best class definition — it is only necessary to calculate a probabilistic description based on the example set (Cheeseman *et al.,* 1987, 1988).

The example we have just discussed is an instance of "supervised" or "tutored" learning: that is, the examples given to the system had been preclassified as either stingers or non-stingers. *Aquinas* can also "learn" in an unsupervised mode by performing a hierarchical classification of the examples using cluster analysis (Figure 20). The junctions in the clusters can be seen as conjectures about possible new classes of elements or constructs. The expert is asked to label nodes and expand clusters where possible (Boose & Bradshaw, 1987a).[52] If the

[50] A joint probability value is significant, in an information theoretic interpretation, when the minimum message length required to encode it, given its expected value, is longer than the minimum message length to encode the observed probability ignoring its expected value.

[51] Since the search procedure becomes combinatorially explosive for higher order joint probabilities, a heuristic is suggested by Cheeseman (1984) that they be investigated only when they have at least one significant lower order (i.e., binary) component joint probability.

[52] The current cluster analysis algorithm in *Aquinas* is a very efficient one based on construct similarity, however we are investigating other approaches. Cheeseman *et al.* (1987, 1988) describe a Bayesian criterion of similarity

clusters do not fit the expert's view of the data, this may be a sign that constructs or ratings are in need of revision.

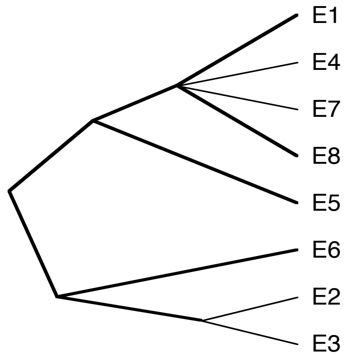

Figure 20: Results of a hierarchical cluster analysis performed on a set of eight examples.

3.4.2 Asking experts to assign probabilities

In addition to deriving probabilities from data, we can ask experts to assign probabilities directly. Figure 21 shows the dog's judgment of the probability of the values of each construct for stinging and non-stinging insects.

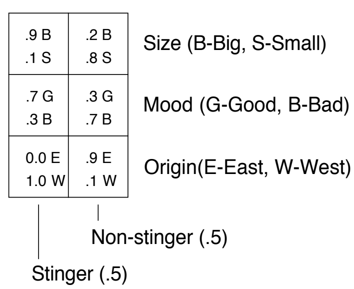

Figure 21: Probabilities obtained directly from the dog. The prior probabilities for Stinger and Non-stinger were also obtained (.5, .5).

Issues in direct assessment. Probability assessment has been the subject of much research (see, e.g., reviews and critiques by Shafer & Tversky, 1985; Spetzler & Stael von Holstein, 1975; Stael von Holstein & Matheson, 1978; Wallsten & Budescu, 1983). Several researchers have focused their attention on identifying and eliminating sources of systematic bias in human judgment under uncertainty (*debiasing*) (e.g., Kahneman, Slovic, & Tversky, 1982).

It is not always easy to get experts to make accurate numerical assessments of uncertainty (Eshelman *et al.,* 1987). Where such precision is not necessary, simpler methods are preferred. For example, in domains with a strong causal model, where there is little uncertainty or conflicting evidence, all reasonable representations of uncertainty (including rough qualitative ones) should perform reasonably well. However, when accurate numerical assessments are required, there are two features of the present approach that make the task somewhat less difficult.

First, by asking for assessments in terms of probabilities rather than other numbers such as certainty factors, anchor points can be clearly defined for the expert. Because the numbers have an unambiguous meaning we can make exact mappings between graphical representations such as probability wheels and histograms in *Aquinas* and probabilities. For example, a pie shaped slice representing 15% of the total area of the circle on a probability wheel has an unambiguous numerical correspondence to a probability judgment of .15. [53] Equivalent mappings for other numerical representations cannot be made as straightforwardly.

Second, we find that experts respond much more readily to a "backward approach to assessment." For example, many rule-based systems represent knowledge in the form:

IF <evidence>
THEN <hypothesis>, CF,

where CF is a measure of change in belief for the hypothesis given the evidence (e.g., IF the insect is small THEN it is a stinger, .5). However, Shachter and Heckerman (1987) have shown that experts are often much more able to give assessments in the opposite, *causal* direction: IF <hypothesis> THEN <evidence> (e.g., IF it is a stinger THEN it is small, .5). When filling in a repertory grid such as the one in Figure 21, experts are doing just that: they provide numbers representing the evidence for each construct, given the element.

While experts often prefer to specify probabilities in the *causal* direction, the direction of usage of evidence to calculate beliefs is usually in the

defined between hypothesized class descriptions and the case, rather than between pairs of cases as in traditional approaches. Boulton and Wallace (1973) present a different criterion based on information theoretic formulations.

[53] Having an unambiguous mapping, however, solves only part of the problem: Shafer and Tversky (1985) emphasize how the metaphor used by probability assessors (e.g., "betting" vs. frequency vs. propensity) can affect results.

diagnostic direction (i.e., from evidence (observable) to hypothesis (unobservable)). A significant advantage of the probabilistic over the rule-based approach to reasoning is the ability to perform inference in both directions from the same set of probabilities.

Obtaining information about dependencies. If the expert gives us only the assessments in the grid of Figure 21, we can say nothing about dependencies between constructs. If the expert is unwilling or unable to provide additional information, the best we can do is assume maximum entropy and combine probabilities as if the constructs were independent. If the expert can provide information about dependencies between constructs, we generate a grid like the one shown in Figure 22. We call this a *hypothetical example set grid* because the expert's task is to estimate the probabilities of occurrence of various combinations of construct values for a hypothetical example set.

This type of grid is nothing more than a multi-dimensional contingency table in a slightly altered format. From such tables for Stingers and Non-stingers, we can compute other common measures of uncertainty if desired. For example, the likelihood ratio is calculated as:

LR = p(Stinger|e) / p(Stinger|not-e)

where e is represents evidence for size, mood, or origin. Alternatively, we can convert to a certainty factor measure as follows:

CF= p(Stinger|e) - p(Stinger) / 1-p(Stinger) when e is evidence *for* the insect being a Stinger;

CF= p(Stinger|e) - p(Stinger) / p(Stinger) when e is evidence *against* the insect being a Stinger.

The default probabilities shown beneath the grid are the products of the probabilities for the corresponding column values of each construct in Figure 21. Experts adjust these probabilities until they are satisfied that the relative proportions of each combination of values are consistent with their experience.[54] The significance of the correlations between constructs can be evaluated in the same way as it would be if we had real observations available.

[54] According to the subjectivist view of probability, we do not need to draw a hard line between probabilities based on direct empirical data and those representing relative strength of belief (in the case of unique (i.e., non-repeatable) events).

Stingers

B	S	B	S	B	S	B	S	Size
G	G	B	B	G	G	B	B	Mood
E	E	E	E	W	W	W	W	Origin
0	0	0	0	.67	.07	.27	.03	

Figure 22: Hypothetical example set grid for Stingers. The probabilities shown beneath the grid are the product of the probabilities for the corresponding column values of each construct in Figure 21.

3.4.3 Combining probabilities from data and experts

The approach we have described makes it convenient to combine probabilities from data with subjective assessments. To do this, we ask experts to express their probabilities as fractions, with the denominator representing an estimate of the total number of cases on which the judgment is based and the numerator representing the fraction of the cases with a particular combination of characteristics (Spiegelhalter, 1986a, 1986b). Then we combine these fractions with those from our data (Figure 23). Again, the same significance test for the joint probabilities can be applied.

Stingers

	B	S	B	S	B	S	B	S	Size
	G	G	B	B	G	G	B	B	Mood
	E	E	E	E	W	W	W	W	Origin
Expert	0/30	0/30	0/30	0/30	19/30	2/30	8/30	1/30	
Data					2/4	1/4		1/4	
Combined	0/34	0/34	0/34	0/34	21/34	3/34	8/34	2/34	

Figure 23: Combining data and expert judgments.

This approach has several desirable features (Hink & Woods, 1987). First, measures of association converge towards "true" values as additional observations are made. Consistent with intuition, this convergence takes longer with initially confident subjective estimates (i.e., large denominators in relative frequencies of expert judgments). Secondly, psychological techniques for debiasing probability assessments are naturally incorporated into this approach. For example, subjects are required to focus on disconfirming as well as confirming conditions, to use a concrete measure of uncertainty (i.e., frequencies rather than probabilities), and to think about the body of experience they possess that underlies their judgments (i.e., estimate how many examples of each type they have encountered).

Calibration of knowledge base with accumulating data. Accumulating data can also be used as a comparison with the original assessments in order to suggest changes to the knowledge base when the system is too often "surprised" by what it sees. One measure of such discrepancies is the 'Brier score,' which is equivalent to the squared predictive error (Spiegelhalter, 1986a), i.e., the difference between an actual and a predicted event (a kind of surprise). Total agreement between an actual and a predicted event gives a score of 0, whereas the maximum discrepancy would produce a score of 1. For example, if an insect turned out to be a Stinger when its current probability was .75, the Brier score would be:

$$(1-.75)^2 = .0625$$

However, if the insect turned out not to be a Stinger, the score would be:

$$(0-.75)^2 = .5625$$

registering increased surprise. If a series of *n* events were recorded, the total Brier score would be $B = \Sigma (1-p)^2$. The expected Brier score if the system were reliable would be $E_0(B) = \Sigma\ p_i\ (1-p_i)$, with variance $V_0(B) = \Sigma\ (1-2p_i)^2 p_i(1-p_i)$. Thus $[B-E_0(B)]/V_0(B)^{1/2}$ can be used as a test statistic to determine whether the knowledge base is in need of calibration.

We have seen how this approach allows for a smooth integration of information derived from data, expert judgment, or both. The maximum entropy approach allows efficient inference under independence assumptions to take place when justified by an information-theoretic criterion. Information about dependencies can either be incorporated into the maximum entropy calculation as constraints or can be used to create new concepts. Direct assessment by experts is made easier by a probabilistic approach that facilitates a "backward" direction in making judgments. As knowledge and data is added to the system, it can be naturally incorporated and exploited for use in calibration of the knowledge base.

3.5 LINKING THE INFORMATION AND PREFERENCE MODELS

Figure 24 is a network representation of a model combining information and preference variables developed earlier in repertory grids. This form has some resemblance to an influence diagram, with a few differences. Unlike grids, a network representation emphasizes the relationships between variables. This is done at the expense of "hiding" details about the distributions within each variable. This combined model can be used to run test consultations such as the one described in section 3.6.

Oval shaped nodes (e.g., **Stinger/Non-stinger, Size**) represent constructs that serve as evidential variables in the information model, and arcs represent the possibility of probabilistic dependence relationships between pairs of nodes. The lack of an arc (e.g., between **Size** and **Mood**) signifies conditional independence. Rounded rectangles represent constructs that serve as objectives in the preference model (**Get-stung?** and **Level-of-effort**). These depend on the decision made (**Avoid/Don't-Avoid,** a rectangular decision variable) and determine the overall expected value (**Value of Avoid/Don't-Avoid**) of the two alternatives. The best alternative at a given point in time is the one with the with the highest expected value.

As we discussed earlier, rating distributions for each of the variables can be assigned directly by the expert or inferred as the result of rating derivation mechanisms. While static rating values were assigned for each alternative for **Level-of-effort**, **Get-stung?** will derive its values by querying the linked information model in a way that is somewhat analogous to a dynamic version of Howard's (1987) *knowledge maps*. Information about the current state of the world flows from the **Get-stung?** variable in the information model to the **Get-stung?** variable in the preference model. In return, direction in model refinement flows from the preference model to the information model when sensitivity analysis reveals value in more accurate modeling. The probability of the outcomes of **Get-stung?** in the information model is derived by combining evidence about the **Length-of-time** an insect is nearby and whether or not the insect appears to be a **Stinger** or **Non-stinger.** The **Stinger/Non-stinger** node in turn derives its values from any evidence that may be available about the insect's **Size, Mood,** or **Origin**. If no evidence is available, the best assessment we can make for **Stinger/Non-stinger** is the prior probability that the insect belongs to either class.

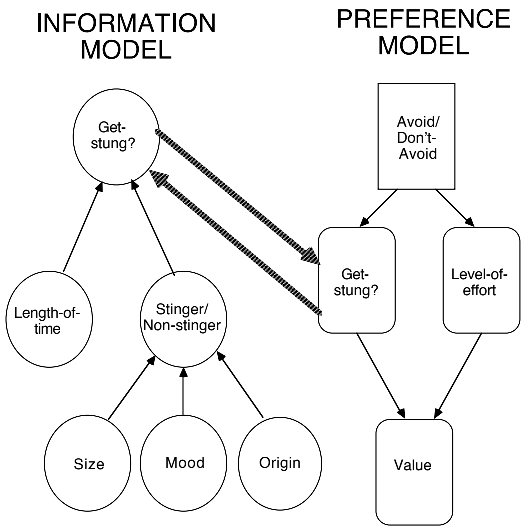

Figure 24: A representation of the combined information and preference models.

3.5.1 Assessing the cost and expected benefit of obtaining information

Our goal in building the information model was to reduce the dog's uncertainty of an insect's stinging potential by providing a framework in which additional relevant information could be obtained. We saw as a result of value of information analysis that this reduction of uncertainty had a direct effect on expected value. The more certain the dogs are about the kind of insect they are dealing with, the more likely they are to achieve a good outcome as a result of their actions.

So far we have ignored the *cost* of asking and answering questions. Sometimes it may be easier to just avoid the insects than to have to make careful observations. Besides, gathering information takes time, and the longer a response is delayed the greater the likelihood of getting stung.

We can represent the alternative of answering questions about the insects to gain additional information as a third column in the preference grid (see Figure 25)[55]. For this third alternative, the rating values for **Level-of-effort** will change dynamically depending on which question is being asked, since some questions may be more difficult to answer than others[56]. The rating values for **Get-stung?** will also be determined dynamically as a function of the current state of information plus the information expected as a result of gathering an additional item of evidence.

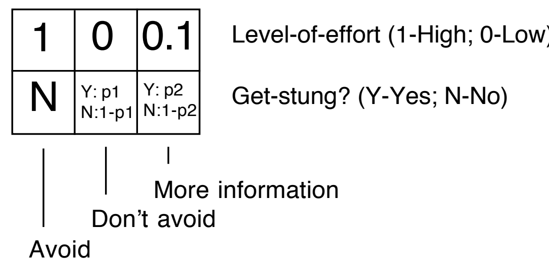

Figure 25: Preference grid incorporating the alternative of getting more information.

The difference between the expected value choosing an alternative after gathering **More-information** and the expected value of the next best current alternative (either avoiding or not avoiding) at a given point in time is the value of information for a that item of evidence. The cost of delaying an avoidance tactic in order to get more information is a function of the increase in probability that the insect will sting, given the length of time it has been in range plus the length of time it will take to answer the question, plus the effort involved in answering it. The objective **Level-of-effort** has been made more sensitive by allowing it to vary between 0 and 1, instead of taking on a simple binary high/low rating.

3.6 TEST CONSULTATION EXAMPLE

We are now ready to use the combined model to produce a recommendation for our situation. If **Avoid** were determined to be the preferred alternative, the initial recommendation would be the final one. However, if the preferred alternative were **Don't-Avoid,** the recommendation would be made but the system would continue to monitor changes to derived rating values over time until a new alternative becomes more highly valued. When the recommendation is **More-information**, the individual is queried about the insect's size, mood, or origin. A re-evaluation will take place to determine the best alternative in light of the new evidence. This pattern will continue until we have run out of questions or until the decision to avoid or not avoid the insect becomes final.

---

[55] Wellman and Heckerman (1987) have provided a more detailed discussion of this issue in which they contrast a "one-shot decision" approach to one in which options other than the primary decision can be explored. These options could include acquiring more knowledge or information, waiting for uncertain events to resolve, consulting other sources of expertise, or designing and performing experiments.

[56] We have assumed independence between the costs of answering individual questions.

3.6.1 Ordering questions using entropy of classification

During a consultation, the system may ask the person questions about one or more items of evidence in the information model. The better a question helps to discriminate between classes, the more informative it is. All other factors being equal, we would like to ask about the more informative questions first.

One heuristic for ordering questions during a consultation is to calculate the entropy of classification for particular constructs. Although not optimal, this “myopic” strategy of considering the information gain from one construct at a time has been widely used because of its robustness and efficiency (Quinlan, 1983, 1987). We compute the entropy of classification, $H(C|a_j)$, of each attribute not yet queried, a_j, by the expression:

$$H(C|a_j) = - \Sigma_{i=1 \text{ to } N}\, p(c_i|a_j) \log_2 p(c_i|a_j)$$

where $p(c_i|a_j)$ is the probability that the class is c_i when the attribute is a_j. The construct with the lowest value for H is the one that should be asked about first, since we assume that the question that gives the most information will be the one that would produce the highest expected value for the **More-information** alternative. We will compute H for size, mood, and origin based upon the example set in Figure 15:

SIZE

$H(\text{insect-type}|\text{small}) = - (2/4 \times \log_2(2/4)) - (2/4 \times \log_2(2/4)) = 1.0$

$H(\text{insect-type}|\text{big}) = - (2/4 \times \log_2(2/4)) - (2/4 \times \log_2(2/4)) = 1.0$

$H(\text{insect-type}|\text{size}) = (4/8 \times 1.0) + (4/8 \times 1.0) = 1.0$

MOOD

$H(\text{insect-type}|\text{good}) = - (2/5 \times \log_2(2/5)) - (3/5 \times \log_2(3/5)) = 0.97$

$H(\text{insect-type}|\text{bad}) = - (2/3 \times \log_2(2/3)) - (1/3 \times \log_2(1/3)) = 0.92$

$H(\text{insect-type}|\text{mood}) = (5/8 \times 0.972) + (3/8 \times 0.916) = 0.95$

ORIGIN

$H(\text{insect-type}|\text{east}) = - (2/2 \times \log_2(2/2)) - (0/2 \times \log_2(0/2)) = 0.0$

$H(\text{insect-type}|\text{west}) = - (2/6 \times \log_2(2/6)) - (4/6 \times \log_2(4/6)) = 0.92$

$H(\text{insect-type}|\text{origin}) = (2/8 \times 0.0) + (6/8 \times 0.916) = 0.69$ ***

In this case, the question about **Origin** would be asked first. If the individual observed the insect coming from the east, we would conclude that it wasn't a Stinger. Otherwise, we would determine the next best question by computing the entropy of classification for **Size** and **Mood** given that **Origin** is west:

SIZE

$H(\text{insect-type}|\text{west, small}) = - (0/2 \times \log_2(0/2)) - (2/2 \times \log_2(2/2)) = 0.0$

$H(\text{insect-type}|\text{west, big}) = - (2/4 \times \log_2(2/4)) - (2/4 \times \log_2(2/4)) = 1.0$

$H(\text{insect-type}|\text{west, size}) = (2/6 \times 0.0) + (4/6 \times 1.0) = 0.67$ ***

MOOD

$H(\text{insect-type}|\text{west, good}) = - (2/5 \times \log_2(2/5)) - (3/5 \times \log_2(3/5)) = 0.97$

$H(\text{insect-type}|\text{west, bad}) = - (0/1 \times \log_2(0/1)) - (1/1 \times \log_2(1/1)) = 0.0$

$H(\text{insect-type}|\text{west, mood}) = (5/6 \times 0.972) + (1/6 \times 0.0) = 0.81$

Note that even though **Mood** was a “better” construct than size at the outset, **Size** given **Origin** is better than **Mood** given **Origin**.

Each time a question is asked, new values for expected information are computed. Our goal is not to create a static decision tree, but only to order questions dynamically in terms of their expected costs and benefits.

3.6.2 Expected value of the alternatives over time

Figure 26 shows the expected value of the three alternatives over time. From the figure we see that the success of the avoidance tactic does not change over time (we assume that efforts to avoid are instantaneous and always successful), while the risk of getting stung while delaying a response increases over time. No action should be taken when the insect first flies within range, but if it hasn't flown off after 7 seconds, one should get information about its origin[57].

[57] To simplify the presentation of figure 26, we have smoothed the lines for each alternative, giving the illusion of continuity for what was essentially a discrete simulation.

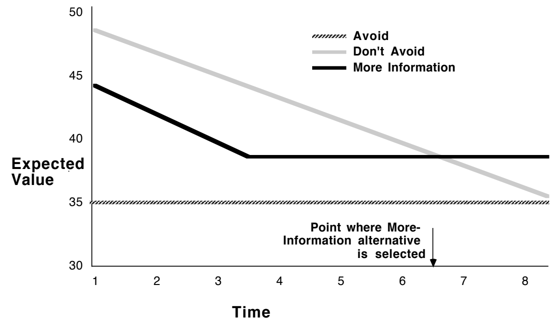

Figure 26: Expected value of the three alternatives over time.

If the individual observes that the insect has come from the east, it is certain that it isn't a stinger. The system can make a final recommendation of **Don't-Avoid**. However, if it has come from the west, the new information is incorporated and the preference model is re-evaluated. On the basis of that re-evaluation, a final recommendation for **Avoid** is made, because neither of the other two alternatives has a higher expected value, and both will continue to diminish in value with time.

The fact that only one of the questions is used might lead one to wonder whether there is any value in the other two potential items of evidence. In fact, there is often value in preserving redundancy in knowledge based systems. Imagine, for example, that the individual was unable to provide information about **Origin** but *was* able to say something about the insect's **Mood**. Figure 27 shows what would happen in this case: the question about **Mood** will be asked, although a little later than the question on **Origin** would have been.

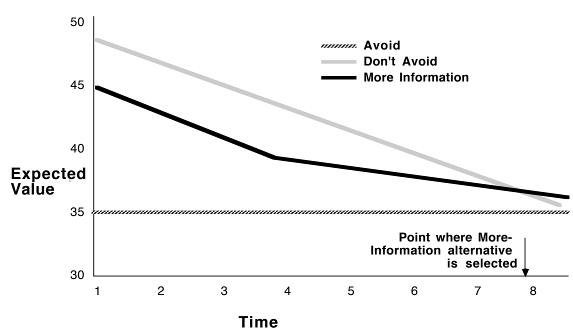

Figure 27: Expected value of the three alternatives if information about origin were unavailable.

4 DISCUSSION

4.1 VALUE OF EXPLICIT PREFERENCE MODELING

Through the example we have seen how the explicit modeling of preferences can be used throughout the knowledge acquisition and decision making processes.

Preferences were assessed and used to guide the selection of alternatives. Unique tradeoffs for a given situation were entered by simple assignment of ratings and weights in the preference grid. These could be easily adjusted when circumstances changed or to accommodate preferences of a different person. The approach allowed the incorporation of additional techniques for more complex preferences or when assessment of risk attitude or time preference is desirable. We also showed how recommendations for alternatives could be made in terms of meaningful units of value, taking both tangible and intangible objectives into account.

Most importantly, the example illustrated how preferences could be used to concentrate efforts during the knowledge acquisition process on the most important variables. The effect of the probability of getting stung on the decision was evaluated through sensitivity analysis. The value of obtaining additional information about the probability of getting stung was ascertained and used as an indicator that additional refinement of this variable would be worthwhile. The value of control for getting stung was also determined. This provided an upper bound on the value of generating and implementing a new alternative to eliminate uncertainty in a particular variable.

The ability of a knowledge-based system to reason about the relative importance of various components in a model is of central importance in light of the increasing volume and complexity of information that individuals are trying to acquire and manipulate via computer. It used to be that computing resources were so scarce and the bandwidth of human-computer interaction was so low one necessarily did all one could to *increase* access to information and make the barriers between the person and the world being modeled more permeable (Nelson, 1980). Now the amount of data that can be manipulated is so potentially overwhelming and the bandwidth sufficiently high that we need to be seriously concerned about actively, selectively keeping the most relevant and valuable information at the forefront of the interaction. To use an analogy, while advances in computing power and user interface have allowed us turning up the *brightness* of the available model image, we now need to to find effective means of increasing the *contrast* to keep the image intelligible. The use of preferences to highlight sensitive variables and the application of personal-construct-based methods to focus attention on similarities and differences are two such means.

In addition to what we have learned about the importance of preferences, our work affirms the worth of two other significant realizations that have begun to inform the efforts of knowledge acquisition researchers. These are:

- the constructive nature of model building;
- insight as the most important product of knowledge acquisition.

There are, however, many difficult, unresolved issues. Perhaps the most troubling ones are those that relate not to the effectiveness of any particular technique, but which question the practical and theoretical limits of any systematic methodology for knowledge acquisition or decision assistance. These include:

- the framing problem;
- the dangers of myopia in value-driven approaches.

Each of these will be discussed briefly below.

4.2 CONSTRUCTIVE NATURE OF MODEL BUILDING

Earlier conceptions of knowledge acquisition as "expertise transfer" (Clancey, 1979; Boose, 1984) have begun to give way to accounts that emphasize the constructive nature of model building: one does not go about extracting pre-existing models from the heads of experts; models are a product of creative design within the constraints of a given representation (Winograd & Flores, 1987). The following statement by Shafer and Tversky (1985) about the tasks of probability assessment applies equally well to the whole process of modeling:

> "Probability judgment is a process of construction rather than elicitation. People may begin a task of probability judgment with some beliefs already formulated. But the process of judgment, when successful, gives greater content and structure to these beliefs and tends to render initial beliefs obsolete. It is useful, in this respect, to draw an analogy between probability and affective notions such as love and loyalty. A declaration of love is not simply a report on a person's emotions. It is also a part of a process whereby an intellectual and emotional commitment is created; so too with probability."

The use of different methods in *Aquinas* for the formulation, analysis, and graphical display of problem solving models allows experts multiple perspectives on their problem. Robustness of the model can be tested through the use of several convergent procedures on the same parts of the problem. The workbench architecture of *Aquinas* allows for the easy incorporation of additional methods and representations.

4.3 INSIGHT AS THE MOST IMPORTANT PRODUCT OF KNOWLEDGE ACQUISITION

In the constructive view of problem solving, the model is not a *picture* of the decision but a *device for the attainment or formulation of knowledge* about it (Kaplan, 1963). The single most important product of knowledge acquisition may not be the model that is produced, nor the answers it generates, but rather the insight generated through the process of articulating, structuring, and critically evaluating one's expertise (Barr, 1985; Howard, 1980b; Moore & Agogino, 1987). The value of a particular knowledge acquisition effort will derive not so much from a final "correct" representation of the decision problem, as from our success in framing of the activity as a self-correcting enterprise that can subject any part of the model to critical scrutiny — including background assumptions. The critical question for knowledge engineers is not "How do you know the model is correct?" (every model is an incorrect simplification) but "How can the model be built so as to expose beliefs, conjectures, and biases — whether or not they are justifiable — to maximum criticism, in order to counteract and eliminate as much error as possible?" (Weimer, 1979).

Agogino and Moore (1987) have pointed out the importance of separating replication of performance from duplication of procedure — "at best, duplication is unlikely to result in performance improvement. Rather than implement actions emulating expert's problem solving actions, we want to use the expert's judgement to construct the model and to evaluate the model's performance." While the descriptive approaches typically used in the building of knowledge-based systems often attempt to duplicate procedure and sometimes are able to replicate expert performance, a normative approach can best realize the potential of actually *improving* the performance of the expert through the generation of insight based on the consistent application of a rational decision rule to which the individual subscribes (Henrion & Cooley, 1987).

4.4 THE FRAMING PROBLEM

Winograd and Flores (1987) have criticized a general bias inherent in techniques such as decision analysis that formulate situations as one of choosing between alternatives. They point out that reframing may create *new* alternatives or *dissolve* rather than *solve* the problem. It is not true that action is always preceded by consideration of alternatives: sometimes we simply commit ourselves, with the exclusion of other possibilities[58]. Rather than talking in terms of *choice,* they prefer to talk about a decision maker going from a state of *irresolution* to *resolution*. Von Winterfeldt and Edwards (1986) give the following example of a framing problem:

> "[A] woman initially believed that she had an apartment selection problem. The real problem turned out to be much more complex and involved financial questions and family interactions. The problem could have been formulated, for example, as a problem about the woman's management of her relationship with her parents. Another problem formulation would have focused on the ways the woman managed her financial situation. After these deeper aspects of the 'apartment selection' problem were identified, they were incorporated into objectives, and apartments were maintained as the alternatives. It was clear, however, that each apartment represented a complex alternative for managing the subject's life style."

Sloan (1987) has written an insightful monograph which shows that committing to an alternative (or set of alternatives) before major lacunae in self-understanding have been eliminated cuts short the process of understanding the roots of the dilemma that motivated the decision in the first place:

> "Issues preventing choice, or forcing impulsive choice, metaphorically express conflicts which are experienced in many spheres of life and repeatedly over the life course. Often a decision will even be framed in a manner which permits the active repetition of strategies which have failed previously, because to overcome the underlying conflict would be, in some sense, more painful." (see also, Corbin, 1980).

Our current approach is still far from satisfactory. We would like to make the process simpler and somewhat closer to the ways in which people ordinarily think about problems. We would like the process of knowledge acquisition to feel less like forcing a (good) methodology on people and more like accenting the natural discourse by giving good advice. In particular, there is a need for greater understanding of the place of alternatives in decision making (Bradshaw & Boose, 1988).

4.5 DANGERS OF MYOPIA IN VALUE-DRIVEN APPROACHES

Bateson (1972) has pointed out some of the most subtle and insidious dangers of decision making methodologies that are explicitly goal driven. He observes that this very purposiveness, like the purposiveness of consciousness itself, makes such approaches vulnerable to misuse by those who consider only what is immediately valuable or expedient when making decisions:

> "Consciousness… is organized in terms of purpose. It is a short-cut device to enable you to get quickly at what you want; not to act with maximum wisdom in order to live, but to follow the shortest logical or causal path to get what you next want, which may be dinner; it may be a Beethoven sonata; it may be sex. Above all, it may be money or power…
>
> If you look at the real situations in our world where the systemic nature of the world has been ignored in favor of purpose or common sense… [you will discover that] the terrible thing about such situations is that inevitably they shorten the time span of all planning. Emergency is present or only just around the corner; and long-term wisdom must therefore be sacrificed to expediency, even though there is a dim awareness that expediency will never give a long-term solution…"

---

[58] This is true, for example, when there is a dominating alternative such as the one expressed in the bumper sticker: "The worst day fishing is better than the best day working".

Von Winterfeldt and Edwards (1986) cite as a case in point the decision by Ford Motor Company to place the Pinto gas tank behind the rear axle:

> "The decision was based on a probabilistic cost-benefit analysis in which the chances of fire in rear-end collisions were traded off against the costs of lives lost in these fires. As it turned out, the expected dollar value of lives saved by placing the tank in front of the rear axle was smaller than the cost of the tank relocation. Ford therefore chose not to relocate it.
>
> The problem with this structure is its myopia, both in foreseeing possible consequences and in identifying relevant dimensions of value. It neglected the possible negative publicity that would result from frequent fires caused by rear-end collisions and it did not consider the possibility of punitive damages in liability suits…"

Beyond the narrow considerations of costs and benefits, however, one is always led to wonder about the extent to which reliance on methodologies of this sort tend to lessen feelings of responsibility for the consequences of our actions. Bateson correctly observes that a rational approach to decision making may keep us from lunacy, but not from sin. At the most important levels of strategic decision making, the question is often not what is the best thing to do within the rules of the game as they are at the moment, but rather making the rules explicit and determining how we can get away from the rules within which we have been operating and look at things from a wider, wiser perspective[59].

Among the corrective factors that serve to restore wisdom to an overly purposive and short-sighted perspectives he mentions: love (Buber's I-Thou as opposed to I-It relationships); the arts, poetry, music, and the humanities; contact between ourselves and animals and the natural world in general; and religion.

5 SUMMARY: THE COMPLEMENTARITY OF DECISION THEORY AND PERSONAL CONSTRUCT THEORY

The progress in *Aquinas* parallels the recent direction of other researchers who realize the value of combining insights from well-established disciplines such as decision analysis and newer methodologies derived from knowledge acquisition research (e.g., Henrion & Cooley, 1987; Holtzman, 1989; Horvitz *et al.,* 1988; Keeney, 1986a; Langlotz, Shortliffe, & Fagan, 1986; Moore & Agogino, 1987; Wellman, 1986). This is a sign that the field is reaching a new level of maturity. Those who have been primarily interested in quantitative problem solving techniques are discovering the central role of problem structuring and representation, and the advantages of heuristic and qualitative approaches for some situations. Those who have previously focused on qualitative and heuristic approaches are finding that complex problems cannot be effectively formulated or resolved without some reliance on quantitative methods.

More specifically, it is interesting to look at the origins and evolution of decision-theory-based and personal-construct-theory-based methodologies for model formulation and evaluation (Figure 28)[60]. Early research in decision theory was primarily concerned with the discovery of useful principles for the consistent use of probabilities and utilities in the evaluation of alternatives (Raiffa & Schlaifer, 1961). While the investigation of these principles constituted

59 On this point, Bateson's (1972) views are worth quoting verbatim: "… a peculiar sociological phenomenon has arisen in the last one hundred years which perhaps threatens to isolate conscious purpose from many corrective processes which might come out of less conscious parts of the mind. The social scene is nowadays characterized by the existence of a large number of self-maximizing entities which, in law, have something like the status of 'persons'—trusts, companies, political parties, unions, commercial and financial agencies, nations, and the like. In biological fact, these entities are precisely *not* persons and are not even aggregates of whole persons. The are aggregates of *parts* of persons. When Mr. Smith enters the board room of his company, he is expected to limit his thinking narrowly to the specific purposes of the company or to those of that part of the company which he 'represents.' Mercifully, it is not entirely possible for him to do this and some company decisions are influenced by considerations which spring from wider and wiser parts of the mind. But ideally, Mr. Smith is expected to act as a pure, uncorrected consciousness—a dehumanized creature."

60 To simplify the discussion, we include methods for model analysis and appraisal in the same dimension as evaluation.

a major breakthrough in the understanding of rational approaches to problem-solving, insights about problem *formulation* really began to come as the field of decision analysis was defined and several key figures began to tackle some of the more difficult problems of practical application (Howard, 1966a; Raiffa, 1968; Keeney & Raiffa, 1976). Decision analysis made advances not only in the area of model formulation, but also developed other useful ideas for model evaluation and appraisal such as value of information and control (e.g., Howard, 1966b). Influence diagrams represented an important breakthrough for the representation of decision models (Howard & Matheson, 1980; Miller, Merkhofer, Howard, Matheson, and Rice, 1976) and recent developments have extended their usefulness as a structuring and communication device (e.g., Horvitz *et al.*, 1988; Howard, 1987; Pearl, 1986; Wiecha & Henrion, 1987). While influence diagrams are primarily a way for structuring information and alternatives, similar progress has been made in the structuring of preferences (e.g., Keeney, 1986b; Keeney & Raiffa, 1976; Saaty, 1980). Recent work in intelligent decision systems continues the effort by providing additional methods and active guidance in decision model formulation and evaluation (Bradshaw & Holtzman, 1987; Holtzman & Breese, 1986; Holtzman, 1989).

Personal construct theory, in contrast to decision theory, had its origins in a concern for how people learn to make useful distinctions about the world (Kelly, 1955) and was only peripherally concerned about numerical evaluation procedures. The rise of cognitive psychology and computer-based approaches to the simulation and assessment of personality led to a number of efforts to automate procedures based on personal construct theory, although early researchers succeeded only in tapping a small part of the ideas latent in the theory through manipulating simple repertory grid representations. The most comprehensive repertory grid tool of its time was PLANET. In PLANET, Gaines & Shaw (1981a) integrated repertory grid elicitation and analysis tools developed by other researchers and added important new ones. ETS (Boose 1984, 1986) was the first personal-construct-theory-based tool to to use repertory grids to create knowledge bases for knowledge-based systems. It incorporated many of the ideas of PLANET and added additional analysis tools, facilities for the transformation of knowledge in grids to rules, and an internal consultation engine. *Aquinas* extended the repertory grid representation to allow superordinate and subordinate constructs as envisioned by Kelly. Tools for grid analysis and evaluation became more sophisticated, allowing individuals to use *Aquinas* for problem solving as well as knowledge acquisition for some types of problems. Gaines (1977, 1987a, 1987b) proposed a mathematical account of the modeling process as part of knowledge acquisition and drew correspondences to personal construct theory. In the current paper, we amplify some of Gaines' ideas to suggest how personal construct theory and decision theory might begin to be integrated in an automated system.

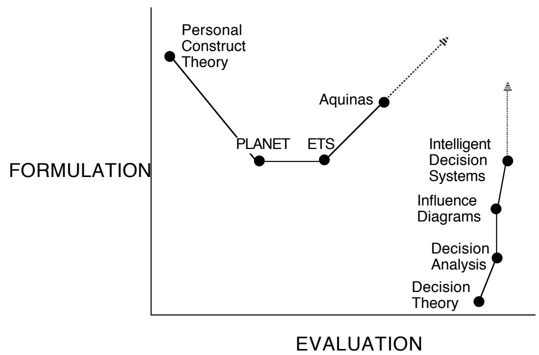

Figure 28: Progress in model formulation and evaluation techniques in personal-construct-theory-based and decision-theory-based methodologies.

In summary, the terms "decision analysis" and "knowledge acquisition" have both become misnomers because of the evolution of emphasis in both lines of research. To decision analysts, the analysis of decisions now takes a back seat to the problem of model formulation; at the same time, knowledge acquisition researchers have realized that the acquisition of knowledge cannot be undertaken effectively without a well-defined theory of model evaluation. One is reminded of Kipling's *Just So Story,* where to confuse their common enemy the painted jaguar, the hedgehog learned to swim like a turtle and the turtle to coil like a hedgehog. Eventually their contortions so changed their nature that they could no longer be distinguished. From that point on, they were called armadillos.

It is our belief that insights from personal construct theory and decision analysis can be effectively combined to model complex problems involving informational and preferential components. We look forward to the exciting developments ahead.

6 ACKNOWLEDGEMENTS

Thanks to Miroslav Benda, Kathleen Bradshaw, Beverly Clark, Stan Covington, Julia Jennings, Cathy Kitto, Sandra Marcus, Art Nagai, Peter Russo, Doug Schuler, Kish Sharma, David Shema, Lisle Tinglof-

Boose, and Bruce Wilson for their contributions and support. Special thanks to Sam Holtzman, Ronald Howard, and Jim Matheson who introduced us to the concepts of decision analysis. This work has also benefitted from conversations and helpful comments from Peter Cheeseman, Brian Gaines, David Heckerman, Max Henrion, Eric Horvitz, and Mildred Shaw. *Aquinas* was developed at the Knowledge Systems Laboratory, Advanced Technology Center, Boeing Computer Services in Seattle, Washington.

7 REFERENCES

ADDIS, T., GAINES, B.R. & BOOSE, J.H. (eds.) (1988). *Advances in Knowledge Acquisition for Knowledge-Based Systems*. Vol. 3: Results of the First European Workshop on Knowledge Acquisition for Knowledge-Based Systems. New York: Academic Press, in press.

ALLPORT, G. W. (1942). *The Use of Personal Documents in Psychological Science*. New York: Social Science Research Council, Bulletin 49.

BATESON, G. (1972). *Steps to an Ecology of Mind*. New York: Ballantine Books.

BATESON, G. (1979). *Mind and Nature: A Necessary Unity*. New York: E. P. Dutton.

BATESON, G. & BATESON, M. C. (1987). *Angels Fear*. New York: Macmillan.

BEACH, L. R. & MITCHELL, T. R. (1988). Image theory: Principles, goals, and plans in decision making. *Journal of Behavioral Decision Making,* in press.

BEIER, E. G. (1966). *The Silent Language of Psychotherapy*. Chicago: Aldine Publishing.

BEIER, E. G., & VALENS, E. G. (1975). *People-Reading*. New York: Stein & Day.

BEYTH-MAROM, R. & DEKEL, S. (1985). *An Elementary Approach to Thinking Under Uncertainty*. London: L. Erlbaum.

BHARATH, R. Information theory. *BYTE,* December, 291-298.

BOOSE, J. H. (1984). Personal construct theory and the transfer of human expertise. *Proceedings of the National Conference on Artificial Intelligence,* Austin, Texas.

BOOSE, J.H. (1985). A knowledge acquisition program for expert systems based on Personal Construct Psychology. *International Journal of Man-Machine Studies,* 23**,** 495-525.

BOOSE, J.H. (1986a). *Expertise Transfer for Expert System Design*. New York: Elsevier.

BOOSE, J.H. (1986b). Rapid acquisition and combination of knowledge from multiple experts in the same domain. *Future Computing Systems Journal,* **1**, 191-216.

BOOSE, J.H. (1987). Uses of repertory grid-centered knowledge acquisition tools for knowledge-based systems. *Proceedings of the Second AAAI Knowledge Acquisition for Knowledge-Based Systems Workshop,* Banff, Canada, October, 1987.

BOOSE, J.H. & BRADSHAW, J.M. (1987a). Expertise transfer and complex problems: Using *Aquinas* as a knowledge-acquisition workbench for knowledge-based systems. *International Journal of Man-Machine Studies,* 26**,** 3-28.

BOOSE, J.H. & BRADSHAW, J.M. (1987b). AQUINAS: A knowledge acquisition workbench for building knowledge-based systems. *Proceedings of the First European Workshop on Knowledge Acquisition for Knowledge-Based Systems,* Reading, England, September, 1987.

BOOSE, J. H., BRADSHAW, J. M., & SHEMA, D. B. (1988). Transforming repertory grids to shell-based knowledge-bases using Aquinas, a knowledge acquisition workbench. Proceedings of the AAAI Workshop on Integration of Knowledge Acquisition and Performance Systems, August 1988.

BOOSE, J.H. & GAINES, B.R. (eds.) (1988a). *Advances in Knowledge Acquisition for Knowledge-Based Systems*. Vol. 1: Interactive Tools. Results of the 1986 First AAAI Knowledge Acquisition for Knowledge-Based Systems Workshop. New York: Academic Press, in press.

BOOSE, J.H. & GAINES, B.R. (eds.) (1988b). *Advances in Knowledge Acquisition for Knowledge-Based Systems*. Vol. 4: Interactive Tools. Results of the 1987 Second AAAI Knowledge Acquisition for Knowledge-Based Systems Workshop. New York: Academic Press, in press.

BOULTON, D.M. & WALLACE, C.S. (1973). An information measure for hierarchic classification. *Computer Journal,* 16**,** 254-261.

BRADSHAW, J. M. & HOLTZMAN, S. (1987). Artificial intelligence and decision analysis. Presentation made to the Nineteenth Annual Meeting of the Society for Information Management, Seattle: October 17, 1987.

BRADSHAW, J. M. & BOOSE, J.H. (1987). Decision analysis techniques for knowledge acquisition: Combining information and preference models using *Aquinas*. *Proceedings of the Second AAAI Knowledge Acquisition for Knowledge-Based Systems Workshop,* Banff, Canada, October, 1987.

BRADSHAW, J. M. & BOOSE, J. H. (1988). How to do with grids what people say you can't. To appear in the *Proceedings of the Third AAAI Knowledge Acquisition for Knowledge-Based Systems Workshop,* Banff, Canada, November, 1988.

BRAND, S. (1987). *The Media Lab: Inventing the Future at MIT*. New York: Viking.

BROWN, B. L., WARNER, C. T., & WILLIAMS, R. N. (1987). Vocal paralanguage without unconscious processes. In A. W. Siegman & S. Feldstein (eds.) *Functions of Non-Verbal Behavior*. Hillsdale, N. J.: Erlbaum.

BROWN, B. L., & BRADSHAW, J. M. (1985). A psychology of vocal patterns. In H. Giles and R. N. St. Clair, Language and the Paradigms of Social Psychology. Hillsdale, N. J.: Lawrence Erlbaum Associates.

BRUNER, J. S., GOODNOW, J. J. & AUSTIN, G. A. (1956). *A Study of Thinking*. New York: Wiley.

CAMPBELL, J. (1982). *Grammatical Man*. New York: Simon and Schuster.

CARROLL, B. (1987). Expert systems for clinical diagnosis: Are they worth the effort? *Behavioral Science,* 32, 274-292.

CHEESEMAN, P. (1983). A method of computing generalized Bayesian probability values for expert systems. *Proceedings of the Eighth International Joint Conference on Artificial Intelligence,* Karlsruhe, West Germany, August 1983.

CHEESEMAN, P. (1984). Learning of expert systems from data. *Proceedings of the Workshop on Principles of Knowledge-Based Systems,* Denver, December 1984.

CHEESEMAN, P. (1985). In defense of probability. *Proceedings of the 9th International Joint Conference on Artificial Intelligence,* Los Angeles, August 1985.

CHEESEMAN, P. (1986). Probabilistic versus fuzzy reasoning. In L.N. Kanal and J. F. Lemmer (eds.), *Uncertainty in Artificial Intelligence*. Amsterdam: Elsevier.

CHEESEMAN, P. (1987). Bayesian prediction for artificial intelligence. *Proceedings of the 3rd AAAI Workshop on Uncertainty in Artificial Intelligence,* Seattle, July 10-12, 1987.

CHEESEMAN, P., KELLY, J., SELF, M. & STUTZ, J. (1987). Automatic Bayesian induction of classes. *Proceedings of the Second Annual AI Research Forum,* Palo Alto, California: NASA Ames Research Center, November 16-18, 1987.

CHEESEMAN, P., KELLY, J., SELF, M., STUTZ, J., TAYLOR, W. & FREEMAN, D. (1988). AutoClass: A Bayesian classification system. To appear in Proceedings of the 1988 AAAI Machine Learning Workshop.

CHEESEMAN, P., SELF, M., STUTZ, KELLY, J., J., TAYLOR, W. & FREEMAN, D. (1988). Bayesian classification. To appear in Proceedings of the 1988 AAAI Conference.

CHEN, P.P., (ed.) (1980). *Entity-Relationship approach to Systems Analysis and Design*. New York: North Holland.

CLANCEY, W. J. (1979). *Transfer of Rule-based Expertise Through a Tutorial Dialogue*. Ph.D. thesis, Stanford University, August, 1979.

CLANCEY, W. J. (1983). The epistemology of a rule-based expert system — a framework for explanation. *Artificial Intelligence,* 20**,** 215-251.

CLANCEY, W. J. (1984). Classification problem solving. *Proceedings of the National Conference on Artificial Intelligence,* Austin, Texas.

COHEN, B. & MURPHY, G. L. (1984). Models of concepts. *Cognitive Science,* 8, 27-58.

COLLINS, A. M. & QUILLIAN, M. R. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior,* 8, 240-247.

COOPER, W. S. (1987). Decision theory as a branch of evolutionary theory: A biological derivation of the Savage axioms. *Psychological Review,* 94(4), 395-411.

CORBIN, R. M. (1980). Decisions that might not get made. In T. S. Wallsten (ed.) *Cognitive Processes in Choice and Decision Behavior*. Hillsdale, New Jersey: Lawrence Erlbaum.

CRAGUN, B.J. & STEUDEL, H.J. (1987). A decision-table-based processor for checking completeness and consistency in rule-based expert systems. *International Journal of Man-Machine Studies,* 26, 633-648.

CRAIK, K. (1943). *The Nature of Explanation*. Cambridge, England: Cambridge University Press.

CURRIE, L. C. (1985). Choice decision and the anticipation of events. In G. Wright (ed.), *Behavioral Decision Making*. New York: Plenum Press.

CYERT, R. M. & MARCH, J. G. (1963). *A Behavioral Theory of the Firm*. Englewood Cliffs, New Jersey: Prentice-Hall.

DAVISSON, A. (1978). George Kelly and the American mind (or why has he been so obscure for so long in the U.S.A. and whence the new interest?). In F. Fransella (ed.), *Personal Construct Psychology 1977*. New York: Academic Press.

DAWES, R. M. & CORRIGAN, B. (1974). Linear models in decision making. *Psychological Bulletin,* 81(2), 95-106.

DE FINETTI, B. (1937). La prévision: ses lois logiques, ses sources subjectives. *Annales de l'Institut Henri Poincaré,* 7:1-68.

DIEDERICH, J., RUHMANN, I. & MAY, M. (1987). KRITON: A knowledge acquisition tool for expert systems. *International Journal of Man-Machine Studies,* **26**(1).

DIEDERICH, J., LINSTER, M., RUHMANN, I. & UTHMANN, T. (1987). A methodology for integrating knowledge acquisition techniques. *Proceedings of the First European Workshop on Knowledge Acquisition for Knowledge-Based Systems,* Reading, England, September, 1987.

DUDA, R.O. & REBOH, R. (1984). AI and decision making: The PROSPECTOR experience. In W. Reitman (ed.) *Artificial Intelligence Applications for Business*. Norwood, New Jersey: Ablex.

ECO, U. (1986). In praise of Thomas Aquinas. *Wilson Quarterly,* Fall 1986.

ESHELMAN, L., EHRET, D., MCDERMOTT, J. & TAN, M. (1987). MOLE: A tenacious knowledge acquisition tool. *International Journal of Man-Machine Studies,* **26,** 41-54.

EVANS, J. St. B. T. (1988). The knowledge elicitation problem: A psychological perspective. *Behaviour and Information Technology,* 7(2), 111-130.

FERGUSON, H. R. P. (1976). Preliminary to catastrophe theory in the behavioral sciences: How to make cusp proverbs. Provo, Utah: Department of Mathematics, Brigham Young University.

FISCHHOFF, B., SLOVIC, P. & LICHTENSTEIN, S. (1980). Knowing what you want: Measuring labile values. In T. S. Wallsten (ed.) *Cognitive Processes in Choice and Decision Behavior*. Hillsdale, New Jersey: Lawrence Erlbaum.

FOUCAULT, M. (1970). *The Order of Things: An Archaeology of the Human Sciences*. London:.

GAINES, B. R. (1977). System identification, approximation, and complexity. *International Journal of General Systems,* 3, 145-174.

GAINES, B. R. (1987a). An overview of knowledge acquisition and transfer. *International Journal of Man-Machine Studies,* **26,** 453-472.

GAINES, B. R. (1987b) How do experts acquire expertise? *Proceedings of the First European Workshop on Knowledge Acquisition for Knowledge-Based Systems,* Reading, England.

GAINES, B.R. & BOOSE, J.H. (eds.) (1988a). *Advances in Knowledge Acquisition for Knowledge-Based Systems*. Vol. 2: Cognition, Expertise, and Knowledge Structure. Results of the 1986 First AAAI Knowledge Acquisition for Knowledge-Based Systems Workshop. New York: Academic Press, in press.

GAINES, B.R. & BOOSE, J.H. (eds.) (1988b). *Advances in Knowledge Acquisition for Knowledge-*

Based Systems. Vol. 5: Cognition and Expertise. Results of the 1987 Second AAAI Knowledge Acquisition for Knowledge-Based Systems Workshop. New York: Academic Press, in press.

GAINES, B. R. & SHAW, M. L. G. (1981a). New directions in the analysis and interactive elicitation of personal construct systems. In Shaw, M. L. G., (ed.), *Recent Advances in Personal Construct Technology*. New York: Academic Press.

GAINES, B. R. & SHAW, M. L. G. (1981b). A programme for the development of a systems methodology of knowledge and action. Reckmeyer, W. J. (ed.) *General Systems Research and Design: Precursors and Futures*. Society for General Systems Research (January), 255-264.

GAINES, B. R. & SHAW, M. L. G. (1986a). Interactive elicitation of knowledge from experts. *Future Computing Systems,* 1(2).

GAINES, B. R. & SHAW, M. L. G. 1986b). Induction of inference rules for expert systems. *Fuzzy Sets and Systems*, 8(3), 315-328 (April).

GARG-JANARDAN, C. & SALVENDY, G. (1987). A conceptual framework for knowledge elicitation. *International Journal of Man-Machine Studies,* 27(1).

GIERE, R. N. (1988). *Explaining Science: A Cognitive Approach*. Chicago: University of Chicago Press.

HAYEK, F. (1952). *The Sensory Order*. Chicago: University of Chicago Press.

HECKERMAN, D. & HORVITZ, E. J. (1987). On the expressiveness of rule-based systems for reasoning with uncertainty. *Proceedings of AAAI,* vol. 1, pp. 121-126. Los Altos, California: Morgan Kaufmann.

HECKERMAN, D. & JIMISON, H. A perspective on confidence and its use in focusing attention during knowledge acquisition. *Proceedings of the 3rd AAAI Workshop on Uncertainty in Artificial Intelligence,* Seattle, July 10-12, 1987.

HENRION, M. (1987). Practical issues in constructing a Bayes' belief network. *Proceedings of the 3rd AAAI Workshop on Uncertainty,* Seattle, July 10-12, 1987.

HENRION, M. (1988). Towards efficient probabilistic diagnosis in multiply connected belief networks. *Proceedings of the Conference on Influence Diagrams for Decision Analysis, Inference, and Prediction*, May 9-11, 1988, Berkeley California, University of California at Berkeley.

HENRION, M. & COOLEY, D. R. (1987). An experimental comparison of knowledge engineering for expert systems and decision analysis. *Proceedings of the National Conference on Artificial Intelligence,* Seattle, Washington.

HINK, R. F. & WOODS, D. L. (1987). How humans process uncertain knowledge: An introduction for knowledge engineers. *AI Magazine,* Fall 1987, 41-53.

HINKLE, D. N. (1970). The game of personal constructs. In D. Bannister (ed.) *Perspectives in Personal Construct Theory*. London: Academic Press.

HOLLAND, J. H., HOLYOAK, K. J., NISBETT, R. E. & THAGARD, P. R. (1986). *Induction: Processes of Inference, Learning, and Discovery*. Cambridge, Massachusetts: The MIT Press.

HOLTZMAN, S. H. & BREESE, J. (1986). Exact reasoning about uncertainty: On the design of expert systems for decision support. In L.N. Kanal and J. F. Lemmer (eds.), *Uncertainty in Artificial Intelligence*. Amsterdam: Elsevier.

HOLTZMAN, S. H. (1989). *Intelligent Decision Systems*. Reading, Massachusetts: Addison-Wesley. In press.

HORVITZ, E. J. (1987a). *Inference Under Varying Resource Limitations*. Knowledge Systems Lab Technical Report KSL-87-16, Stanford University, Stanford, California, February, 1987.

HORVITZ, E. J. (1987b). Reasoning about beliefs and actions under computational resource constraints. *Proceedings of the Third AAAI Workshop on Uncertainty in Artificial Intelligence,* Seattle, July 10-12, 1987.

HORVITZ, E. J. (1987c). Problem-solving design: Reasoning about computational value, tradeoffs, and resources. *Proceedings of the Second Annual AI Research Forum,* Palo Alto, California: NASA Ames Research Center, November 16-18, 1987.

HORVITZ, E. J., BREESE, J. S. & HENRION, M. (1988). Decision theory in expert systems and artificial intelligence. *Journal of Approximate Reasoning,* in press.

HOWARD, R. A. (1966a). Decision analysis: Applied decision theory. In D. B. Hertz and J. Melese (eds.), *Proceedings of the Fourth International Conference on Operational Research*. Reprinted in Howard & Matheson, 1984.

HOWARD, R. A. (1966b). Information value theory. *IEEE Transactions on Systems Science and Cybernetics,* SSC-2, August, 22-26. Reprinted in Howard & Matheson, 1984.

HOWARD, R. A. (1980a). On making life and death decisions. R. C. Schwing & W. A. Albers, Jr. (eds.), *Societal Risk Assessment*. Plenum Publishing. Reprinted in Howard & Matheson, 1984.

HOWARD, R. A. (1980b). An assessment of decision analysis. *Operations Research,* **28,** 4-27.

HOWARD, R. A. (1984). On fates comparable to death. *Management Science,* 30(4). Reprinted in Howard & Matheson, 1984.

HOWARD, R. A. (1987). Knowledge maps. Department of Engineering-Economic Systems, Stanford University, Stanford, California. To be published in *Management Science*.

HOWARD, R. A. (1988). From influence to relevance to knowledge. *Proceedings of the Conference on Influence Diagrams for Decision Analysis, Inference, and Prediction*, May 9-11, 1988, Berkeley California, University of California at Berkeley.

HOWARD, R. A. & MATHESON. J. E. (1980). Influence Diagrams. Reprinted in Howard & Matheson, 1984.

HOWARD, R. A. & MATHESON, J. E. (eds.) (1984). *Readings on the Principles and Applications of Decision Analysis*. Menlo Park, California: Strategic Decisions Group.

HUNTER, D. (1986). Uncertain reasoning using maximum entropy inference. *Proceedings of the 1st AAAI Workshop on Uncertainty in Artificial Intelligence*.

HUSAIN, M. (1983). To what can one apply a construct? In J. Adams-Webber and J. C. Mancuso (eds.), *Applications of Personal Construct Theory*. New York: Academic Press.

JANKOWICZ, A. D. (1987). Whatever became of George Kelly? *American Psychologist,* 42(5), 481-487.

JOHNSON-LAIRD, P. N. (1988). *The Computer and the Mind*. Cambridge, Massachusetts: Harvard University Press.

JOHNSON-LAIRD, P. N., HERRMAN, D. J., & CHAFFIN, R. (1984). Only connections: a critique of semantic networks. *Psychological Bulletin*, 96, 292-315.

KAHNEMAN, D., SLOVIC, P. & TVERSKY, A. (1982). *Judgment Under Uncertainty: Heuristics and Biases*. Cambridge: Cambridge University Press.

KAPLAN, A. (1963). *The Conduct of Inquiry*. New York: Harper and Row.

KEENEY, R. L. (1986a). *Value-driven Expert Systems For Decision Support*. Los Angeles: Systems Science Department, University of Southern California.

KEENEY, R. L. (1986b). *Insights From Value Assessments*. Los Angeles: Systems Science Department, University of Southern California.

KEENEY, R. L. & RAIFFA, H. (1976). *Decisions With Multiple Objectives: Preferences and Value Tradeoffs*. New York: Wiley.

KELLY, G. A. (1955). *The Psychology of Personal Constructs*. 2 Volumes. New York: Norton.

KELLY, G. A. (1957). Hostility. In B. Maher (ed.) *Clinical Psychology and Personality: The Selected Papers of George Kelly*. New York: John Wiley.

KELLY, G. A. (1961). A mathematical approach to psychology. In B. Maher (ed.) *Clinical Psychology and Personality: The Selected Papers of George Kelly*. New York: John Wiley.

KELLY, G. A. (1963). Discussion: Aldous, the personable computer. In S. S. Tomkins & S. Messick (eds.) *Computer Simulation of Personality*. New York: John Wiley.

KELLY, G. A. (1966). Ontological acceleration. In B. Maher (ed.) *Clinical Psychology and Personality: The Selected Papers of George Kelly*. New York: John Wiley.

KIM, J. & PEARL, J. (1983). A computational model for combined causal and diagnostic reasoning in inference systems. *Proceedings of the 8th International Joint Conference on Artificial Intelligence,* Karlsruhe, West Germany, August 1983.

KING, A. H. (1979). *Mechanization*. Flea Market Lecture. Provo, Utah: Brigham Young University, September 13, 1979.

KITTO, C. & BOOSE, J. H. (1987). Heuristics for expertise transfer: The automatic management of complex knowledge acquisition dialogs. *International Journal of Man-Machine Studies,* 26, 183-202.

KLIR, G. J. (1976). Identification of generative structures in empirical data. *International Journal of General Systems,* 3, 89-104.

KLIR, G. J. (1985). *Architecture of Systems Problem Solving*. New York: Plenum Press.

LANGLOTZ, C. P., SHORTLIFFE, E. H. & FAGAN, L. M. (1986). Using decision theory to justify heuristics. *Proceedings of the National Conference on Artificial Intelligence,* Philadelphia.

LEVI, I. (1986). *Hard Choices: Decision Making Under Unresolved Conflict*. Cambridge: Cambridge University Press.

MALVESTUTO, F. M. (1987). Answering queries in categorical databases. *Proceedings of the 6th ACM-SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems*. San Diego, March 23-25, 1987.

MARCH, J. G. (1978). Bounded rationality, ambiguity, and the engineering of choice. *The Bell Journal of Economics,* 9(2), 587-608.

MARCUS, S. (1987). Taking backtracking with a grain of SALT. *International Journal of Man-Machine Studies*, 26, 383-398.

MATHESON, J. E. (1968). The economic value of analysis and computation. *IEEE Transactions on Systems Science and Cybernetics,* SSC-4(3), September. Reprinted in Howard & Matheson, 1984.

MATHESON, J. E. (1988). Using influence diagrams to value information and control. *Proceedings of the Conference on Influence Diagrams for Decision Analysis, Inference, and Prediction,* May 9-11, 1988, Berkeley California, University of California at Berkeley.

MAVROVOUNIOTIS, M. L. & STEPHANOPOULOS, G. (1987). Reasoning with orders of magnitude and approximate relations. *Proceedings of the National Conference on Artificial Intelligence,* Seattle.

MCCOY, M. M. (1981). Positive and negative emotion: A personal construct theory interpretation. In H. Bonarius, R. Holland, & S. Rosenberg (eds.), *Personal Construct Psychology: Recent Advances in Theory and Practice*. London: Macmillan.

MEDIN, D. L. & SMITH, E. E. (1984). Concepts and concept formation. *Annual Review of Psychology,* 35, 113-138.

MELGES, F. T. (1982). *Time and the Inner Future: A Temporal Approach to Psychiatric Disorders*. New York: John Wiley.

MICHALSKI, R. S. & CHILAUSKY, R. L. (1980). Knowledge acquisition by encoding expert rules versus computer induction from examples — A case involving soybean pathology. *International Journal of Man-Machine Studies,* 12, 63-87.

MILLER, A. C., MERKHOFER, M. W., HOWARD, R. A. & RICE, T. R. (1976). Development of automated aids for decision analysis. SRI Technical Report. Menlo Park, California: Stanford Research Institute.

MOORE, E. A. & AGOGINO, A. M. (1987). INFORM: An architecture for expert-directed knowledge acquisition. *International Journal of Man-Machine Studies*, 26, 213-230.

NELSON, T. (1980). Interactive systems and the design of virtuality. Parts one and two. *Creative Computing,* November, 1980, 56-62; December 1980, 95-106.

PEARL, J. (1985). How to do with probabilities what people say you can't. *Proceedings of the 2nd Annual Conference on Artificial Intelligence Applications,* Miami, December 11-13, 1985.

PEARL, J. (1986). Fusion, propagation, and structuring in belief networks. *Artificial Intelligence,* 29, 241-288.

PEARL, J. (1987). *Uncertainty Management in Expert Systems*. Tutorial #10 Notes, Tenth

International Conference on Artificial Intelligence, Milan, August 23-28, 1987.

PEARL, J., GEIGER, P. & VERMA, T. (1988). The logic of influence diagrams. *Proceedings of the Conference on Influence Diagrams for Decision Analysis, Inference, and Prediction*. May 9-11, 1988, Berkeley California, University of California at Berkeley.

PEARL, J. & VERMA, T. (1987). The logic of representing dependencies by directed graphs. *Proceedings of the National Conference on Artificial Intelligence,* Seattle.

PIAGET, J. (1981). *Intelligence and Affectivity: Their Relationship During Child Development*. Palo Alto, California: Annual Reviews.

PRIBRAM, K. (1971). *Languages of the Brain*. Englewood Cliffs, New Jersey: Prentice-Hall.

QUINLAN, J. R. (1983). Learning efficient classification procedures and their application to chess endgames. In R. S. Michalski, J. G. Carbonell, & T. M. Mitchell (eds.) *Machine Learning: An Artificial Intelligence Approach,* Palo Alto, California: Tioga.

QUINLAN, J. R. (1986). The effect of noise on concept learning. In R. S. Michalski, J. G. Carbonell, & T. M. Mitchell (eds.) *Machine Learning: Volume 2*. Los Altos, California: Morgan Kaufmann.

RAIFFA, H. (1968). *Decision Analysis: Introductory Lectures On Choices Under Uncertainty*. Reading, Massachusetts: Addison-Wesley.

RAIFFA, H. & SCHLAIFER, R. (1961). *Applied Statistical Decision Theory*. Cambridge, Massachusetts: Harvard University Press.

RAIMAN, O. (1986). Order of magnitude reasoning. *Proceedings of the National Conference on Artificial Intelligence,* Philadelphia.

ROSCH, E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General,* 104, 192-223.

ROSCH, E. & MERVIS, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology,* 7, 573-605.

RUBENSTEING, M. F. (1975). *Patterns of Problem Solving*. Englewood Cliffs, N. J.: Prentice-Hall.

RUCKER, R. (1984). *The Fourth Dimension*. Boston: Houghton Mifflin.

SAATY, T. L. (1980). *The Analytic Hierarchy Process*. New York: McGraw-Hill.

SAVAGE. L. J. (1954). *The Foundations of Statistics*. New York: Wiley.

SAVAGE, L. J. (1972). *The Foundations of Statistics*. 2nd Edition, revised. New York: Dover.

SCHEIER, M. F. & CARVER, C. S. (1987). Dispositional optimism and physical well-being: The influence of generalized outcome expectancies on health. *Journal of Personality*, 55:2 (June).

SCHULER, D., BOOSE, J., RUSSO, P., and BRADSHAW, J. (1988). Collaborative Evaluation Using Personal Construct Techniques. Manuscript submitted for publication, 1988.

SHACHTER, R. D. & HECKERMAN, D. E. (1987). Thinking backward for knowledge acquisition. *AI Magazine,* Fall, 55-61.

SHAFER, G. & TVERSKY, A. (1985). Languages and designs for probability judgment. *Cognitive Science,* 9, 309-339.

SHANNON, C. E. & WEAVER, W. (1964). *The Mathematical Theory of Communication*. Urbana, Illinois: University of Illinois Press.

SHASTRI, L. (1985). *Evidential Reasoning in Semantic Networks: A Formal Theory and its Parallel Implementation*. Technical Report 166, Rochester, New York: University of Rochester, September 1985.

SHASTRI, L. (1987). A connectionist encoding of semantic networks. *Proceedings of the Cognitive Science Society,* Seattle.

SHASTRI, L. & FELDMAN, J. (1985). Evidential reasoning in semantic networks: A formal theory. *Proceedings of the Ninth International Joint Conference on Artificial Intelligence,* Los Angeles.

SHAW, M. L. G., BRADSHAW, J. M., GAINES, B. R., and BOOSE, J. H. (1988). Rapid prototyping techniques for expert systems. *Proceedings of the Fourth IEEE Conference on Artificial Intelligence Applications,* San Diego, CA, March 1988.

SHAW, M. L. G. & GAINES, B. R. (1986). Cognitive foundations of knowledge-based systems. Calgary, Alberta, Canada: University of Calgary.

SHAW, M. L. G. & GAINES, B. R. (1987). Techniques for knowledge acquisition and transfer. *International Journal of Man-Machine Studies,* 26.

SHAW, R. & MCINTYRE, M. (1974). Algoristic foundations to cognitive psychology. In W. B. Weimer and D. S. Palermo (eds.) Cognition and the Symbolic Processes. New York: Erlbaum.

SHEMA, D. & BOOSE, J.H. (1987). Refining problem-solving knowledge in repertory grids using a consultation mechanism. *Proceedings of the 2nd AAAI Knowledge Acquisition for Knowledge-Based Systems Workshop,* Banff, Canada, October, 1987.

SIGEL, I. E. & HOLMGREN, A. (1983). A constructivist dialectic view of the development of the person: An update. In J. Adams-Webber & J. C. Mancuso (eds.) *Applications of Personal Construct Theory*. New York: Academic Press.

SLOAN, T. S. (1987). *Deciding: Self-Deception in Life Choices*. New York: Macmillan.

SPETZLER, C. S. & STAEL VON HOLSTEIN, C. A. S. (1975). Probability encoding in decision analysis. *Management Science,* 22, 340-358.

SPIEGELHALTER, D. J. (1986a). A statistical view of uncertainty in expert systems. In W. Gale, (ed.), *AI and Statistics*. Reading, Massachusetts: Addison-Wesley.

SPIEGELHALTER, D. J. (1986b). Probabilistic reasoning in predictive expert systems. In L. N. Kanal and J. F. Lemmer (eds.), *Uncertainty in Artificial Intelligence*. Amsterdam: Elsevier.

STAEL VON HOLSTEIN, C. A. S. & MATHESON, J. E. (1978). *A Manual for Encoding Probability Distributions*. SRI Project 7078. Menlo Park, California: SRI International.

STICH, S. P. & NISBETT, R. E. (1984). Expertise, justification and the psychology of inductive reasoning. In T. L. Haskell (ed.) *The Authority of Experts*. Bloomington, Indiana: Indiana University Press.

TUCKER, D. M. & WILLIAMSON, P. A. (1984). Asymmetric neural control systems in human self-regulation. *Psychological Review,* 91(2), April, 185-215.

TVERSKY, A., SATTATH, S., & SLOVIC, P. (1987). Contingent weighting in judgment and choice. Department of Psychology, Stanford University, Stanford, California.

VON BERTALANFFY, L. (1968). *General System Theory: Foundations, Development, Applications*. New York: George Braziller.

VON NEUMANN, J. & MORGENSTERN, O. (1953). *Theory of Games and Economic Behavior*. New York: Wiley.

VON WINTERFELDT, D. & EDWARDS, W. (1986). *Decision Analysis and Behavioral Research*. Cambridge: Cambridge University Press.

WALLSTEN, T. S. & BUDESCU, D. V. (1983). Encoding subjective probabilities: A psychological and psychometric review. *Management Science,* 29(2), February, 151-173.

WARNER, C. T. (1984). Anger and similar delusions. Presentation to the Utah Academy of Arts and Sciences, May, 1984.

WEIMER, W. (1975). The psychology of inference and expectation: Some preliminary remarks. In G. Maxwell and R. M. Anderson, Jr. (eds.) *Minnesota Studies in the Philosophy of Science,* vol. 6, pp. 430-486. Minneapolis: University of Minnesota Press.

WEIMER, W. (1979). *Notes on the Methodology of Scientific Research*. Hillsdale, New Jersey: Erlbaum.

WEINBERG, G. M. (1975). *An Introduction to General Systems Thinking*. New York: Wiley.

WELLMAN, M. (1986). Reasoning about assumptions underlying mathematical models. In J. S. Kowalik (ed.), *Coupling Symbolic and Numeric Computing*. Amsterdam: Elsevier.

WELLMAN, M. & HECKERMAN, D. (1987). The role of calculi in uncertain reasoning. In *Proceedings of the 3rd AAAI Workshop on Uncertainty in Artificial Intelligence,* Seattle, July 10-12, 1987.

WIECHA, C. & HENRION, M. (1987). *Demaps: A Graphical Decision Support System.* Pittsburgh: Department of Engineering and Public Policy, Carnegie-Mellon University.

WINOGRAD, T. & FLORES, F. (1987). *Understanding Computers and Cognition*. Reading, Massachusetts: Addison-Wesley.

WISE, B. P. (1987). Satisfaction of assumptions is a weak predictor of performance. *Proceedings of the 3rd AAAI Workshop on Uncertainty in Artificial Intelligence*. Seattle, July 10-12, 1987.

WISE, B. P. & HENRION, M. (1986). A framework for comparing uncertain inference systems to probability. In L. N. Kanal and J. F. Lemmer (eds.), *Uncertainty in Artificial Intelligence*. Amsterdam: Elsevier.

WISE, B. P., PERRIN, B., VAUGHAN, D., & YADRICK, R. (1987). The role of tuning uncertain inference systems. *Proceedings of the 3rd AAAI Workshop on Uncertainty in Artificial Intelligence,* Seattle, July 10-12, 1987.