REASONING WITH PARTIAL PREFERENCE MODELS

by

VU ANH HA

A Dissertation Submitted in
Partial Fulfillment of the
Requirements for the Degree of
Doctor of Philosophy in Engineering

at

The University of Wisconsin-Milwaukee

August 2001

REASONING WITH PARTIAL PREFERENCE MODELS

by

VU ANH HA

A Dissertation Submitted in
Partial Fulfillment of the
Requirements for the Degree of
Doctor of Philosophy in Engineering

at

The University of Wisconsin-Milwaukee

August 2001

Major Professor	(Signature)	Date
Graduate School Approval		Date

REASONING WITH PARTIAL PREFERENCE MODELS

by

VU ANH HA

The University of Wisconsin-Milwaukee, 2001
Under the Supervision of Peter Haddawy

Classical decision theory provides a normative framework for representing and reasoning with complex preferences. Straightforward application of this theory to automate decision making is difficult due to the high cost of eliciting preferences. The objective of this thesis is to develop a flexible decision-theoretic framework for eliciting and reasoning with preferences and to apply the tools and techniques of this framework to build a practical decision support system. This framework incorporates two orthogonal, complementary approaches to eliciting and reasoning with partial preference information. The first approach is grounded in classical multi-attribute utility theory, and can make effective use of qualitative preferential statements represented by logical comparative sentences. The second approach applies ideas from case-based reasoning and collaborative filtering to address the elicitation problem. The theoretical results of this framework have been tested extensively in a number of experiments involving real-world data, as well as found an application in a recommender system.

Major Professor (Signature) Date

© Copyright by Vu Ha, 2001
All Rights Reserved

ACKNOWLEDGEMENTS

I would like to thank my advisor, Peter Haddawy, for his continuous guidance and support from the very first day of my Ph.D. study in Milwaukee. He has always shown patience and encouragement in helping me learn how to approach a problem, how to write a research paper, and how to give a presentation. I am grateful to have him as my mentor.

I would like to thank my thesis committee members for their time. Prof. Jay Beder, from whom I took two excellent graduate courses in statistics, has been especially supportive of my graduate career in later years. I am also indebted to Prof. K. Vairavan for his help and encouragement in my teaching assignments at UWM over the years.

Two persons deserve special acknowledgements for inspiring and persuading me to pursue a Ph.D.: Prof. László Babai at University of Chicago, and my good friend AnHai Doan at University of Washington. Without these two persons, my career would have probably taken a completely different turn for the worse.

Special thanks are also due to my research collaborators and friends who have one way or another contributed to my successful graduation: John Miyamoto, Van H. Vu, Ben Geisler, Angelo Restificar, Tri Le, Hien Nguyen, Son Hai Ha, Du Truong Nguyen, Trung Nguyen, Hung Quang Nguyen, Son Truong Vu.

Finally, I thank my family for giving me love, support, and believing in me all these years.

Contents

List of Figures

CHAPTER 1
INTRODUCTION

Classical decision theory provides a normative framework for representing and reasoning with complex preferences. Straightforward application of this theory to automate decision making is difficult due to high cost of eliciting preferences. The objective of this thesis is to develop a flexible decision-theoretic framework for eliciting and reasoning with preferences, and to take the first steps toward applying the tools and techniques of this framework to build practical decision-theoretic advisory systems.

The theories, tools and techniques developed in the field of decision theory over the last fifty years seem to provide adequate foundations for building such advisory systems. Within this normative framework, we can represent and reason with complex preferences in the face of uncertainty and multiple objectives: the framework can capture important aspects of preference such as attitudes toward risk and tradeoffs among conflicting objectives. Lying at the heart of decision theory is the paradigm of *maximum expected utility (MEU)*: assuming a utility function that quantifies the desirability of the outcomes of the decisions, and a probability distribution over outcomes that captures the effects of each decision alternative, MEU dictates that the optimal decision alternatives are those that maximize expected utility (Savage 1954).

Built upon this deceptively simple paradigm, decision theory and its companion methodology of *decision analysis*, which we henceforth refer to collectively as *decision science*, have had a major impact in a number of disciplines. Decision theory provides a unified theory for statistical inferences (Wald 1950), and mathematical foundations for micro-economics (von Neumann & Morgenstern 1944). Recent advances in decision analysis over the past thirty years (Raiffa 1968; Keeney & Raiffa 1976; von Winterfeldt & Edwards 1986) have been applied to solve practical problems in fields as diverse as public policy and finance.

In spite of these successes, the tools and techniques of decision science have not proven fully adequate for building automated decision making systems. In fact, a large number of decision support systems are based on foundations other than formal decision theory. The art of decision analysis, as Glenn Shafer noted, "has not lent itself so readily to computer implementation" (Shafer & Pearl 1990) [1].

Perhaps the most difficult aspect of applying the techniques of decision analysis is the high cost of eliciting preferences, i.e., getting the utility numbers. In practice, decision analysts obtain these numbers by asking the decision maker a systematically selected set of questions, and analyzing the answers. This procedure is intellectually demanding – it makes the decision maker think hard about his objectives – and thus is often quite time-consuming. However, different decision making situations have

[1] Holtzman (Holtzman 1989) discusses the difficulties of using computers to implement decision analysis.

different perspectives with regard to this difficulty. I will discuss two such situations with contrasting perspectives.

The first situation involves a typical business decision problem where decision analytic techniques can be effectively used. Mr Ali, the manager of product planning at a company must decide whether to approve the introduction of a new product. This decision could have a major impact on the market share, and eventually on the profitability of the company. In addition to the bottom line profit, Mr Ali also has to consider several other political and personal factors involved in this decision. Because the problem is complex and the stakes are high, he decides to hire Mr Brown, a consultant who is an expert in decision analysis. The two spend several weeks analyzing various reports on market research, company resources, etc. Using the tools of decision analysis, Mr Brown helps Mr Ali clarify his objectives with regard to the introduction of the new product. In the end, Mr Ali decides to approve the new product.

The second situation is the sequel of the first. The new product turns out to be a blockbuster, generating substantial profit for the company. Because of his contributions, Mr Ali is promoted to the position of vice president of the company and awarded a four-week paid vacation. To plan for this well-deserved break, Mr Ali turns to his secretary, Ms Carol, for assistance. He tells Ms Carol that he intends to spend the next four weeks relaxing, scuba diving, and enjoying seafood, preferably on a location not too far from the US. Accustomed to planning vacations for her boss, Ms Carol immediately recommends a special vacation package on the Cayman Islands that she just heard about, emphasizing its attractive features. Mr Ali finds this plan satisfactory and tells Ms Carol to go ahead with the air ticket and hotel booking.

There are two factors that contrast the above two examples and motivated my work in this thesis. The first factor is S: *stake*. In the first problem, the decision maker only wants the absolutely best solution available, and is willing to bear the extra cost of hiring a consultant and then spending weeks clarifying the situations, identifying the objectives, and refining the model before arriving at a conclusion. The stakes here are high. In the second problem, the stakes are much lower and thus the decision maker makes a quick decision. He would certainly be opposed to spending a large amount of time analyzing his objectives or exploring the available alternative options for his vacation.

The second factor is R: *reusability*. High-stakes decision problems such as the one in the first example tend to be unique; they are formulated and solved only once, and their solutions are rarely reused. In contrast, a personal planning assistant (e.g., a software program that plays the role of Ms Carol) may service the same user in a number of similar decision making situations (e.g., Mr Ali usually takes a vacation at least twice a year). Moreover, such an advisory system may service a large number of users. While theoretically the system has to elicit the preferences of each user, the fact is that many of these users may have very similar preferences. Consequently, solutions for one user may be reused with little modification.

In this thesis, I set out to provide a decision-theoretic framework for building advisory systems that addresses the problem with factor S while exploiting factor R. These systems are designed as special-purpose advisory systems that provide advice for a particular decision making problem. Examples of such problems are personal planning problems (vacation, investment, etc), recommendation (book/film/music), and medical decisions. Below is the outline of this thesis.

- In Chapter 2, I discuss some background to decision theory and decision analysis.

- In Chapter 3, I address stake factor S, when the stake does not justify the the high cost of eliciting preferences, by developing a framework for reasoning with *partial* preference information. This approach is grounded in the well-understood theory of multi-attribute utility functions. It assumes that the user's utility function decomposes according to the *multi-linear form* with unknown scaling coefficients. The system then uses comparative statements by the user, represented by *qualitative* logical constructs, to constrain these coefficients. These constraints can in turn induce that certain decision alternatives are sub-optimal. This hybrid approach is one of the first attempts to enhance the quantitative theory of multi-attribute utilities with some recently developed qualitative, logic-oriented approaches to reasoning with preferences.

- In Chapter 4, I investigate a case-based approach to preference elicitation and decision making that exploits factor R, the reusability of situations in advisory systems. To help a user A make a decision, the system may choose to elicit some preference information from A. Based on this information, the system then uses utility models of users who have similar preferences to A's to recommend candidate solutions. This approach, which was partly inspired by recent advances in collaborative information filtering, attempts to incorporate ideas and techniques from the area of case-based reasoning into automated decision making. The focus of my investigation is the probabilistic distance, a measure of similarity among people's preferences that has its roots in the Kendall's tau function. I show that this measure has attractive theoretical properties, can be approximated efficiently in all situations, and has a number of advantages over existing similarity measures on preferences.

CHAPTER 2
BACKGROUND TO DECISION SCIENCE

2.1. The Essence Of Decision Theory

The field of Decision Theory is concerned with the principles of making rational decisions. Nowadays, it is often identified with *Bayesian Decision Theory*, which is based on a subjective interpretation of probability theory, and which has distilled ideas from statistical decision theory and utility theory. Thanks to its elegant intuitions and rigorous mathematical foundations, Bayesian Decision Theory, referred to simply as Decision Theory (DT) in this discussion, has now become the dominant theory based on which many practical decision making problems are formulated and solved. It has become a reference point to which any other general theories of decision making must relate.

While the origins of DT can be traced back several hundred years, the most important developments took place about half century ago, culminating in the work of Leonard Savage, published in his book titled *The Foundations of Statistics* (Savage 1954). In this book, Savage presented an axiomatic treatment of principles of decision making under uncertainty that leads to the paradigm of maximum expected utility. This paradigm is drawn from the axioms of probability theory and utility theory. While probability theory provides a framework for coherent assignment of *beliefs* in the face of incomplete information, utility theory provides a set of principles for consistency among *preferences and decisions*. In the following, we briefly describe the essence of this axiomatization, while at the same time introduce several important concepts and terminologies that will be used throughout this thesis. The reader is referred to Savage's book, or several others (de Finetti 1974; Fishburn 1981; Kreps 1988) for a complete exposition.

When a decision problem involves no uncertainty, the consequences of decisions are called *outcomes*. We denote the set of outcomes by Ω. For ease of exposition, we will assume throughout the paper that Ω is finite and $\Omega = \{1, 2, \ldots, n\}$. As a highly simplified example, we will meet Mr. Ali, an about-to-graduate MBA student with a hobby of watching the television series *Baywatch*. When time comes for Mr. Ali to make a career decision, he thinks of two options: he can choose to be either a lifeguard or a corporate manager. With regard to this decision, he considers to factors: health (H) and wealth (W), which are assumed to be propositional (*healthy/not-healthy, wealthy/not-wealthy*). According to his belief, a lifeguard career would provide him a healthy lifestyle but a modest income. The outcome of this decision is $H\overline{W}$. On the other hand, a corporate manager career would make him rich but the business stress leave him in poor health. The outcome of this decision is $\overline{H}W$.

When uncertainty is involved, the consequences of decisions are probability distributions over outcomes, which we will refer to as *prospects*. A prospect is thus a

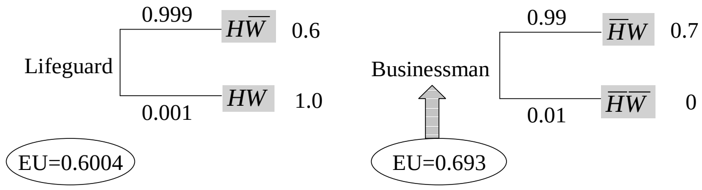

Figure 2.1. The Maximum Expected Utility Paradigm

probability distribution over Ω. As an example, let assume that a lifeguard may have a (small) chance of .1% of getting rich (e.g., by saving a millionaire from drowning and getting a big reward). The prospect of a lifeguard career is thus $(H\overline{W} : .999, HW : .001)$. On the other hand, a corporate manager could go bankrupt with a small probability of 1% (e.g., because of bad business decisions). The prospect of a corporate manager career for Mr. Ali is thus $(\overline{H}W : .99, \overline{H}\overline{W} : .01)$.

From now on, the terms decision alternatives and consequences will be used interchangeably. Outcomes are used to indicate certain consequences, and prospects are used to indicate uncertain consequences. We use $\mathcal{D}$ to denote the set of decision alternatives.

In the axiomatic framework of Savage, preference of the decision maker is postulated to be a *weak order* among decision alternatives. This means that the preference of the decision maker is an *asymmetric* ($a \prec b \Rightarrow b \not\prec a$), and *negatively transitive* ($a \not\prec b, b \not\prec c \Rightarrow a \not\prec c$) binary relation on the set of alternatives. We will call this order the *preference order*, or sometimes *preference structure* of the decision maker. The meaning of this order is that $a \prec b$ indicates that the decision maker *prefers* alternative b to alternative a. When neither of the two alternatives is preferred to the other ($a \not\prec b, b \not\prec a$), we say that the decision maker is *indifferent* between them and denote this relation by $a \sim b$. We also use the notation $a \preceq b$ to denote that the decision maker either prefers b to a or is indifferent between them.

In the case of certainty, it can be proven (Kreps 1988) that for any preference order $\prec$ over Ω there exists a function v, called a *value function*, that is *consistent* with $\prec$, i.e., for any outcomes i, j, $i \prec j \Leftrightarrow v(i) < v(j)$. We sometimes write $\prec$ as $\prec_v$ to emphasize this relationship. Two value functions that induce identical orders are said to be *strategically equivalent*. Otherwise, they are said to be *strategically different*. Furthermore, if $(\Omega, \prec)$ is a strict order, then a value function consistent with $\prec$ is the permutation $\pi : \Omega \rightarrow \Omega$ that satisfies that $i \prec j \Leftrightarrow \pi(i) < \pi(j), \forall i, j \in \Omega$ [1]. We will refer to this permutation as the *canonical value function* for the strict preference

[1]We abuse notation a little bit here. In the proposition $i \prec j$, i and j are decision outcomes, while in the proposition $\pi(i) < \pi(j)$, $\pi(i)$ and $\pi(j)$ are integer numbers.

order $\prec$. In many applications, decision outcomes are assigned non-negative numbers called *ratings* (for example, movie ratings, news article ratings, etc). These ratings can be viewed as a value function if we assume that higher ratings are preferred to lower ratings, and equal ratings are equally preferred.

In the case of uncertainty, Savage's postulates about the preference of the decision maker are the following. The *monotonicity postulate* says that, when comparing two prospects, each with the same two alternative outcomes but different probabilities, a decision maker should prefer the prospect that has the higher probability of the preferred outcome. The *decomposability postulate* says that a decision maker should be indifferent between prospects that have the same set of eventual outcomes and probabilities, even if they are reached by different means. The *substitutability postulate* says that, if a decision maker is indifferent between a prospect and some outcome, then substituting one for the other in some more complex prospect should not affect her preference for that prospect. Finally, the *continuity postulate* says that, if one prefers outcome x to y, and y to z, then there is some probability p such that one is indifferent between getting the intermediate outcome y for sure and a prospect with a p chance of x (the preferred outcome) and $(1-p)$ chance of z (the inferior outcome).

It follows from accepting the above postulates that there exists a *utility function* $u : \Omega \rightarrow R$ such that preference order among prospects can be established based on the expectation of u over outcomes. Formally, this means that $p \prec q$ iff $E_p(u(s)) < E_q(u(s))$. Such a utility function is also said to be *consistent* with the preference structure of the decision maker. We sometimes write $\prec$ as $\prec_u$ to emphasize this relationship. Two utility functions that induce identical orders are said to be *strategically equivalent*. Otherwise, they are said to be *strategically different*. The power of this result is that it allows preferences over complex *prospects*, including *attitudes toward risk*, to be captured by a real-valued function over *outcomes*. Thus, it may be used as a tool to help people cope with complex decision problems by identifying a utility function.

Continuing the previous example, suppose the utility function u for Mr. Ali is such that $u(HW) = 1, u(H\overline{W}) = .6, u(\overline{H}W) = .7, u(\overline{HW}) = 0$. Mr. Ali should choose the corporate manager career because its expected utility (.693) is higher than that of a lifeguard career (.6004).

2.2. The Field Of Decision Analysis

How do we obtain those utility numbers that seemingly magically reduce the complex problem of making a decision to a simple expectation calculation? The answer to this question lies in the realm of Decision Analysis. Decision Theory provides an intuitively appealing axiomatic framework for reasoning with preferences in the face of uncertainty. Decision analysis, on the other hand, provides the *methodology* to ap-

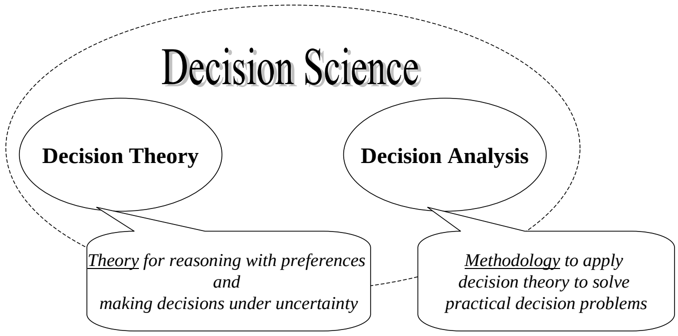

Figure 2.2. Decision Theory + Decision Analysis = Decision Science

ply this framework to solve practical decision making problem. Decision Theory and Decision Analysis are together referred to as *Decision Science* (see Figure 2.2).

Since its inception around the 1960s, decision analysis has grown into an established academic and professional discipline. It has developed a number of powerful ideas and techniques to help people and organizations analyze and solve decision problems. Among the major advances of decision analysis are the following.

- Decision trees (Raiffa 1968), a tool for structuring complex decision problems that contain many sequential actions with uncertain outcomes.
- Influence diagrams (Howard & Matheson 1984), a tool that provides a compact representation for decision trees. Enhanced with the recent advances in Bayesian networks (Pearl 1988), influence diagrams now provides an attractive graphical representation that encodes probabilistic and causal relationships in a decision problem. This expressiveness greatly simplifies the tasks of clarifying the decision alternatives and belief, identifying the objectives, and computing the expected utility (Shachter 1986). The representation also facilitates sensitivity analysis.
- Multi-attribute utility theory (MAUT) (Keeney & Raiffa 1976). This theory is developed to help people make tradeoffs between multiple objectives. MAUT is discussed in more detail in Chapter 3.

2.3. Decision Science as a Prescriptive Tool for Decision Making

Mathematical rigor and aesthetic appeal notwithstanding, DT is not many things. For the purpose of this thesis particularly, there are three points that are relevant.

First, it is widely accepted that decision theory is *not descriptive*; it does not describe how people make decisions. Psychological experiments, particularly those carried out by Tversky and Kahneman, showed that people often deviate from the MEU paradigm in making their everyday decisions, and these deviations are *substantial* (Tversky & Kahneman 1986). In fact, compelling arguments have been presented that suggest that no adequate normative theory can at the same time be descriptively accurately (Tversky & Kahneman 1986).

Second, critics of DT have also argued that it is *not normative* either. These scholars either object to the Bayesian interpretation of probability theory and its use to represent personal belief (see, for example, (Efron 1986)), or challenge Savage's axioms as the guideline for rationality (see, for e.g. Glenn Shafer's article (Shafer 1986) and the follow-up commentaries that appeared in (Shafer & Pearl 1990)).

What is a theory of decision making good for if it is neither descriptive nor normative? A response for this descriptive inadequacy and normative controversy was offered by Ralph Keeney and Howard Raiffa in their classic work on multi-attribute utility theory (Keeney & Raiffa 1976):

> ... In this sense, the [decision-theoretic] approach is not *descriptive*, because most people do not attempt to think systematically about hard choices under uncertainty. It is also not *normative* since it is not an idealized theory designed for the superrational being with an all-powering intellect. It is, instead, a *prescriptive* approach designed for normally intelligent people who want to think hard and systematically about some important real problems.
>
> Keeney and Raiffa (1976, Preface).

The research carried out in this thesis takes this prescriptive approach as its normative guideline.

The third point about decision theory that is relevant to this thesis is the importance of the distinction between good decisions and good outcomes. This point is best illustrated by a quote from a paper by Warner North (North 1990):

> We might distinguish between a good decision and a good outcome. We are all familiar with situations in which careful management and extensive planning produce poor results, while a disorganized and badly managed competitor achieved spectacular success. As an extreme example, place yourself in the position of the company president who has discovered that a valuable and trusted subordinate whose past judgements had proved unfailingly accurate actually based his decisions upon the advice of a gypsy fortune teller. Would you promote this man or fire him? The answer, of course, is to fire him and hire the gypsy as a consultant. The availability

of such a clairvoyant to provide perfect information would make decision theory unnecessary. But we should not confuse the two. Decision theory is not a substitute for the fortune teller. It is rather a procedure that takes account of all available information to give us the best possible logical decision. It will minimize the consequences of getting an unfavorable outcome, but we cannot expect our theory shield us from all “bad lucks”. The best protection we have against a bad outcome is a good decision.

North (1990, Introduction).

CHAPTER 3
A CONSTRAINT-BASED APPROACH TO REASONING WITH PARTIALLY SPECIFIED MULTI-LINEAR UTILITY FUNCTIONS

In this chapter, we develop a constraint-based framework for reasoning with *partial* preference information. This approach is grounded in the well-understood theory of multi-attribute utility functions. The user's utility function is assumed to decompose according to the *multi-linear form* with unknown scaling coefficients. Qualitative, pairwise comparisons among decision consequences are used as constraints on these coefficients, which in turn can induce dominance and potential optimality

3.1. Some Preliminaries

Similar to the early days when probability theory was considered epistemologically inadequate for AI (McCarthy & Hayes 1969), utility theory for decision analysis these days still faces several epistemological and computational problems of its own. In particular, it is often quite difficult to elicit the required utility function, especially when the outcomes of the decisions are complex. Decision analysts have addressed this issue by developing a comprehensive framework, generally known as multi-attribute utility theory (MAUT) (Keeney & Raiffa 1976), for reasoning with such complex utilities. This theory exploits the fact that outcomes of decisions can often be described in terms of a set of *features* or *attributes*. (The reader has already encountered the concept of attributes in an earlier example in Chapter 2. In that example, the outcomes of Mr Ali's career decision are described by two attributes, *health* and *wealth*.) Based on this framing of the problem, utility functions can often be decomposed into simpler, easy-to-obtain sub-utility functions over individual attributes. In this section, we shall present the key concepts and results of MAUT that are relevant to this work. The reader is referred to (Keeney & Raiffa 1976) for a comprehensive treatment of this subject.

Utility Independence

Lying at the heart of MAUT are the notions of *preferential independence* for decision making under certainty, and its generalization to the case of uncertainty, *utility independence*. Suppose that an outcome can be described by a set of attributes $X = \{X_1, X_2, ..., X_n\}$, meaning that an outcome x is a value assignment $(X_1 = x_1, X_2 = x_2, ..., X_n = x_n)$ to the attributes. Abusing notation, we denote the set of values an attribute X_i can take simply by X_i, and thus the outcome space Ω is just the Carthesian product $X_1 \times X_2 \times \cdots \times X_n$. A subset $Y \subseteq X$ is also referred to as an attribute (it can be viewed as a composite attribute).

Definition 1. (Preferential Independence) *Given a preference order $\succeq$ over outcomes,* *an attribute* $Y \subset X$ *is* preferentially independent *of its complement* Z*, or, for short,* Y *is PI, if the preference* $\succeq$ *over outcomes that are fixed in* Z *at some level does not depend on this level.*

Definition 2. (Utility Independence) *Given a preference order* $\succeq$ *over the prospects, an attribute* $Y \subset X$ *is* utility independent *of its complement* Z*, or, for short,* Y *is UI, if the preference* $\succeq$ *over prospects that are fixed in* Z *at some level does not depend on this level.*

Utility independence occurs quite often in real-life decision making situations, and in general can be detected easily. For example, the utility function of Mr. Ali in our example in Chapter 2 ($u(HW) = 1$, $u(H\overline{W}) = .6$, $u(\overline{H}W) = .7$, and $u(\overline{HW}) = 0$) suggests utility independence of H and W. Intuitively, with respect to Mr Ali's preferences over H, no matter what his *wealth* is fixed to, he will always prefers prospects that yield *healthy* with higher probability. The same can be said about his preferences over W.

Additive, Multiplicative, and Multi-Linear Utility Functions

When an attribute Y is UI, the following simple theorem (Keeney & Raiffa 1976) shows that we can write the utility function $u(x)$ as an expression that consists of two functions over $X - Y$ and one function over Y, achieving a reduction of dimensionality (and hence complexity).

Theorem 1 (Basic Decomposition). *If some attribute* $Y \subset X$ *is UI, then the utility function* $u(x)$ *must have the form:*

$$u(x) = u(y, z) = g(z) + h(z).u(y, z^+),$$ [1]

where $g(\cdot)$ *and* $h(\cdot) > 0$ *depends only on* z *but not on* y*, and* z^+ *is some fixed value of* z*.*

Utility independence thus plays just as fundamental a role in utility theory as does probabilistic independence in probability theory: it provides modularity and decomposition. In particular, if every subset Y of X is UI, a condition called *mutual utility independence (MUI)*, then we can write $u(x)$ either in an *additive form*

$$u(x) = \sum_{i=1}^{n} k_i u_i(x_i),$$

or in a *multiplicative form*

[1]The function $u(y, z^+)$, defined over variable y is called the *subutility function* for the attribute Y, and is sometimes designated by $u_Y(y)$.

$$u(x) = \prod_{i=1}^{n}(1 + k_i u_i(x_i)),$$

where u_i are so-called *sub-utility* functions that capture the decision maker's preference with regard to attribute X_i when holding the attributes $X_j, j \neq i$ at some fixed level, and k_i are constants that ensures proper global scaling. Usually, the constants k_i are scaled so that both u and the sub-utility functions u_i have 0 as minimum and 1 as maximum.

Whenever MUI is applicable, the decision analyst can obtain the utility function in the following two steps.

1. The subutility functions $u_i, i = 1, ..., n$ are assessed. For this purpose, the decision maker is asked to rank decision alternatives that have the values for $X_j, j \neq i$ fixed. This step is relative simple. For example, when X_i is propositional, the cost of assessing u_i is zero; u_i is either 0 or 1, depending on x_i being the inferior or superior value.

2. The scaling coefficients $k_i, i = 1, ..., n$ are assessed, typically by determining the relative ratio for each pair of coefficients. The user is asked to rank alternatives that differ in the two corresponding attributes.

In real-world situations, however, MUI is often not applicable. For example, we may only be able to structure the state space in terms of attributes that are individually UI, i.e., $\{X_i\}$ is UI, $i = 1, 2, ..., n$. In such cases, the utility function takes on the *multilinear* form

$$u(x) = \sum_{\emptyset \neq Y \subseteq X} k_Y \prod_{X_i \in Y} u_i(x_i), \tag{3.1}$$

where u_i are sub-utility functions, and $k_Y, \emptyset \neq Y \subseteq X$ are scaling coefficients [2].

3.2. Outline Of the Constraint-Based Approach

In the case when the utility function is assumed to be multi-linear, while obtaining the sub-utility functions u_i is still relatively easy, assessing a total of $2^n - 1$ scaling coefficients is quite daunting [3]. This complexity poses a difficult dilemma to the decision analyst: she can either work with an additive or a multiplicative function

[2] See Keeney and Raiffa (1976, Chapter 6) for a detailed comparisons of various independence assumptions and their implications.

[3] In fact, in the case when there are more than 3 attributes, assessing multi-linear utility functions is usually abandoned.

even when evidence suggests that MUI is violated, in effect obtaining an approximate model of the decision maker's preference, or work with a multilinear utility function, placing a sizable elicitation burden on herself and on the decision maker.

Our perception is that in real-world applications of decision theory, decision analysts often choose the former option: to assume MUI. In contrast, we choose to work with multi-linear utility functions [4]. First, we will present a set of techniques that help identifying sub-optimal decision alternatives without assessing all the $2^n - 1$ scaling coefficients (expected utilities thus are not explicitly computed). These techniques assume the following:

(i) The set of viable decision alternatives is finite. Each decision alternative results in a completely specified probability distribution over the states of the world.

(ii) The decision maker's utility function is multilinear. The sub-utility functions have already been elicited, but the scaling coefficients are still unknown.

(iii) The decision maker can provide a set of preferential comparisons of the form $p_j \preceq q_j, j$, where p_j and q_j are decision alternatives.

The essence of this approach is that each of the above assumptions can be translated into a constraint, or a set of constraints on the unknown scaling coefficients k_Y. These constraints can then be used to deduce further preferential information such as *induced dominance* and *potential optimality.* They can also be used to detect user preferential inconsistency. We are thus able to make useful preferential inference without the knowledge of the *scaling coefficients* but instead with the knowledge of certain *constraints on them.*

The key assumption in this approach is (iii), which assumes that the decision maker can provide preferential comparisons between (real or fictious) decision alternatives. This assumption seems quite reasonable in certain circumstances where it might be easier for the decision maker to express preferences among decision alternatives than to introspect about the attributes describing each one. For example, in expressing preferences about movies, most people can readily express their preferences over two films they have seen in the past but may have difficulty describing preferences over attributes like director, leading actor, or costume designer. In fact, most people would not even recognize the names of the costume designers, even when they may have a preference for films with nice costumes.

The more comparisons the decision maker can provide, the more conclusive inference can be made. In particular, if the decision maker can provide a succinct,

[4]Bacchus and Grove also take this stand: "We conjecture that [multilinear utility models] might be worth studying in the context of artificial intelligence applications, and in particular for giving a better decision-theoretic account of goal" (Bacchus & Grove 1997).

qualitative statement about her preference that implicitly encoded a *set* of comparison statements, then we may be able to quickly identify a large set of sub-optimal alternatives.

To capture such preferential statements and derive efficient inference mechanisms using them is one of the aims of the field of qualitative decision theory. Recent work from this field has attempted to address the elicitation problem by providing formal languages in which partial preference information can be conveniently expressed (Doyle, Shoham, & Wellman 1991; Doyle & Wellman 1994; Tan & Pearl 1994; Boutilier 1994; Bacchus & Grove 1996). For example, the languages proposed by Doyle and Wellman (Doyle & Wellman 1994) and by Tan and Pearl (Tan & Pearl 1994) attempt to provide a semantic to *ceteris paribus* (all else being equal) comparative statements. These are preferential statements concerning classes of decision consequences.

While these languages have successfully addressed a number of expressiveness issues, the inferential mechanisms available have not been sufficiently powerful for building practical decision making systems. In Section 3.4, we propose using ceteris paribus comparative statements, as presented in (Doyle & Wellman 1994), as a means to represent comparative statements made by the decision maker, to be used in conjunction with the assumptions (i), (ii), and (iii) described above. In Section 3.5, we discuss related literature on reasoning with partial utility information. Finally, we give a summary and discuss future work in Section 3.6.

3.3. Representing And Reasoning With Multi-Linear Utility Functions Using Polyhedral Cones

In this section we explore the idea of using explicit pairwise comparisons of decision consequences to identify sub-optimal alternatives, as outlined in Section 3.2. The premises of this analysis are assumptions (i), (ii), (iii). We begin by introducing some basic concepts of convex geometry.

3.3.1. Preliminary: Some Basic Concepts of Convex Geometry

Polyhedra, Cones, and Polyhedral Cones. The *d-dimension Euclidean Space* is the vector space $\Re^d$ equipped with the inner product $\langle\rangle$. Given a vector $\mathbf{n}$, and $\alpha \in \Re$, the set $\mathbf{n}_\alpha = \{\mathbf{x} | \langle \mathbf{n}, \mathbf{x} \rangle = \alpha\}$ is called a *hyperplane*, the set $\mathbf{n}_\alpha^- = \{\mathbf{x} | \langle \mathbf{n}, \mathbf{x} \rangle \leq \alpha\}$ is called a *closed halfspace* with *outward normal* $\mathbf{n}$. Similarly, the set $\mathbf{n}_\alpha^+ = \{\mathbf{x} | \langle \mathbf{n}, \mathbf{x} \rangle \geq \alpha\}$ is called a *closed halfspace* with *inward normal* $\mathbf{n}$. For simplicity of notations, the subscript α is omitted when $\alpha = 0$. The intersection of a finite number of closed halfspaces is called a *polyhedron*.

A set $K \subseteq \Re^d$ is called a *cone with apex* $\mathbf{0}$ if $\lambda\mathbf{x} \in K$ whenever $\lambda \geq 0$ and $\mathbf{x} \in K$. A set K is a cone with apex $\mathbf{a} \in \Re^d$ if $K - \mathbf{a} := \{\mathbf{x} - \mathbf{a} | \mathbf{x} \in K\}$ is a cone with apex $\mathbf{0}$. In this thesis, cones are all $\mathbf{0}$-apexed, unless indicated otherwise. Given a set $K \subseteq \Re^d$, the set of all points that can be expressed as non-negative linear combinations of points of K can be shown to be a convex cone, and is denoted by C_K. This cone is called the convex cone generated by K and can be equivalently defined as the smallest convex cone containing K. A cone that is also a polyhedron is called a *polyhedral cone*. It is well-known that polyhedral cones are precisely convex cones generated by finite sets of points, and can be shown to be *closed*.

Dual Cones. Given a set $K \subseteq \Re^d$, let K^* be defined as $K^* := \{\mathbf{y} | \langle \mathbf{x}, \mathbf{y} \rangle \leq 0, \forall \mathbf{x} \in K\}$. K^* is easily shown to be a convex cone and is referred to as the *dual cone of* K. For example, if K contains a single point $\mathbf{n}$, then the dual cone of K is $\mathbf{n}^-$, the closed halfspace with outward normal $\mathbf{n}$. The following theorem, which is standard in convex cone analysis, will be used later in this thesis.

Theorem 2. *Let* $K \subseteq \Re^d$. *Then*

(a) $K^* = (C_K)^*$. *Any set and the convex cone it generates share the same dual cone.*

(b) $K^{**} := (K^*)^* = \overline{C}_K$. *The dual cone of the dual cone of* K *is equal to the closure of the convex cone it generates. In particular, if* K *is finite, then* $K^{**} = C_K$.

3.3.2. Linear Constraints on the Scaling Coefficients of Multi-Linear Utility Functions

First, note that the multilinear form of the utility function, as formalized in equation (3.1) does not fully capture the assumption that the attributes X_i are UI; the multilinear form is only a necessary but not sufficient condition for X_i to be UI. We need to add constraints on the scaling coefficients k_Y in order to obtain a necessary and sufficient condition.

Take, for example, the assumption that X_i is UI. Let $t_Y(x) = \prod_{X_j \in Y} u_j(x_j)$, $\emptyset \neq Y \subseteq X$. Then the multi-linear utility function $u(x)$ in equation 3.1 can be written as

$$\begin{aligned} u(x) &= \sum_{\emptyset \neq Y \subseteq X} k_Y t_Y(x) \\ &= \left(\sum_{Z \subseteq X - \{X_i\}} k_{\{X_i\} \cup Z} t_Z(x) \right) u_i(x_i) + \sum_{\emptyset \neq Y \subseteq X - \{X_i\}} k_Y t_Y(x), \end{aligned}$$

and thus can be viewed as a linear function of $u_i(x_i)$. Then to say that X_i is UI is equivalent to say that the coefficient for $u_i(x_i)$ in this linear function must be non-negative. Formally, this means

$$\sum_{Z \subseteq X - \{X_i\}} k_{\{X_i\} \cup Z} t_Z(x) \geq 0. \tag{3.2}$$

Moreover, this inequality must be satisfied for *any* value assignment to the attributes in the set $X - \{X_i\}$. Inversely, if this constraint is satisfied, then X_i is UI. In other words, the utility independence of the attributes X_i is precisely captured by the multilinear form as in equation (3.1), with the additional linear, homogeneous constraints about the scaling constants, as expressed in inequality (3.2).

To further simplify the expositions, we introduce the following notations. Let $\mathbf{k}$ denote the $(2^n - 1)$-dimensional vector with components $k_Y, \emptyset \neq Y \subseteq X$. For any $i = 1, 2, ..., n$, let $\mathbf{s}^i$ denote the same-dimension vector whose components are functions $s_Y^i : X \rightarrow \Re, \emptyset \neq Y \subseteq X$, defined as:

$$s_Y^i(x) = \begin{cases} 0 & \text{if } X_i \notin Y \\ -t_{Y - \{X_i\}}(x) & \text{otherwise.} \end{cases}$$

Note that the functions $s_Y^i(x)$ do not depend on x_i, the i-th component of x. Then inequality (3.2) can be written as

$$\langle \mathbf{k}, \mathbf{s}^i(x) \rangle \leq 0, \forall i = 1, 2, ..., n; x,$$

where $\langle ., . \rangle$ denotes the inner product of two vectors.

One may at first think that these constraints are by themselves strong enough to imply non-trivial preferences over decision alternatives. However, the following theorem suggests that further preferences can be deduced in only very special cases.

Theorem 3. *Let p and q be two decision alternatives. Further assume that $E_p[u_i(x_i)] \leq E_q[u_i(x_i)]$, for all i, i.e., q would be preferred to p if we took the sub-utility function $u_i(x_i)$ as our overall utility function. Then from these so-called* local dominance *conditions, we can infer* overall dominance, *i.e. $p \preceq q$ if either:*

(a) the utility function is additive, or

(b) in the probability distributions p and q, the attributes X_i, when viewed as random variables, are probabilistically independent.

Furthermore, the inference ($p \preceq q$) is not *sound if the utility function is multiplicative.*

Proof

To prove Part (a), suppose that the utility function u is additive: $u(x) = \sum_{i=1}^n k_i u_i(x_i)$. Note that the scaling constants k_i of an additive utility function are non-negative.

$$
\begin{aligned}
E_p[u(x)] &= \int_\Omega u(x)p(x)dx \\
&= \int_\Omega (\sum_{i=1}^n k_i u_i(x_i))p(x)dx \\
&= \sum_{i=1}^n k_i \left(\int_\Omega u_i(x_i)p(x)dx \right) \\
&= \sum_{i=1}^n k_i E_p[u_i(x_i)] \\
&\leq \sum_{i=1}^n k_i E_q[u_i(x_i)] \text{ (Because of local dominances)} \\
&= E_q[u(x)] \text{ (Using analogous derivations) (q.e.d)}
\end{aligned}
$$

To prove Part (b), first we note that the multi-linear function u in Equation 3.1 can be viewed as the composition $u = h \circ w$, where $w : \Omega \to R^n$, $w(x) = (u_1(x_1), \ldots, u_n(x_n))$, and $h : R^n \to R$, $h(w_1, w_2, \ldots, w_n) = \sum_{\emptyset \neq Y \subset X} k_Y \prod_{X_i \in Y} w_i$. From the Basic Decomposition (Theorem 1), and the well-known fact that utility functions are unique up to a positive linear transformation, it follows that h is a monotonically non-decreasing function with respect to *each* of its variables.

With the introduction of this decomposition, the expected utility of p can be written as:

$$
\begin{aligned}
E_p[u(x)] &= E_p[h \circ w(x)] \\
&= \sum_{\emptyset \neq Y \subset X} k_Y E_p[\prod_{X_i \in Y} u_i(x_i)] \\
&= \sum_{\emptyset \neq Y \subset X} k_Y \prod_{X_i \in Y} E_p[u_i(x_i)] \text{ (because } x_i \text{ are prob. indep. wrt } p\text{)} \\
&= h(E_p[u_1(x_1)], \ldots, E_p[u_n(x_n)]) \\
&\leq h(E_q[u_1(x_1)], \ldots, E_q[u_n(x_n)]) \text{ (Because of local dominances)} \\
&= E_q[u(x)] \text{ (Using analogous derivations) (q.e.d)}
\end{aligned}
$$

Finally, we present a counterexample that shows that if the utility function has multiplicative form, i.e. when we assume only MUI, it is possible that q dominates p with respect to each individual attribute, yet p dominates q overall. Let $X = \{Y, Z\}$,

$\Omega = \{(y_1, z_1), (y_2, z_2), (y_3, z_3)\}$. The overall utility function u is the product of the subutility functions, u_Y and u_Z: $u(y, z) = u_Y(y)u_Z(z)$. This can be derived from the multiplicative form, whenever k is positive. The utility functions and the two density functions p and q are specified in the following table.

Y	Z	u_Y	u_Z	u	p	q
y_1	z_1	1	0	0	1/4	1/2
y_2	z_2	0	1	0	1/4	1/2
y_3	z_3	1/3	1/3	1/9	1/2	0

Clearly, q dominates p with respect to both Y and Z ($E_p[u_Y] = E_p[u_Z] = 5/12 < E_q[u_Y] = E_q[u_Z] = 1/2$), but overall p dominates q ($E_p[u] = 1/18 > E_q[u] = 0$) (q.e.d) [5].

3.3.3. Linear Constraints for Pairwise Comparisons

In order to be able to infer a pairwise preference, we need to impose very strong conditions, either about the form of the utility function (it must be additive), or about alternatives (they must be probabilistically independent), in addition to having the local dominances. When these conditions do not hold, we need other sources of preferential information in order to be able to identify sub-optimal alternatives and to narrow down the set of candidate alternatives. One such source is pairwise comparison statements made by the decision maker.

Note that the statement that $p \preceq q$ translates into the following inequalities:

$$\begin{aligned} E_p[u(x)] &\leq E_q[u(x)] \Leftrightarrow \\ E_p\left[\sum_Y k_Y t_Y(x)\right] &\leq E_q\left[\sum_Y k_Y t_Y(x)\right] \Leftrightarrow \\ \sum_Y k_Y E_p[t_Y(x)] &\leq \sum_Y k_Y E_q[t_Y(x)]. \end{aligned}$$

Now, denoting $t_Y(p) = E_p[t_Y(x)] = \sum_x p(x)t_Y(x)$, and $t_Y(q) = E_q[t_Y(x)] = \sum_x q(x)t_Y(x)$, we then have

$$\begin{aligned} p \preceq q &\Leftrightarrow \sum_Y k_Y t_Y(p) \leq \sum_Y k_Y t_Y(q) \\ &\Leftrightarrow \langle \mathbf{k}, \mathbf{t}p - \mathbf{t}q \rangle \leq 0, \end{aligned}$$

[5] It is interesting to ask if there is a condition that would ensure the inference, yet is weaker than additive independence.

where $\mathbf{t}p$ (respectively $\mathbf{t}q$) denotes the $(2^n - 1)$-dimensional vector whose components are $t_Y(p)$ (respectively $t_y(q)$). This last inequality is also a linear, homogeneous constraint over the unknown constants k_Y.

3.3.4. Induced Dominance and Potential Optimality

In summary, the analysis in Subsections 3.3.2 and 3.3.3 show that the assumptions (i), (ii), and (iii) described in Section 3.2 can be precisely captured by the following inequalities

$$\begin{cases} \langle \mathbf{k}, \mathbf{s}^i(x) \rangle & \leq \; 0, \forall i, x \\ \langle \mathbf{k}, \mathbf{t}p_j - \mathbf{t}q_j \rangle & \leq \; 0, \forall j. \end{cases}$$

Assuming that the domain of each attribute X_i is finite, the above inequalities are equivalent to a finite set of linear homogeneous constraints over the scaling coefficients k_Y. From now on, we denote these constraints as follows $\{\langle \mathbf{k}, \mathbf{w}_j \rangle \leq 0 | j\}$, and denote $W = \{\mathbf{w}_j | j\}$. Thus, the utility function u, when represented as a vector $\mathbf{k}$ with coordinates k_Y, must lie in the intersection of the closed halfspaces with outward normal vectors $\mathbf{w}_j$:

$$\mathbf{k} \in K := \bigcap_{j=1}^{m} \mathbf{w}_j^- .$$

This intersection is a polyhedral cone, which is the *dual cone* $K = W^*$ of W. Using Theorem 2, we deduce that $K^* = W^{**} = C_W$, i.e., the dual cone of K, the set of admissible $\mathbf{k}$, is the polyhedral cone generated by the constraint vectors $\mathbf{w}_j$. Thus given a pair of alternatives (p, q), we can deduce $p \preceq q$ if and only if $\mathbf{t}p - \mathbf{t}q \in C_W$. The "only if" part means that if $\mathbf{t}p - \mathbf{t}q \notin C_W$, then there exists $\mathbf{k} \in K$ such that $\langle \mathbf{k}, \mathbf{t}p - \mathbf{t}q \rangle > 0$, i.e. $p \succ q$ under such $\mathbf{k}$. Figure 3.1 illustrates this observation. Based on this observation, we have the following algorithm for testing induced dominance. This "inference rule" is sound and complete based on the above observation.

Algorithm 1. Input: *Two prospects p and q.*
Output: *Return*

$$\begin{cases} 1 & \textit{if } p \preceq q, \\ -1 & \textit{if } q \preceq p, \\ 0 & \textit{if the relationship between } p \textit{ and } q \textit{ cannot be determined.} \end{cases}$$

1. *Determine a set of generators $\{\mathbf{k}_l | l\}$ of the polyhedral cone K. These vectors are outward normal vectors of the of the closed halfspaces whose intersection is C_W, and can be computed using several methods (see for e.g. (Matheiss & Rubin 1980)).*

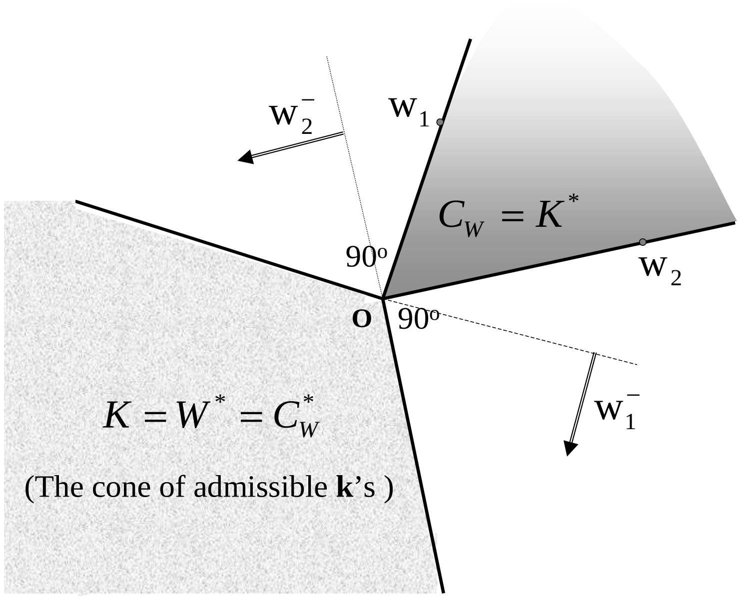

Figure 3.1. A 2-dimensional illustration of the dual cone observation. Here $W = \{\mathbf{w}_1, \mathbf{w}_2\}$.

2. *Return* 1, *if* $\langle \mathbf{k}_l, \mathbf{t}p - \mathbf{t}q \rangle \leq 0, \forall l$.

3. *Return* -1, *if* $\langle \mathbf{k}_l, \mathbf{t}q - \mathbf{t}p \rangle \leq 0, \forall l$.

4. *Return* 0, *otherwise.*

Complexity Analysis. The complexity of the Algorithm 1 is determined by the complexity of determining the generators $\{\mathbf{k}_l | l\}$ of the polyhedral cone K (step 1). These vectors are outward normal vectors of the $(d-1)$-dimension facets of the polyhedron C_W. This problem is essentially the same problem as finding the facet normals of the convex hull of a set of points in the $(d-1)$-dimension space, which can be computed in time $O(m^{\lceil d/2 \rceil - 1})$ using algorithms from computational geometry (Seidel 1997), where m is the number of the elements of W, and $d = 2^n - 1$ is the dimension of the space (which is the number of the scaling coefficients).

Potential Optimality. We can also test for potential optimality in a straightforward way. Given a set of decision alternatives $\{a_1, a_2, ..., a_m\}$ resulting in the corresponding prospects $\{p_1, p_2, ..., p_m\}$, decision alternative a_r is potentially optimal if $\mathbf{t}p_r - \mathbf{t}p_s \notin C_W, \forall s = 1, 2, ..., m, s \neq r$, or, equivalently, the polyhedral cone generated by $\{\mathbf{t}p_r - \mathbf{t}p_s | s = 1, 2, ..., m, s \neq r\}$ intersects with C_W at the origin only.

Consistency Checking. In complex decision making problems, the decision maker may easily exhibit inconsistent preferences. For example, she may assert a set of comparative statements that results in an empty preference cone ($K = \emptyset$). In such situations, we would like the system to restore consistency by eliminating one or more "problematic" comparative statements. But how to identify such problematic statements is an open question and needs further research and experiment.

3.4. Integrating Qualitative Comparative Statements

In the previous section we have shown how comparative statements about decision consequences can be exploited to test for induced dominance and potential optimality. Since there are $2^n - 1$ unknown scaling coefficients, chances are that we would need an inordinately large number of pair-wise comparisons from the decision maker in order to make useful inferences.

Qualitative preference logics such as those proposed by Doyle and Wellman (Doyle, Shoham, & Wellman 1991; Doyle & Wellman 1994) and Tan and Pearl (Tan & Pearl 1994) provide languages that can express comparative statements about classes of decision consequences. Such a qualitative expression of preferences gives us a large number of pair-wise preferences among individual decision consequences, which can be used to effectively constrain the space of utility functions.

In this section we give an example to illustrate this idea. In this example, the state space has three propositional attributes $\{X_1, X_2, X_3\}$ with domains $\{0, 1\}$. Suppose also that the attributes are utility independent in such a way that for each attribute, 1 is the preferred value. This means that the sub-utility functions $u_i(x_i)$ for the attributes are given by $u_i(x_i) = x_i$. The overall utility function can be written as

$$\begin{gathered} u(x_1, x_2, x_3) = k_1x_1 + k_2x_2 + k_3x_3 + \\ k_{12}x_1x_2 + k_{13}x_1x_3 + k_{23}x_2x_3 + k_{123}x_1x_2x_3. \end{gathered}$$

Next, we translate the assumption about utility independence of the attributes into constraints about the scaling constants. Let us consider attribute X_1. To say that X_1 is UI is equivalent to say that the overall utility function

$$\begin{aligned} u(x_1, x_2, x_3) = (k_1 + k_{12}x_2 + k_{13}x_3 + k_{123}x_2x_3)x_1 + \\ k_2x_2 + k_3x_3 + k_{23}x_2x_3, \end{aligned}$$

when viewed as a linear function of x_1 must have positive coefficient for x_1, i.e.

$$k_1 + k_{12}x_2 + k_{13}x_3 + k_{123}x_2x_3 > 0, \forall x_2, x_3.$$

Considering all value assignments for x_2 and x_3, the UI assumption for X_1 implies the following inequalities involving the scaling constants:

$$\begin{cases} k_1 & > 0 \\ k_1 + k_{12} & > 0 \\ k_1 + k_{13} & > 0 \\ k_1 + k_{12} + k_{13} + k_{123} & > 0 \end{cases} \tag{3.3}$$

The utility independence for X_2 and X_3 can be expressed similarly:

$$\begin{cases} k_2 & > 0 \\ k_2 + k_{12} & > 0 \\ k_2 + k_{23} & > 0 \\ k_2 + k_{12} + k_{23} + k_{123} & > 0 \end{cases} \tag{3.4}$$

$$\begin{cases} k_3 & > 0 \\ k_3 + k_{13} & > 0 \\ k_3 + k_{23} & > 0 \\ k_3 + k_{13} + k_{23} + k_{123} & > 0 \end{cases} \tag{3.5}$$

Note that the inequality $k_1 > 0$ is equivalent to the comparison statement $(0,0,0) \prec (1,0,0)$, and the inequality $k_1 + k_{12} > 0$ is equivalent to $(0,1,0) \prec (1,1,0)$.

Now suppose that the decision maker states that she prefers to have $(x_1 = 1, x_2 = 0)$ to $(x_1 = 0, x_2 = 1)$, all else being equal, i.e. regardless of the value of x_3. This statement is equivalent to the following constraints:

$$\begin{cases} k_1 & > k_2 \\ k_1 + k_{13} & > k_2 + k_{23} \end{cases} \tag{3.6}$$

Then using algorithm 1 with the Constraints (3.3), (3.4), (3.5), and (3.6), we will be able to obtain that, for example,

$$(1,0,1) \succ \{(.5,(0,1,1),(.3,(1,0,0)),(.2,(0,1,0)\},$$

where the right hand side is a probability distribution giving the probabilities .5, .3, .2 to the states $(0,1,1),(1,0,0),(0,1,0)$, respectively.

3.5. Related Work

The idea of representing partial preference information using polyhedral cones has appeared in work in the field of Multiple Criteria Decision Making (MCDM). In this work, decision alternatives are scored according to a finite number of criteria, and the overall score for each alternative is a (value) function of the individual scores. In this sense, all decision alternatives result in *certain* outcomes that have scores as attributes. In contrast, in this paper, the consequences of decisions are uncertain. Furthermore, in work in MCDM, the value function is usually assumed to have some tractable form such as (in increasing order of generality) linear (Zionts & Wallenius 1976), quasiconcave (Korhonen, Wallenius, & Zionts 1984; Prasad, Karwan, & Zionts 1997), or monotonic (Koksalan & Sagala 1995). In our approach, the decision maker's utility function is assumed to have multi-linear form. Since multi-linear functions can be non-monotonic (see Theorem 3), it is not immediately clear if the preference cone techniques from the work in MCDM mentioned above can be used in our approach.

The preference cone technique was also proposed by Hazen in the context of reasoning with partially specified additive or multiplicative utility functions (which are special cases of multi-linear utility functions) (Hazen 1986).

3.6. Summary and Discussion

Classical Decision Theory provides a normative framework for representing and reasoning about complex preferences. Straightforward application of this essentially quantitative theory to automate decision making is difficult due to high cost of eliciting utility functions. Recent work from the field of qualitative decision theory offers several alternative solutions. These approaches focus on developing formal languages that can express qualitative, partial preference information. However, the inference mechanisms offered by these languages remain rather weak.

It is thus highly desirable to develop a framework that can exploit different strengths of these different quantitative and qualitative approaches. In this chapter, we set out to provide such a framework. Assuming a multilinear utility function with known sub-utility functions, we show how ceteris paribus comparative statements by the decision maker can be used to infer further preferential information such as induced pairwise preference and sub-optimality. There are several open problems that future work can address.

The first issue is efficiency. Note that the time complexity of Algorithm 1 is exponential in terms of d, the number of scaling coefficients: $(O(m^{\lceil d/2 \rceil - 1}))$. The base of the time complexity, which is the number of the constraint vectors $\mathbf{w}_j$, can be large. A possible solution for this is to investigate ways to *effectively* use sets of comparisons, represented by qualitative logical constructs. For example, a ceteris paribus comparison, which is equivalent to a *set* of comparisons between individual decision consequences, can sometimes be captured by a *single* linear inequality involving some (but not all) scaling coefficients. This would result in fewer constraint vectors. With regard to the exponent of the time complexity, since in practice, the number of attributes (n) is usually very small (3-8), we expect the exponent in the time complexity (which is $\lceil d/2 \rceil - 1$) not to exceed 127. We can also try to exploit further utility independencies, if applicable, to reduce the number of scaling coefficients and to reduce the exponent. For example, if Y is UI of $X - Y$ and $|Y| = r$, then the number of scaling coefficients can be reduced to $2^r + 2^{n-r+1} - 4$. If both Y and $X - Y$ are UI of the other, then the number can be reduced to $2^r + 2^{n-r} - 2$ (Keeney & Raiffa 1976, Chapter 6.10.3). Finally, we point out that since the motivation of our work is to reduce the elicitation time, which in most cases is much larger than the computation time, this approach to reasoning with partial preference information could sometimes provide an attractive option.

Second, instead of working with only the supplied preferences, we may want the system to take the initiative and ask the user to make comparisons between decision consequences, or sets of decision consequences (using some qualitative logic constructs). To this end, the most interesting issue is identifying the questions whose answers would lead to the most conclusive inference, e.g., induced optimality of a particular decision alternative, or induced sub-optimality of a large number of decision alternatives [6]. A difficulty with this approach is that the user may not be able to answer some of those queries.

Third, it is highly desirable to provide the decision maker with some explanation for any conclusion that the system makes. For example, the user may want to know why one decision consequences is inferior to another, or why some decision consequences cannot be optimal, according to his own preference. Our framework could provide such capability to generate explanations via geometric illustration of the preference cone. The key issue here that we need to address is how to interpret the geometric concepts in such a way that the user is comfortable with.

[6] Our previous work (Ha & Haddawy 1997) has addressed this issue in a special case when the utility function is additive.

CHAPTER 4
A CASE-BASED APPROACH TO REASONING WITH PARTIAL PREFERENCES

Imagine that Ms. Xaviera (let's call her X), an avid cineaste, is watching the Ebert & Roeper show for their reviews of *Kiss of the Dragon*, a recently released movie. Ebert gives it a thumb-up, but Roeper a thumb-down. Who should X listen to in deciding whether to see the movie? While both are great film critics with whom she agrees most of the time, recently X tends to agree more with Ebert, and thus decides to go out and see the movie. X's preference is more "similar" to Ebert's than to Roeper's.

This approach to decision making is ubiquitous in our everyday life. We listen to advice, judgement, recommendations from people with whom we share common interests, tastes. In addition, research in market segmentation (Frank, Massy, & Wind 1972) showed that people tend to form *clusters* according to their preferences. Starting from this simple observation, the fledgling area of collaborative filtering attempts to build systems that recommend items of interest (e.g. movies, music, books, news articles, etc.) to people in a virtual community. Each user in the community rates various alternatives according to a numeric scale. The filtering system correlates the ratings in order to determine which users' ratings are most similar to each other. Finally, the system predicts how well users will like new items based on ratings from similar users.

It is this recent work in collaborative filtering that inspired our efforts to develop a case-based approach to eliciting and reasoning with partial preferences. In Chapter 3, we have seen that the process of eliciting and reasoning with preferences is usually aided with various kinds of assumptions, mostly preferential/utility independence assumptions. But even with the leverage obtained by making various independence assumptions about user preference, elicitation of utility functions can still be too time-consuming. Instead of performing the elicitation in a vacuum, it would be

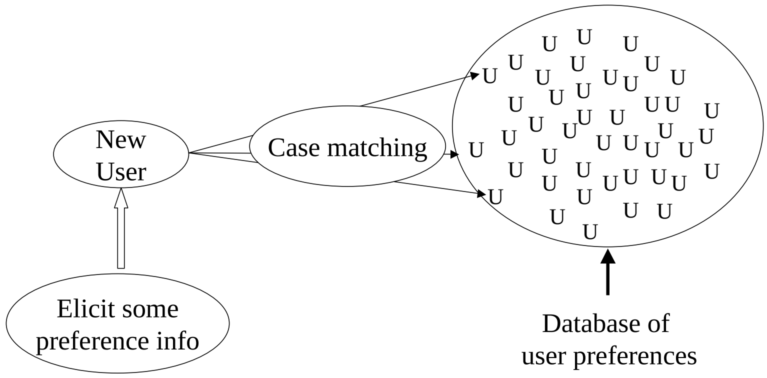

Figure 4.1. The Case-Based Approach to Preference Elicitation

useful if we could augment directly elicited preferences with some appropriate default information. The case-based approach to preference elicitation is a perfect candidate for this purpose. Assuming the existence of a population of users from whom we have elicited complete or incomplete preferences, we propose eliciting the preferences of a new user interactively and incrementally, using the closest existing preference structures as potential defaults. We envision our system to maintain a population of users with their preferences partially or completely specified in a given domain. When encountering a new user A, the system elicits some preference information from A and then determines which user in the population has the preference structure that is closest to the preference structure of A. The preference structure of that user will be used to determine an initial default representation of A's preferences. Figure 4.1 provides a schematic view of this approach.

A key issue common to this approach as well the work in collaborative filtering is the choice of a distance measure on preference orders. Existing distance measures on preferences suffer from a number of limitations and thus are unsuitable for our case-based preference elicitation approach. To address this problem, we introduce a novel distance measure, called the *probabilistic distance*, according to which the distance between two preference orders is determined by the probability that they disagree in their relative rankings of two randomly picked decision consequences. This Chapter's main focus is on the probabilistic distance: its inception, properties, computation, and several applications, some of which are not related to case-based preference elicitation.

The rest of this Chapter is organized as follows. We start out by discussing existing approaches to defining distance measures on preference orders, highlight their inadequacies for the purpose of the case-based approach, and introduce the probabilistic distance. We then discuss the probabilistic distance in the case of certainty and in the case of uncertainty separately. We conclude with a summary of the main contributions of this chapter, and discuss future work.

4.1. The Probabilistic Distance on Preference Orders

So Ms. Xaviera listened to Mr. Ebert's recommendation and went to see *Kiss of the Dragon*, and enjoyed it. Her preference is perhaps closer to Mr. Ebert's than Mr. Roeper's. Perhaps subconsciously, Xaviera has a measure of similarity between her cinematic taste and those of Ebert and Roeper. How should we capture this measure? How exactly should preference similarity be modeled? [1]

Until now, similarity measures used in existing recommender system are defined on rather sketchy models of preference. For example, Amazon.com has a feature called *Customers who Bought* that suggests a user browsing a particular item A a list of items

[1] Technically, distance measures should be referred to as *dissimilarity* measures, in reverse scale against similarity measures. For the sake of simplicity, we use distance measures and similarity measures interchangeably whenever no misunderstanding is foreseen.

bought by users that bought A. The assumption here is that people who express some form of interest in a common item (by buying, or just browsing) may have common interest in other items as well. In other examples, systems such as Grouplens usenet article recommender (1994), the Ringo music recommender (Shardandand & Maes 1995), and the Bellcore video recommender (1995), use a similarity measure that is defined on explicit numeric ratings that people assign to items. While numeric ratings provide a more accurate model to capture preferences than the naive *Customers who Bought*, they are still a far cry from a model capable of representing complex aspects of preferences such as attitude toward risk, and tradeoffs among competing objectives. For example, if a person ranks a movie as a 5 on a 1-5 scale and later decides he likes some other movies better, there is no way to accommodate this, without possibly having to reassign ratings to many movies. A similarity measure based on a numeric rating scheme is unable to accommodate even a simple pairwise preferential statement such as "Xaviera likes *Black Orpheus* more than *Lambada, the forbidden dance*".

In this thesis, we study the issue of defining similarity measures on preferences from a fundamental, decision-theoretic point of view. Preferences are represented by preference orders, or consistent value/utility functions. Two requirements that mainly concern us are 1) extensibility to accommodate *partial* preferences and 2) amenability to efficient computation. While the second requirement is self-evidently important, the first requirement is important because most of the time, the available models of user preferences are incomplete. The probabilistic distance, our proposed measure has its roots in the Kendall's tau function, a well-known concept in the statistical literature [2]. In a nutshell, the probabilistic distance between Xaviera's cinematic taste and that of Ebert is the probability that they disagree on a "randomly chosen" pair of movies (i.e. Xaviera prefers, say, the first to the second, while Ebert prefers the second to the first).

4.1.1. Probabilistic Distance on Complete Preferences

We first formally define the probabilistic distance on complete preference orders. Let the *conflict indicator function* $c_{\prec_1,\prec_2} : \mathcal{D}^2 \to \{0,1\}$ be defined as follows:

$$c_{\prec_1,\prec_2}(a,b) := \begin{cases} 1 & \text{if } (a \precsim_1 b \wedge b \prec_2 a) \vee (a \prec_1 b \wedge b \precsim_2 a) \\ & \vee\ (a \precsim_2 b \wedge b \prec_1 a) \vee (a \prec_2 b \wedge b \precsim_1 a) \\ 0 & \text{otherwise.} \end{cases}$$

The probabilistic distance is defined as:

$$\delta(\prec_1,\prec_2) := \Pr(c_{\prec_1,\prec_2}(a,b) = 1) = E[c_{\prec_1,\prec_2}(a,b)]. \tag{4.1}$$

[2]Although at the time of the conception of the probabilistic distance, we ourselves were not aware of the existence of Kendall's tau function.

Here the expectation is taken with respect to a and b, which are two independent identically distributed uniform random variables on $\mathcal{D}$. The following result says that the probabilistic distance on complete preference orders is a metric.

Theorem 4. *In the case of certainty, the probabilistic distance on the set of weak orders on Ω is a metric with range $[0,1]$. In the case of uncertainty, the probabilistic distance on the set of "rational" weak orders on $\mathcal{S}$ is a metric with range $[0,1]$.*

Proof Recall that the conditions for being a metric are:

1. **Reflexivity.** $\delta(a,b) \geq 0$, "=" iff $a = b$.
2. **Symmetry.** $\delta(a,b) = \delta(b,a)$.
3. **Triangle Inequality.** $\delta(a,b) + \delta(b,c) \geq \delta(a,c)$.

It is evident that the probabilistic distance only takes values between 0 and 1, the distance between two identical orders is zero, and zero distance implies two identical weak orders. The symmetry of the distance function trivially follows from the symmetry of the conflict function. Finally, to prove the triangle inequality, we note that for all weak orders $\prec_i, i = 1,2,3$ and alternatives a, b, $c_{\prec_1,\prec_3}(a,b) = 1$ implies either $c_{\prec_1,\prec_2}(a,b) = 1$ or $c_{\prec_2,\prec_3}(a,b) = 1$, and for all events X, Y, $\Pr(X \vee Y) \leq \Pr(X) + \Pr(Y)$.

4.1.2. Probabilistic Distance on Partial Preferences

Partial Preference Orders. How should we represent partial preferences? For the purpose of the case-based preference elicitation, a partial preference of a person is obtained via an incomplete elicitation. For example, in the case of certainty, we encounter partial preferences when a person assigns ratings to a *subset* of the set of decision outcomes. In the case of uncertainty, we may have determined that the utility function of a person has a certain parametric form (e.g. multiplicative form), but have yet to assess some parameters (e.g. a scaling coefficient, or a sub-utility function of a multi-attribute utility function).

For the most generality, we assume that a partial preference order $\prec$ is a *binary relation* on the set $\mathcal{D}$ of decision consequences. Furthermore, it is reasonable to assume that this binary relation is *asymmetric*: if we know that a person prefers a to b, then it is not the case that he prefers b to a. We may also assume *transitivity*: if he prefers a to b, and b to c, then he prefers a to c. In the theory of orders, an asymmetric, transitive binary relation is a called a *partial order*, or a *poset*. We thus represent partial preferences using partial orders[3].

[3]Note that the difference between the definition of complete preference order and that of partial preference order is the difference between negative transitivity and transitivity. Given asymmetry,

We have a slightly different concept of consistent functions for partial orders. A real-valued function $f : \mathcal{D} \rightarrow \Re$ over the decision consequences is said to be *consistent* with a partial preference order $\prec$ if for any decision consequences a, b, $a \prec b \Rightarrow f(a) < f(b)$ and $a \sim b \Rightarrow f(a) = f(b)$. The set of all functions that are consistent with $\prec$ is denoted as $C(\prec)$. Intuitively, consistent functions capture all information contained in the partial orders, and they might contain more than that. Consequently, functions that are consistent with a partial preference order $\prec$ may be strategically different, as they induce weak orders that are different extensions of $\prec$. There is however a one-to-one correspondence between the weak order extensions of $\prec$ and the equivalence classes of $(C(\prec), \simeq)$.

We can extend the definition of probabilistic distance to partial orders in the following way. Let $\prec_1$ and $\prec_2$ be two partial orders with corresponding sets of weak order extensions E_1 and E_2. Recall that E_i can be viewed as a set of strategically different value/utility functions f_i consistent with $\prec_i$, for $i = 1, 2$. These functions form a one-to-one correspondence with the weak order extensions of $\prec_i$ (note that in the uncertainty case, the correspondence is with only extensions that satisfy the "rational properties" required for the existence of a utility function). We define the probabilistic distance $\delta(\prec_1, \prec_2)$ to be the average of the probabilistic distance between pairs of extensions of $\prec_1$ and $\prec_2$, respectively. Formally,

$$\begin{aligned} \delta(\prec_1, \prec_2) &= E\left[\delta(\prec_{f_1}, \prec_{f_2})\right] \\ &= E\left[E[c_{\prec_{f_1}, \prec_{f_2}}(a, b)]\right], \end{aligned}$$

where f_i are uniform random variables on E_i, $i = 1, 2$, and a and b are uniform random variables on $\mathcal{D}$. Note that this distance is *not* a metric on the set of partial orders, since the distance between two identical partial orders that are not complete orders is always positive (which violates the "distinguishability of non-identicals" property). This, however, is desirable if the two orders represent the preferences of two different users, since the complete preference orders for the two may actually differ.

transitivity is weaker than negative transitivity, i.e. the latter implies the former. This "weakness" reflects the incompleteness of our information about the person's preference. For any pair of decision consequences $\{a, b\}$, *exactly one* of the following four scenarios can arise: 1) $a \prec b$, 2) $b \prec a$, 3) $a \sim b$, and 4) $a \parallel b$. The interpretations for the first three scenarios are 1) b is preferred to a, 2) a is preferred to b, 3) a and b are equally preferred (indifference). The last scenario corresponds to the case when the decision maker is not sure about his preference regarding the two decision consequences a and b. (The general interpretation is that they are incomparable.) Note that this *lack of preference* is not the same as *indifference* (case 3)).

4.2. The Case of Certainty

4.2.1. The Probabilistic Distance on Complete Preferences

When the decision problem does not involve uncertainty, the distance $\delta(\prec_1, \prec_2)$ can be computed by averaging the conflict function $c_{\prec_1,\prec_2}(i,j)$ over all n^2 pairs $(i,j) \in \Omega^2$. In the case when both $\prec_1$ and $\prec_2$ are strict orders with corresponding canonical value functions π_1, π_2, we have:

$$\delta(\prec_1, \prec_2) = \frac{\tau(\pi_1, \pi_2)}{n^2},$$

where $\tau(\pi_1, \pi_2)$ is *Kendall's tau function* (Kendall 1962) which simply returns the number of conflicts. We can divide the Kendall's tau function by $n(n-1)$ instead of n^2 in order to properly scale the probabilistic distance to the range of [0,1]. According to that scale, the distance between a strict order and its complete reverse is 1.

Example 1. *Suppose that there are 3 decision outcomes* ($\Omega = \{1, 2, 3\}$) *and* $1 \prec_1 2 \prec_1 3$ *and* $2 \prec_2 3 \prec_2 1$*. Then* $\delta(\prec_1, \prec_2) = 4/9$*, since* $\{(1,3),(3,1),(2,3),(3,2)\}$ *are the four pairs of outcomes that cause conflict between* $\prec_1$ *and* $\prec_2$*.*

Another popular metric on the set of permutations of $\{1, 2, \ldots, n\}$ is *Spearman's rank order correlation coefficient*, or *Spearman's rho* (Spearman 1904):

$$\rho(\pi_1, \pi_2) = \frac{\sum_{i=1}^{n}(\pi_1(i) - \frac{n+1}{2})(\pi_2(i) - \frac{n+1}{2})}{\frac{n^2-1}{12}}. \tag{4.2}$$

Here, $\frac{n+1}{2}$ is the mean, and $\sqrt{\frac{n^2-1}{12}}$ is the standard deviation of the both π_1 and π_2. Spearman's rho is often computed using the following form:

$$\rho(\pi_1, \pi_2) = 1 - \frac{6R^2(\pi_1, \pi_2)}{n^3 - n},$$

where $R(\pi_1, \pi_2) = (\sum_{i=1}^{n}(\pi_1(i) - \pi_2(i))^2)^{1/2}$ is the Euclidean distance between the two vector π_1 and π_2. Spearman's rho and Kendall's tau have been studied extensively in the statistical literature. Other commonly used metrics include:

Spearman's footrule (Spearman 1906):

$$F(\pi_1, \pi_2) = \sum_{i=1}^{n} |\pi_1(i) - \pi_2(i)|. \tag{4.3}$$

Ulam's distance (Ulam 1981):

$$U(\pi_1, \pi_2) = n \text{ - the max number of items ranked in the same order by } \pi_1 \text{ and } \pi_2.$$

Kemeny's distance (Kemeny 1959):

$$K(\pi_1, \pi_2) = \text{the min number of pairwise inversions to obtain } \pi_2 \text{ from } \pi_1.$$

Ulam's distance is used in DNA research to measure the distance between two strings of molecules. Kemeny's distance is used to define *Kemeny ranking*, which is a ranking that aggregates a set of rankings in such a way that minimizes the total Kemeny's distances from the members of that set. Kemeny's ranking is often used in social choice theory as the best compromise between the possibly conflicting views of the judges (e.g. judges in figure skating). See Critchlow (Critchlow 1980) for a discussion these metrics from a statistical point of view.

These metrics can be used as distance measures on strict orders. But since their definitions are all based on the canonical value functions of *strict* orders, it is not straightforward to extend these metrics to define similarity measures on *weak* orders. Furthermore, while it is possible to define distance measures between two weak orders $\prec_1$ and $\prec_2$ based on Spearman's rho (see Equation 4.2) or Spearman's footrule (see Equation 4.3) using some functions π_1 and π_2 that are consistent with $\prec_1$ and $\prec_2$, respectively, the resulting measures will apparently be sensitive to the choice of those functions. For example, in the equation for Spearman's rho (Equation 4.2), $\pi_1(i)$ and $\pi_2(i)$ may be replaced with ratings that the two persons assign to item i, and $\frac{n+1}{2}$ may be replaced with corresponding means $\overline{\pi}_1, \overline{\pi}_2$ and $\sqrt{\frac{n^2-1}{12}}$ with the corresponding variances σ_1, σ_2 of these ratings. The resulting p is the well-known *Pearson's correlation coefficient*.

$$p(\pi_1, \pi_2) = \frac{\sum_{i=1}^{n}(\pi_1(i) - \overline{\pi}_1)(\pi_2(i) - \overline{\pi}_2)}{\sigma_1 \sigma_2}.$$

The Pearson correlation coefficient measures the degree to which a *linear* relationship exists between two variables (or in this case, two sets of ratings), and thus may be unsuitable as a similarity measure between two variables having a close but non-linear relationship. In addition, the Pearson correlation coefficient does not meet the "distinguishability of nonidenticals" and the "triangle inequality" property. Many researchers have argued against the routine use of similarity measures that do not meet the requirements of a metric (Jardine & Sibson 1971; Clifford & Stephenson 1975). Nevertheless, the Pearson's correlation coefficient and several of its mutants have been used as similarity measures quite extensively in many fields. Within AI, notable recommender systems that make use of the Pearson correlation coefficient include the Grouplens' collaborative filtering system (Resnick *et al.* 1994; Konstan *et al.* 1997), the Ringo music recommender (Shardandand & Maes 1995), and the Bellcore video recommender (Hill *et al.* 1995).

4.2.2. The Probabilistic Distance on Partial Preferences

Let $\prec_1, \prec_2$ be two partial orders with corresponding sets of weak order extensions E_1, E_2. Recall that the probabilistic distance is defined as:

$$\begin{aligned}\delta(\prec_1, \prec_2) &= E\left[\delta(\prec_{v_1}, \prec_{v_2})\right] \\ &= E\left[E[c_{\prec_{v_1}, \prec_{v_2}}(i,j)]\right],\end{aligned}$$

A simplistic approach to compute the probabilistic distance would be to evaluate the conflict function c for all possible 4-tuples $\{(f_1, f_2, i, j) | f_1 \in E_1, f_2 \in E_2, i, j \in \Omega\}$ and take the average. This however is impractical because the number of weak order extensions of a partial order can be exponentially large (the number of strict order extensions of a vacuous partial order - a partial order in which everything is incomparable with everything - is $n!$). In fact, the much easier problem of *counting linear extensions* [4] of finite posets was shown to be #P-complete [5] (Brightwell & Winkler 1991). Given the hardness of counting and generating linear extensions, we turn to approximation techniques to estimate $\delta_p(\prec_1, \prec_2)$. For example, we can use the Monte Carlo simulation method to estimate $\delta_p(\prec_1, \prec_2)$, provided that we have an efficient algorithm to generate v_i uniformly randomly from V_i.

It turns out that counting (approximately) and generating (uniformly randomly) elements of large combinatorial sets are two closely related problems. In fact, Sinclair (Sinclair 1993) showed that an efficient algorithm for one problem can be used to construct an efficient algorithm for the other, provided the combinatorial sets have a certain structural property called *self-reducibility.* The set of linear extensions of a poset has this property and, not surprisingly, a number of algorithms for generating (almost) uniformly randomly linear extensions of posets have been developed (Karzanov & Kachiyan 1991; Bubley & Dyer 1998). These algorithms are all randomized algorithms based on the Markov chain Monte Carlo technique [6]. In Appendix A we describe the best known algorithm, due to Bubley and Dyer (Bubley & Dyer 1998) that has a running time of $O(n^3 \log n\epsilon^{-1})$, where n is the poset's cardinality, and ϵ is the desired accuracy.

Now with the help of the routine that almost uniformly randomly generates linear

[4]This is a fundamental problem in the theory of ordered sets with applications in computer science (sorting) and social sciences.

[5]The complexity class #P, introduced by Valiant (Valiant 1979), consists of all counting problems whose solutions are the number of accepting states of some non-deterministic polynomial-time Turing Machine. A counting problem is *#P-complete* if the problem of counting the number of satisfying assignments to a 3-SAT problem can be reduced to it in polynomial time. #P-complete problems, which are analog counting counterparts of NP-complete problems, are considered very difficult, especially in the view of Toda's results (Toda 1989), which implies that one call to a #P-complete oracle suffices to solve any problem in the polynomial hierarchy in deterministic polynomial time.

[6]See (Jerrum & Sinclair 1996) for a recent survey of this method.

extensions of a poset, we can estimate $\delta_p(\prec_1, \prec_2)$ by randomly generating $v_{ij} \in V_i(i = 1, 2; j = 1, \ldots, k)$, computing $\delta_p(v_{1j}, v_{2j}), j = 1, \ldots, k$, and taking the sample mean $\hat{\delta}_p = \frac{1}{k}\sum_{j=1}^{k} \delta_p(v_{1j}, v_{2j})$. This sample mean is an unbiased estimator [7] of $\mathrm{E}\ Y = \delta_p(\prec_1, \prec_2)$ with variance $(\mathrm{Var}\ Y)/k$. We can derive a confidence interval for δ_p as follows. Let t be the ratio of Y's variance and square of its expectation: $t = \mathrm{Var}\ Y/(\mathrm{E}\ Y)^2$, a non-negative quantity that can usually be bounded above by τ, which is polynomial in terms of n, the input size. Thus $\mathrm{Var}\ Y \leq \tau(\mathrm{E}\ Y)^2$ and $\mathrm{Var}\ \hat{\delta}_p \leq (\tau(\mathrm{E}\ Y)^2)/k = (\tau\delta_p^2)/k$. For any positive number c, Chebysev's inequality states that:

$$\Pr((\hat{\delta}_p - \delta_p)^2 > c\mathrm{Var}\ \hat{\delta}_p) \leq 1/c,$$

and thus:

$$\Pr((\hat{\delta}_p - \delta_p)^2 > c\tau\delta_p^2/k) \leq 1/c,$$

or equivalently:

$$\Pr((1 - \sqrt{c\tau/k})\delta_p \leq \hat{\delta}_p \leq (1 + \sqrt{c\tau/k})\delta_p) \geq 1 - 1/c.$$

As a consequence, if we want our estimator $\hat{\delta}_p$ to be within a multiplicative factor of $1 + \epsilon$ of δ_p with probability of at least $1 - 1/c$, it is sufficient to take a sample of size $k = \lceil 4c\tau/\epsilon^2 \rceil$.

4.2.3. Some Applications of the Probabilistic Distance

The Decision-Theoretic Video Advisor

Nguyen and Haddawy (1999) describe the Decision-Theoretic Video Advisor (DIVA), an interactive, low-stake decision support systems. DIVA implements the case-based preference elicitation approach, as outlined at the beginning of this chapter. It uses the probabilistic distance as a measure of similarity between user preferences. Partial preference orders of users consist of pair-wise preferences, which are obtained by having the user classify movies into like, ok, and dislike categories. In addition, DIVA can differentiate user's long- and short-term preference. It is also able to incorporate feedback from A and revise its representation of A's preference to provide revised, better recommendation.

The experiments with DIVA were performed using the EachMovie database of 72916 users and 1628 movies, courtesy of the Digital Equipment Corporation. The results

[7] Strictly speaking, $\hat{\delta}_p$ is *not* an unbiased estimator for δ_p, since the routine only generates *almost* uniform linear extensions. The incurred bias is insignificant and often simply ignored in Markov chain Monte Carlo analysis.

showed that the probabilistic distance provided better performance in term of both *precision* and *recall*, two metrics commonly used in information retrieval (Salton & McGill 1983) [8]. The precision of DIVA using the probabilistic distance is 85%, versus 65% when using the Pearson's correlation coefficient. The recall of DIVA using the probabilistic distance is 40%, versus 35% when using the Pearson measure. This confirms our intuition that this simple generalized Pearson correlation coefficient is inaccurate as a measure of similarity between two partial ratings.

Preference Elicitation via Theory Refinement

Another practical application of the probabilistic distance is found in the work of Haddawy et. al. (2001) that is not related to case-based preference elicitation.

To reduce the complexity of preference elicitation, traditional approaches from Decision Analysis and Multiple Criteria Decision Making make assumptions concerning the structure of preferences (e.g. monotonicity or independence) and then perform elicitation within the constraints of those assumptions. But inaccurate assumptions can result in inaccurate elicitation. Nevertheless, assumptions can be a useful guide if they at least approximately apply to some large segment of the population. Ideally we need a method of using assumptions to guide but not constrain the elicitation process. This kind of functionality is provided by theory refinement techniques. The basic idea behind theory refinement is that we can start with a domain theory that may be approximate and incomplete and then correct for inaccuracies and incompleteness by training on examples. If the domain theory is at least approximately correct, we can learn faster with it than without it.

Haddawy et al. (2001), explore the use of one particular theory refinement technique, Knowledge-Based Artificial Neural Networks (KBANN), to learn user preferences. In the case of certainty, they describe the problem of choosing a flight, where it is reasonable to make several assumptions about preferential independence and monotonicity. They then show how to represent these assumptions as Horn-clause theories that can be encoded in a KBANN network. This KBANN network can be trained to learn fine-grained preference structures from a variety of preferential data, including numeric ratings and simple binary classification. One of the main hypothesized advantages of the KBANN technique is its robustness to noise: the domain theory only needs to be approximately correct for KBANN to be useful. To evaluate this hypothesis in our flight selection domain, Haddawy et al. examine the performance of KBANN in learning preferences using examples generated from a number of value functions that violate the independence assumptions to various degrees. It is expected that the more a value function violates the domain theory, the worse the performance. The

[8] In the context of recommending movies, precision indicates how many movies in a recommended list were actually liked by the user. Recall indicates how many movies out of all movies liked by the user were predicted correctly.

hypothesis is confirmed if this decrease in performance does not occur in a precipitous manner.

The main issue to be addressed in this robustness analysis is the definition of the *degree of violating the independence assumptions*, or *DOVI*. Haddawy et al. define this measure in the following way. Given a domain theory $\mathcal{D}$ and a value function u, note that $\mathcal{D}$ can be viewed as a set of value functions that satisfy $\mathcal{D}$. The DOVI measure of u violating $\mathcal{D}$ is defined to be the distance between u and the member of $\mathcal{D}$ that is closest to u:

$$V_{\mathcal{D}}(u) = \min_{f \in \mathcal{D}} \delta(f, u),$$

where δ is the probabilistic distance between value functions. Imagine this approach as an analogy to the definition of a distance from a point to a set of points in Euclidean geometry. This definition of degree of violating the domain theory is *semantic*, a departure from existing *syntactic* approaches used in robustness analysis of theory refinement techniques.

Figure 4.2 shows the results of Haddawy et al.'s robustness analysis using seven preference orders with DOVI varying from 0.05 to 0.92. They analyze the performance of KBANN on these seven preference orders, using a training set size of 30, 50, 100, and 150, and keeping the test set size constant at 50. Notice that for any given number of training examples (30, 50, 100, 150), the performance of KBANN decreases as the DOVI increases. KBANN's performance when the DOVI is 0.24 is still very close to when the DOVI is 0.05. This implies that KBANN is fairly robust to some small amount of noise in the domain theory. The sharpest performance decrease occurs when the DOVI hits the range of 0.3 to 0.35. The fewer examples we use to train KBANN, the sharper this performance decrease. Fewer training examples means that KBANN is relying more on the (inaccurate) domain theory. In addition to confirming the robustness hypothesis, this experiment also justifies the probabilistic distance as a proper measure of distance on preferences: if it were not, the performance curves of KBANN would have been much more random.

4.2.4. Related Work

We now review some related work on similarity measures on partial orders. To our knowledge, there is no general theory that addresses the problem of defining distance measures on partial orders. The closest to such theory we found is Critchlow's monograph (Critchlow 1980). Recall that on the strict orders on a finite set, there are a number of well-studied metrics such as Spearman's rho, Spearman's footrule, Kendall's tau, Ulam's distance, Hamming distance, and Cayley's distance. In (Critchlow 1980), strict orders are referred to as *fully ranked data*, since they fully rank a set of items of interest. Critchlow extended the aforementioned six metrics to *partially ranked data*.

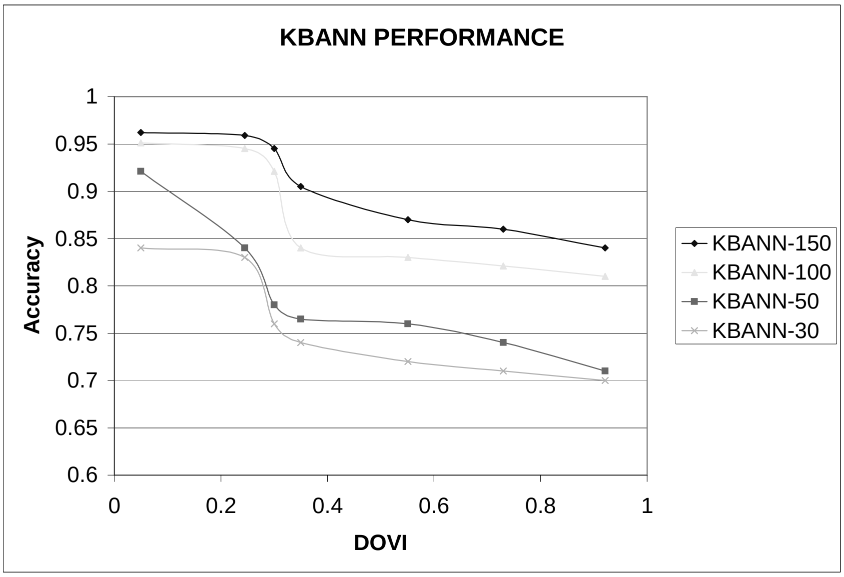

Figure 4.2. Performance of KBANN for value functions of various degrees of violation of the independence domain theory. The number of training examples are 30, 50, 100, and 150. The size of the set of test examples is kept constant at 50.

Here, partially ranked data refer to certain *special cases* of partial orders. For example, they may correspond to the case when a person lists his first through k-th choices, where $k < n$ (n is the number of decision outcomes). Note that if fully ranked data are identified with the elements of the permutation group (or, in layman's term, the permutations of Ω), then the above partially ranked data can be identified with points in a *coset space* of the permutation group. As a consequence, this restriction of partial orders facilitates several group-theoretic techniques and thus makes the extended six metrics more amenable to analysis and computation.

Without a general theory of metrics on partial orders, researchers often extend metrics such as the Pearson correlation coefficient to partial orders in some simple way. For example, consider the Grouplens collaborative filtering system. Each user of the system have rated a set of news articles, and different users have rated different sets of articles. The similarity weight between two users is taken to be the Pearson correlation coefficient between the ratings over the articles both have rated (i.e. the intersection of the two rating sets). This solution could be unsatisfactory because it is insensitive to the number of articles rated by both users. Two users having the same rating on the only one article they have both rated would be maximally correlated, while their preferences may conceivably be quite different. This intuition is confirmed by the experiments with DIVA (Nguyen & Haddawy 1999) (see Section 4.2.3 for a description of DIVA and summary of the findings). Recent research on Grouplens acknowledged this difficulty and proposed a "significance weighting" scheme to account for the size

s of the intersection set (Herlocker *et al.* 1999). Basically, the Pearson correlation coefficient is multiplied with a significance weight of $s/50$ if $s < 50$. The modified similarity measure was empirically shown to provide more accurate recommendation (Herlocker *et al.* 1999). Another approach to address the problem of small intersection was proposed by Breese et al (Breese, Heckerman, & Kadie 1998). In this approach, the correlation is computed over the *union*, instead of the intersection of the two sets of ratings. This is made possible by assigning some default ratings to items that are in the union but not in the intersection (i.e. the symmetric difference of the two sets).

4.3. The Case of Uncertainty

In the area of collaborative filtering, recommender systems such as GROUPLENS (Resnick *et al.* 1994) and the DIVA video recommender (Nguyen & Haddawy 1999) all require the use of a distance measure on preferences. Because all of these systems concern with decision making under *certainty*, it is not clear whether a study of distance measures on preferences is warranted in the case of *uncertainty*. We argue that it is. The concept of "how different is my preference from yours" is intuitive, but far from well-understood, especially when the preferential information is incomplete, or the choices are uncertain, or both. We shall now describe an example to illustrate this point.

Miyamoto and Eraker (Miyamoto & Eraker 1989) described a psychology experiment with 44 undergraduate students at the University of Michigan. The experiment is designed to test several assumptions about people's preferences and attitudes towards risks with regards to survival duration. The subjects were asked to assign certainty equivalences (CE) to a total of 42 standard gamble questions involving duration of survival. Below is a typical question:

> *For any non-negative number n, let n be the event that you will live exactly n more years in good health, and then have a sudden and relatively painless death. Let $(m, .5, n), 0 \leq m < n$, be a lottery of 50% chance for m and 50% chance for n. What is the number p for which you regard $(m, .5, n)$ and p as equivalent (denoted $(m, .5, n) \sim p$)?*

Suppose that u denotes the utility function of a subject. Each answer of the form $(m, .5, n) \sim p$ translates into the following constraint on u: $u(m)+u(n) = 2u(p)$. Thus for each subject, we have a set of 42 constraints on his/her utility function u. Given two subjects with utility functions u and u', how should we define a distance measure between u and u'? A simplistic approach may use some well-known statistical measures such as Spearman's footrule, Ulam distance, or various correlation coefficients. The problem with this approach is twofold. First, it typically requires that the constraints on u and u' are obtained from exactly the same set of CE questions, which substantially reduce its applicability. Second, this approach has to address sensitivity issues with respect to additional available constraints. Another possible approach is to completely determine u and u' (using methods such as interpolation, curve-fitting, or parameter estimation), and compute the distance on two completely specified utility functions. We believe that because of the strong assumptions required to compute the complete utility functions, the suitability of this approach can only be determined on a case-by-case basis. As we shall show in this chapter, the probabilistic distance provides a principled solution for this problem that can be used in a wide range of other problems as well.

4.3.1. The Probabilistic Distance on Complete Preferences

Let $\prec_1$ and $\prec_2$ be two preference orders on the set $\mathcal{S}$ of prospects. The probabilistic distance is defined as:

$$\begin{aligned}\delta(\prec_1, \prec_2) &= E[c_{\prec_1,\prec_2}(p,q)] \\ &= \int_{\mathcal{D}}\int_{\mathcal{D}} c_{\prec_1,\prec_2}(p,q)\partial p \partial q.\end{aligned}$$

where p and q are two independent identically distributed uniform random variables on the set $\mathcal{D}$ of decision consequences [9].

Example 2. *Let $\Omega = \{1,2,3\}$ and $\prec_1$ and $\prec_2$ be two preference orders on the set $\mathcal{S}$ of all prospects over Ω. Suppose that u_1 and u_2 are two utility functions that are consistent with $\prec_1$ and $\prec_2$ respectively, and $u_1 = (0,1,2)$ and $u_2 = (0,2,3)$. Then $\delta(\prec_1, \prec_2) = 1/9$.*

When $\mathcal{D} = \mathcal{S}$, i.e. the set of decision consequences is the same as the set of all prospects, computing the above integral amounts to computing the volume of a polytope in the $(2n-2)$-dimension space (both p and q have $n-1$ coordinates that can vary). While computing the exact volume of a polytope in general is computationally complex (Barany & Furedi 1986; Elekes 1986), there is a simple Monte Carlo approximation algorithm for this particular problem. This algorithm works by sampling $p^{(i)}, i = 1,2,\ldots,k$ and $q^{(i)}, i = 1,2,\ldots,k$ according to the uniform distribution on $\mathcal{S}$, and taking the average $\overline{c} = \frac{1}{k}\sum_{i=1}^{k} c_{\prec_1,\prec_2}(p^{(i)}, q^{(i)})$. With a sufficiently big sample size k, the sample mean $\overline{c}$ can approximate $\delta(\prec_1, \prec_2)$ with arbitrary degree of precision, according to the Central Limit Theorem. Sampling $p^{(i)}$ and $q^{(i)}$ according to the uniform distribution on $\mathcal{S}$ is basically the well-studied problem of *random division of the*

[9] Note that in the case of uncertainty, the set $\mathcal{D}$ of decision consequences is a simplex in a multi-dimensional Euclidean space. A uniform random variable on $\mathcal{D}$ can be defined using the standard method of measure and probability theory.

Table 4.1. Algorithm for uniform sampling on $\mathcal{S}$.

1. *Generate* $n-1$ *numbers* $x_i, i = 1,2,...,n-1$ *according to* $n-1$ *independent uniform random variables on* $[0,1]$.

2. *Sort* x_i*'s:* $0 \leq x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(n-1)} \leq 1$. *This is the* order statistics *of the sample. Let* $x_{(0)} = 0$ *and* $x_{(n)} = 1$.

3. *Let* $p_i = x_{(i)} - x_{(i-1)}, i = 1,2,\ldots,n$. *(*$p_i$*'s are called the* spacings *of the sample). Return* $(p_1, p_2, \ldots, p_n)$.

unit interval and can be performed using the algorithm in Table 4.1. See (Pyke 1965; David 1981) for more details.

The probabilistic distance between two preference orders, defined this way, depends only on the orders. It can be computed given the two orders, or two utility functions that are consistent with the two orders. This definition can be useful when the two preference orders, or the two consistent utility functions are given, but little is known about the available decision alternatives. When we have more information about the decision alternatives and their consequences, it is desirable that we tailor the definition of the probabilistic distance to reflect this knowledge. So if the set $\mathcal{D}$ of decision consequences is finite and known, the probabilistic distance can be defined as:

$$\delta(\prec_1, \prec_2) = \frac{\sum_{(p,q)\in\mathcal{D}^2} c_{\prec_1,\prec_2}(p,q)}{|\mathcal{D}|^2}. \tag{4.4}$$

The computation of this (discrete) formula is obviously much simpler than the integral formula of Equation 4.4, *provided that* we know the set of decision alternatives $\mathcal{D}$ [10]. Also, it is a subtle issue to determine which decision alternatives to include in $\mathcal{D}$ in the above definition.

4.3.2. The Probabilistic Distance on Partial Preferences

Let $\prec_1$ and $\prec_2$ be the partial preference orders of two persons, A_1 and A_2. Recall that the probabilistic distance $\delta(\prec_1, \prec_2)$ is defined as:

$$\delta(\prec_1, \prec_2) = E[\delta(\prec_{f_1}, \prec_{f_2})],$$

where f_1, f_2 are uniform random variables on E_1, E_2, the sets of weak order extensions of $\prec_1, \prec_2$, respectively. Exactly how should we interpret this definition? In the

[10] Here we can also replace $|\mathcal{D}|^2$ with $|\mathcal{D}|(|\mathcal{D}|-1)$ so that the distance may scale to the range of $[0,1]$.

certainty case, this is easy since E_1 and E_2 are finite sets (a finite poset has only finitely many extensions) and we can just take the average of $\{\delta(\prec_{f_1}, \prec_{f_2}) | f_1 \in E_1, f_2 \in E_2\}$. But in the case of uncertainty, the set E_1 and E_2 are typically infinite. For example, consider a typical partial preference elicitation process. We may have determined that the utility function of A_1 is additive over two binary attributes $\{x_1, x_2\}$:

$$u(x) = k_1 u_1(x_1) + k_2 u_2(x_2), k_1, k_2 \geq 0. \tag{4.5}$$

In addition, we have also elicited the sub-utility functions u_1, u_2. We have not, however, assessed the scaling constants (or tradeoff coefficients) k_1, k_2. The set E_1 is thus the set of all utility functions of the form in Equation 4.5, which is obviously infinite.

Partial Utility Functions As Polyhedral Cones

Defining the expectation of a quantity involving random variables over infinite, multi-dimensional domains often requires the language and formalism of measure theory. With a simplifying assumption, however, we can define the probabilistic distance δ using more elementary concepts. Note that since a utility function $u : \Omega \to \Re$ can be viewed as a point in the n-dimensional Euclidean space $\Re^n$: $u = (u(1), u(2), \ldots, u(n))$, we can (and will) talk about the sets E_1, E_2 of consistent utility functions as sets of points in $\Re^n$. The simplifying assumption we shall make regarding E_1, E_2 is that they are determined by linear, homogeneous inequalities. Formally, they are sets of the forms

$$\{\vec{u} \in \Re^n | A\vec{u} \leq \vec{0}\}, \tag{4.6}$$

where A is some $m \times n$ matrix of real numbers, and $\overrightarrow{0}$ is the $m \times 1$ zero vector. In geometric terms, such a set is the intersection of m *half-spaces*, each of which crosses the origin and having one of the rows of matrix A as its *outward normal vector*, and is called a *polyhedral cone*. Partial utility functions satisfying the above assumption encompass most of the common kinds of partial utility functions encountered in the practice of decision analysis. For example, a multi-linear utility function with known sub-utility functions and unknown scaling coefficients, a model studied in Chapter 3, satisfies this assumption. It is not difficult to see that the same is true for multiplicative and additive utility functions with known sub-utility functions and unknown scaling constants. Furthermore, a constraint on the partial preference order $\preceq$ of the form $p \preceq q$, for some $p, q \in \mathcal{S}$ would also translate to a homogeneous linear inequality: $\langle u, p - q \rangle \leq 0$.

The nice thing of having E_1 and E_2 as polyhedral cones is that in the defining formula of the probabilistic distance

$$\begin{aligned}\delta(\prec_1, \prec_2) &= E\left[\delta(\prec_{f_1}, \prec_{f_2})\right] \\ &= \int_{E_1}\int_{E_2}\int_{\mathcal{D}}\int_{\mathcal{D}} c_{\prec_{f_1}, \prec_{f_2}}(p, q)\partial f_1 \partial f_2 \partial p \partial q,\end{aligned}$$

we can interpret the integral on the right hand side as the *volume* of a bounded polyhedral cone in some multi-dimensional Euclidean space. But more importantly, we can reduce the problem of computing the probabilistic distance on partially specified utility functions to the well-studied problem of computing the volume of polyhedral cones. (In fact, the problem of computing the probabilistic distance on partial orders in the certainty case can also be reduced to the volume-computing problem, using some elementary geometric arguments.)

Computing the Volume of Convex Bodies

The problem of computing the volume of convex bodies has received considerable interest in the theoretical computer science community in the past fifteen years. Early results were negative for the prospect of finding an efficient deterministic algorithm (Barany & Furedi 1986). But randomization techniques once again come to the rescue. The first work that uses randomization to obtain a polynomial time algorithm for this problem is due to Dyer et al (Dyer, Frieze, & Kannan 1991). A series of work followed and refined the algorithm of Dyer et al, substantially reducing its complexity (Lovasz, Kannan, & Simonovits 1997). These works are all based on various Markov chain-based sampling techniques that sample points from the convex body according to a nearly uniform distribution. The convex body is input to the algorithm by means of a *membership oracle*, i.e. a black box that provides the answer whether a given point belongs to the convex body. Note that this requirement fits excellently with the assumption that the set E_1, E_2 are polyhedral cones determined by a set of homogeneous linear inequalities as in Equation 4.6: we can check if a utility function $\vec{u}$ is consistent if $A\vec{u} \leq \vec{0}$ in time $O(m)$ (recall that m is the number of rows of A).

We now sketch out the main ideas behind the sampling algorithm. To sample uniformly from a convex body K, we perform a random walk on the points of K. Starting at an arbitrary point inside K, we move at each step to a uniformly selected random point in a ball of radius ϵ about the current point (if this remains inside K, if the new point is outside K, we remain where we were). The size ϵ of the radius is typically $1/\sqrt{n}$. It follows from elementary Markov chain theory that the distribution of the point after t step tends to the uniform distribution as t tends to infinity. The crucial issue is, how long to walk before the walking point becomes nearly uniformly distributed? There are two reasons for needing a long walk: we have to get to the "distant parts" of K, and we may get stuck in "corners", especially "sharp corner" of K. The first reason suggests that we choose a step-size that is large enough relative

Table 4.2. Standard gamble questions. X/Y denotes a 50/50 gamble between X and Y years of survival.

Basic	Times 2	Times 3	Plus 10	Plus 20	Zero
1/10	2/20	3/30	11/20	21/30	0/32
2/10	4/20	6/30	12/20	22/30	0/36
3/10	6/20	9/30	13/20	23/30	
4/10	8/20	12/30	14/20	24/30	
1/12	2/24	3/36	11/22	21/32	
2/12	4/24	6/36	12/22	22/32	
3/12	6/24	9/36	13/22	23/32	
4/12	8/24	12/36	14/22	24/32	

to the diameter of K, while the probability of the second can be reduced by choosing a small step-size. A number of advanced techniques have been developed to address this dilemma to ensure that the Markov chain settles quickly to a nearly uniform distribution (in technical terms, such a chain is called *rapidly mixing*). See Lovász et al (Lovasz, Kannan, & Simonovits 1997) for a comprehensive treatment of this topic.

While this Markov chain-based sampling algorithm was developed for the purpose of computing the volume of convex bodies (and thus can be used to compute the volume of the polyhedron that is $\delta(\prec_1, \prec_2)$), we can use it directly to perform a Monte Carlo estimation of the probabilistic distance on partial utility functions. Specifically, we can estimate $\delta(\prec_1, \prec_2)$ by sampling $f_1^{(i)}, i = 1, 2, \ldots, k$ and $f_2^{(i)}, i = 1, 2, \ldots, k$ according to nearly uniform distributions on E_1 and E_2 respectively, and taking the average $\overline{\delta} = \frac{1}{k} \sum_{i=1}^{k} \delta(f_1^{(i)}, f_2^{(i)})$. Again, the Central Limit Theorem ensures that with a sufficiently big sample size k, the sample mean $\overline{\delta}$ can approximate $\delta(\prec_1, \prec_2)$ with arbitrary degree of precision.

4.3.3. An Illustrative Example

We illustrate the algorithm to compute the probabilistic distance on partially specified utility functions. The data we use are taken from the psychology experiment by Miyamoto and Eraker (Miyamoto & Eraker 1989), as described at the beginning of Section 4.3. Out of the 44 subjects, 6 were dropped due to failure to complete the interview in the allocated time, or failure to understand the CE task. The effective sample size is thus 38. There are a total of 42 CE questions (see Table 4.2). Note that with this data set, it is not possible to define a distance measure that requires the knowledge of the decision alternatives (Equation 4.4).

Computing the Probabilistic Distance on the Subjects for Clustering

Since the survival duration in the CE questions ranges from 0 to 36, we scale the utility functions so that $u(0) = 0$ and $u(36) = 1$. The next step is to discretize the outcome space, which is discretizing the number of years of survival. Because each subject gave 4 different answers (at 4 different time points) to each CE questions, we take the average of the 4 answers as the CE. Because each answer is either integers or integers plus 0.5 (e.g. $(1, .5, 10) \sim 4.5$), we discretize the number of years of survival to the granularity of 1/8, resulting in $36 \times 8 + 1 = 289$ outcomes. We also assume that all subjects prefer longer survival to shorter survival: $u(\frac{i}{8}) \leq u(\frac{i+1}{8}), i = 0, 1, \ldots, 287$. Framed this way, the utility function u of each subject has a total of 288 inequality constraints and $42 + 2 = 44$ equality constraints. It is easy to see that these linear constraints determine a convex set of consistent utility functions.

To find a starting point for the random walk, we need to find a consistent utility function, i.e, a feasible solution for the linear constraints. For this we use the linear programming facility LINPROG of Matlab® Optimization Toolbox, with some randomly generated target function. Interestingly, we found that out of the 38 subjects, only 3 provided consistent answers; the rest provided answers that lead to linear programs that are infeasible. This inconsistency can be attributed to the fact that the expected utility paradigm is normative but not descriptive (Kahneman & Tversky 1979). An example of this school of thought is the approach called *subjective expected utility* (SEU) (Tversky & Kahneman 1967), according to which a CE statement $(m, .5, n) \sim p$ translates into the equation: $(1 - w(.5))u(m) + w(.5)u(n) = u(p)$. Here $0 < w(.5) < 1$ is the *probability distortion* for a .5 probability applying to the superior outcome. Note that in the standard expected utility paradigm, $w(.5) = .5$.

But even with more general utility models such as SEU, it is likely that subjects will have inconsistent preferences, due to *variations in subject responses*. Our approach is to stay within the standard expected utility paradigm and account for the inconsistency in some way. While the fact that random error in judgement exists is well-known, the question of how to deal with it remains open. For the purpose of our experiment, we take the following simple approach. We keep all of the 288 inequality constraints that capture the "longer survival is better" assumption. For each subject, from the set of the 42 equality constraints provided by the CE answers, we incrementally randomly add one at a time to LINPROG and keep doing this as long as a feasible solution exists. Note that due to differences between subjects' responses and the randomness of this method, different sets of CE answers may be taken into account for different subjects. Fortunately, this is not a problem for the probabilistic distance.

Now that a set of consistent CE answers is selected for each subject, we simultaneously start 38 random walks from 38 consistent utility functions, one for each subject. The radius ϵ of the ball is initialized to 0.001. At each iterations, we generate a random point in each ball of radius ϵ. If the generated point is consistent with the constraints, we move to the new point and mark the iteration a *success*; otherwise we stay at the current location and call the iteration a *failure*. If two successes occur consecutively, we double the radius. If two failures occur consecutively, we halve the radius. We stop

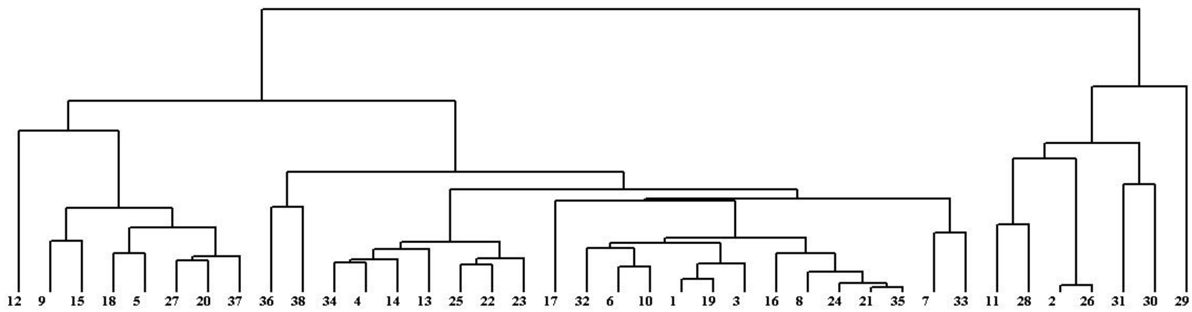

Figure 4.3. Hierarchical cluster of the 38 subjects.

the random walk after 1000 iterations, at which point we obtain a random sample of consistent utility functions for the 38 subjects. We compute the distance between any two consistent utility functions and record the distances in a square dissimilarity matrix of size 38×38. This computation is performed by a routine that implements the algorithm in Table 4.1. We repeat the whole process for a total of 1000 times, updating the averages of the distances as we go. Finally, we input the average distance matrix to the hierarchical clustering algorithm of ClustanGraphics® to obtain the hierarchical clustering shown in Figure 4.3. The method used was average-linkage. [11]

Relation to Attitude Toward Risk. Observe that in this psychology experiment, since there is a single attribute - the duration of survival in number of years - the difference of one subject's preferences from another's is basically the difference in *attitudes toward risk (ATR)*. It is thus interesting to see if there is some correlation between ATR the probabilistic distance we just computed.

How should we define and measure ATR? We take the following straightforward approach. Consider a SGQ $(l, .5, u)$ to which the answer from a subject X is ce, i.e. $(l, .5, u) \sim ce$, and $l < ce < u$. Define the proportional match of this SGQ as:

$$pm = \frac{ce - l}{u - l}.$$

Intuitively, subject X is risk-averse, risk-neutral, and risk-seeking respectively if $pm < .5$, $pm = .5$, and $pm > .5$ respectively. We thus may define the attitude toward risk of X as the sample average of pm, averaging over 42 SGQ's:

$$ATR(X) = \text{avg}\{pm|42 \text{ answers}\}.$$

Now that we have defined and computed ATR for each of the 38 subjects, we analyze the correlation between ATR and the probabilistic distance in the following

[11] All of the codes were written in Java™ and the mathematical programming language of MatLab®. The computations were performed on an Athlon™@850Mhz system with 512MB RAM running Windows® 2000, and took about an 30 minutes to finish.

way. We consider each ordered triple of subjects (X, Y, Z). We look at the probabilistic distance $\delta(X, Y)$ and $\delta(X, Z)$ to see which of the two subjects Y and Z is closer to subject X. We then look at the ATR of X, Y, and Z to see which of the two ATR's of subjects Y and Z is closer to the ATR of X (i.e. which of the two quantities $|ATR(Y) - ATR(X)|$ and $|ATR(Z) - ATR(X)|$ is smaller). If the answers in both instances match, we mark the ordered triple (X, Y, Z) as OK. We then compute the percentage of ordered triple of subjects that is marked OK out of all possible ordered triple. This percentage is in between 72% and 80%, depending on the number of iterations used in computing the probabilistic distance. This shows that there is some strong correlation between the probabilistic distance and the ATR, since the correlation between ATR and a random distance measure can be shown to be approximately 33%.

4.3.4. Related Work on Similarity Measures on Preferences

The only existing similarity measure on preferences in the case of uncertainty that we are aware of is defined in Chajewska et al (1998). This measure is also based on a finite set of decision alternatives. But in contrast to the probabilistic distance that is defined based on the *preference orders*, this measure is defined based on *consistent utility functions*. Specifically, let u_1 and u_2 be two utility functions, and the decision alternatives be $\{p_1, p_2, \ldots, p_m\}$, indexed in such a way that $i = \operatorname{argmax}_{j:1\leq j\leq m}\langle u_i, p_j\rangle, i = 1, 2$. This means that according to the utility function u_i, p_i is an optimal decision alternative, for $i = 1, 2$. The distance between u_1 and u_2 is defined as

$$d(u_1, u_2) = \frac{\langle u_1, p_1\rangle - \langle u_1, p_2\rangle + \langle u_2, p_2\rangle - \langle u_2, p_1\rangle}{2}.$$

The difference of the first two terms in the above numerator, $\langle u_1, p_1\rangle - \langle u_1, p_2\rangle$, is called the *utility loss* of u_1 with respect to u_2, and that of the last two terms is the *utility loss* of u_2 with respect to u_1. The utility loss of one utility function with respect to another is the difference of, or loss in expected utility by choosing a decision alternative that is optimal according to the latter instead of the former.

There are several issues with this definition of distance between utility functions. First, this is a similarity measure between utility functions. If it is to be used a measure of similarity between preference orders, one must deal with the issue of choosing the corresponding consistent utility functions. A standard solution is to scale the consistent functions to the range of $[0, 1]$ (a sort of the equivalence of canonical value functions for strict orders in the definition of Spearman's rho in the case of certainty). The second issue is that since this measure focuses on optimal decision consequences, it can become vacuous if there is a clear optimal decision alternative, a "clear winner" among the competing candidates. Formally, suppose that p_1 is the optimal decision alternative according to both u_1 and u_2, then $d(u_1, u_2) = 0$. The implication here is that the two preference orders are maximally similar, while the truth beneath is that they only agree on choosing the "clear winner". On the other hand, depending on the intended use, this measure can still be useful - imagine that it is used only under circumstances when there is no "clear winner". The third, mainly technical issue is that, strictly speaking, this distance is not well-defined, since $\text{argmax}_{j:1\leq j\leq m}\langle u_i, p_j\rangle, i = 1, 2$ are not well-defined. Suppose for example that in addition to p_1, p_3 is also an optimal decision alternative according to the utility function u_1. Replacing p_1 with p_3 in the defining formula for this distance may result in a different value, since $\langle u_2, p_3\rangle$ may be different from $\langle u_2, p_1\rangle$. But again, depending on the intended use of the distance measure, this variance may not play an important role. Finally, as Chajewska et al noted, this distance measure is not a metric since it does not satisfy the triangle inequality.

What other kinds of similarity measure can be defined on preferences in the uncertainty case? An immediate thought that comes to one's mind is to extend measures such as Pearson's coefficient, Spearman's footrule, etc. to accommodate utility functions. This generalization, however, has another difficulty beside the issue of scaling the utility functions. To see why, let us consider how we might define Spearman's footrule on utility functions:

$$F(u_1, u_2) = \sum_{i=1}^{n} |u_1(i) - u_2(i)|. \tag{4.7}$$

As Chajewska et al noted, this approach gives all outcomes equal weight: a difference in utility for a highly probable outcome contributes the same to F as the same difference in utility for a highly improbable outcome. In other words, generalizing similarity measures to utility functions this way cannot account for any *a priori* knowledge about the resulting decision consequences.

It is interesting to see what happens to the Pearson correlation coefficient on utility functions. Since strategically equivalent utility functions are positive linear transformation of each other, they are maximally similar (or perfectly correlated) according to the Pearson's measure.

4.3.5. Related Work in Case-Based Utility Assessment

In a recent paper, Chajewska *et al* (Chajewska *et al.* 1998) discuss an approach to preference elicitation similar to ours. Given a data base of user utility functions, they propose clustering them and describing each cluster by a prototype. They propose building a decision tree for associating a user with a prototype utility function based on some elicited pairwise preferences. Their approach requires having a data base of complete utility functions. Their retrieval scheme depends on asking the user questions and ruling out utility functions that conflict with the user's answers. Since no prototype is likely to exactly match the user's preferences, this approach has the problem that the utility function retrieved is sensitive to the order in which questions are asked. In contrast, our approach would retrieve the closest matching preference structure, independent of the order of questions. Since Chajewska *et al* are initially focusing on the problem of building a working system and we are initially focusing on the theoretical underpinnings, we see their work as complementary to ours.

Chajeswska et al. (Chajewska & Koller 2000) pursue another approach to utility elicitation using the same set of utility functions. The novelty of this approach is that utilities are treated as random variables, and if drawn from a mixture of Gaussians, as they were postulated to, their density functions can be learned from the utility database using Bayesian learning techniques. Also, using standard Bayesian techniques, it is possible to determine the relevance of an elicitation question based on its *value of information* (Chajewska, Koller, & Parr 2000). In contrast, our case-based approach requires fewer structural assumptions and as such has an edge over Chajewska et al.'s approach in those situations where these assumptions are not applicable.

4.4. SUMMARY

We investigate a case-based approach to preference elicitation and decision making. The focus of our investigation is the probabilistic distance, a measure of similarity among people's preferences that has its roots in the Kendall's tau function. We showed that this measure has attractive theoretical properties, can be approximated efficiently in all situations, and has a number of advantages over existing similarity measures on preferences.

- **Attractive Theoretical Properties.** The probabilistic distance is a metric on complete preference orders, regardless of whether the decision problem involves uncertainty or not (Theorem 4).
- **Amenability to Efficient Computation.** In the case of certainty, the probabilistic distance on partial preferences can be approximated using a randomized algorithm that samples uniformly randomly from the set of linear extensions of a partial order (Section 4.1.2). Under uncertainty, the problem is innately harder, because of the complexity introduced by probabilities and utilities. We have shown that with the reasonable assumption that the set of consistent utility functions is linearly bounded, computing the probabilistic distance can be reduced to the well-studied problem of computing the volumes of convex bodies for which efficient approximate algorithms exist (Section 4.3.2).
- **Favorable Comparison with Existing Similarity Measures on Preferences.** In the case of certainty, the probabilistic distance theoretically appears to be better suited for use in recommendation systems than the predominant Pearson correlation coefficient measure. This is confirmed by the experiments

with DIVA the Decision-Theoretic Video Advisor (Section 4.2.3). In the case of uncertainty, the probabilistic distance is the first similarity measure that is defined on partial preference orders. Because of its reliance on orders instead of utilities, the probabilistic distance can be defined and computed in a wide range of situations (Section 4.3).

In addition, we believe that the implication of the probabilistic distance goes beyond the context of case-based preference elicitation, since it is in its most general form a *distance measure on partial orders* - a topic that has not been received adequate treatment. A manifestation of this statement is the use of the probabilistic distance in the robustness analysis of the KBANN network for user preference modeling (Section 4.2.3).

We are currently investigating several medical decision problems as potential candidates for implementing the case-based preference elicitation approach. For such candidates, the basic requirement is that a database of patient utilities is available. Since utility data are routinely collected for a wide range of medical decision problems, and since the standard gamble CE method is one of the most widely used techniques to elicit utilities, we believe that the case-based approach using the probabilistic distance has serious potential to see real-world application.

There are two main research issues that future work can address. The first issue if *clustering*. Our analysis suggests that computing the distance between preference structures can be fairly complex. For this reason, in the case retrieval step, it may be desirable to reduce the number of preference structures to be matched with the preference structure of the new user. One possibility is to perform clustering of the case base using the probabilistic distance measure and to store prototype representations of each cluster. Then instead of matching each preference structure in the case base, the system needs to match only the prototypical preference structures with A's, thus achieving some computational savings. There are various schemes to clustering the case base and it is an question which one is most suitable for our case-based approach. The most popular clustering methods are: 1) *Hierarchical agglomerative*, 2) *Iterative partitioning*, 3) *Factor analytic*, 4) *Hierarchical divisive*, 5) *Density search*, 6) *Clumping*, and 7) *Graph theoretic*. Hierarchical agglomerative methods are frequently used in biological sciences, whereas factor analytic methods are popular in psychology. In addition to selecting a clustering method to cluster the case base, we need to address other issues such as determining the number of clusters and representing the prototype of each cluster. Addressing these technical issues will facilitate the process of fine-tuning the system.

The second issue is incorporation of user feedback. Using the retrieved case, the system may be able to give recommendations to A. These recommendations of course may or may not be what A finds useful. Thus, it is desirable if the system can incorporate user feedback and update the working model of user preferences accordingly, in order to give better recommendations. One possible approach for updating the working model is to regard user feedback as a correction to the working preference model and should override preferences in that model. We plan to explore protocols that facilitate this interactive process of learning user preferences.

CHAPTER 5
THESIS SUMMARY

The objective of this thesis is to develop a flexible decision-theoretic framework for eliciting and reasoning with preferences and to apply the tools and techniques of this framework to build a practical decision support system. To this end, I propose two orthogonal, complementary approaches to eliciting and reasoning with partial preference information.

The first approach is grounded in the well-understood theory of multi-attribute utility functions. It assumes that the user utility function decomposes according to the multi-linear form with unknown scaling coefficients. The system then use comparative statements by the user, represented by *qualitative* logical constructs, to constrain these coefficients. These constraints can in turn induce that certain decision alternatives are sub-optimal. The representation of logical constructs is facilitated by the languages of qualitative preference logics developed in the field of qualitative decision theory.

The second approach adapts ideas from the field of case-based reasoning and collaborative filtering to the problem of eliciting user preferences. I envision a that system maintains a population of users with their preferences partially or completely specified in a given domain. When encountering a new user, called A, the system first elicits some preference information from A, and then determines which user in the population has the preference structure that is closest to A's. The preference structure of that user will be used to determine an initial default representation (or working model) of A's preferences. I define and investigate the probabilistic distance, a measure of similarity among people preferences that has its roots in the Kendall's tau function. I show that this measure has attractive theoretical properties, can be computed efficiently in all situations, and has a number of advantages over existing similarity measures on preferences. I believe that the implication of the probabilistic distance goes beyond the context of case-based preference elicitation, since it is in its most general form a *distance measure on partial orders* - a topic that has not been received adequate treatment

It is interesting to see how the techniques of the above two approaches can be integrated to build a single decision support system. One possible way is to use the techniques of the first approach to elicit some initial preference information from the decision maker and to eliminate some sub-optimal decision alternatives. If the set of remaining candidates is still large, the system can use the case-based approach to make recommendations for the decision maker.

BIBLIOGRAPHY

Bacchus, F., and Grove, A. 1996. Utility independence in a qualitative decision theory. In *Proceedings of the Fifth International Conference on Knowledge Representation and Reasoning (KR '96).* Morgan Kaufmann.

Bacchus, F., and Grove, A. 1997. Independence and qualitative decision theory. In *Working Notes of the Stanford Spring Symposium on Qualitative Decision Theory.*

Barany, I., and Furedi, Z. 1986. Computing the volume is difficult. In *Proceedings of the 18th Annual Symposium on Theory of Computing*, 442–447.

Boutilier, C. 1994. Toward a logic for qualitative decision theory. In *Principles of Knowledge Representation and Reasoning: Proceedings of the Fourth International Conference (KR94).* San Mateo, CA: Morgan Kaufmann.

Breese, J.; Heckerman, D.; and Kadie, C. 1998. Empirical analysis of predictive algorithms for collborative filtering. In *Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence*, 43–52.

Brightwell, G., and Winkler, P. 1991. Counting linear extensions. *Order* 8(3):225–242.

Bubley, R., and Dyer, M. 1998. Faster random generation of linear extensions. In *Proceedings of the Ninth Annual ACM–SIAM Symposium on Discrete Algorithms*, 175–186.

Chajewska, U., and Koller, D. 2000. Utilities as random variables: Density estimation and structure discovery. In *Proceedings of the Sixteenth Conference on Uncertainty in Artificial Intelligence*, 63–71.

Chajewska, U.; Getoor, L.; Norman, J.; and Shahar, Y. 1998. Utility elicitation as a classification problem. In *Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence.* To appear.

Chajewska, U.; Koller, D.; and Parr, R. 2000. Making rational decisions using adaptive utility elicitation. In *Proc AAAI 2000*, 363–369.

Clifford, H., and Stephenson, W. 1975. *An Introduction to Numerical Taxonomy.* New York: Academic Press.

Critchlow, D. E. 1980. *Metric Methods for Analyzing Partially Ranked Data.* Springer-Verlag. Lecture Notes in Statistics.

David, H. 1981. *Order Statistics, 2nd Edition.* John Wiley.

de Finetti, B. 1974. *Theory of Probability.* New York: Wiley and Sons.

Doyle, J., and Wellman, M. 1994. Representing preferences as ceteris paribus comparatives. In Hanks, S.; Russel, S.; and Wellman, M., eds., *Working Notes of the AAAI Spring Symposium on Decision-Theoretic Planning.* 'Stanford, CA: American Association of Artificial Intelligence.

Doyle, J.; Shoham, Y.; and Wellman, M. 1991. A logic of relative desires (preliminary report). In *Proceedings of the 6th International Symposium on Methodologies for Intelligent Systems*, 16–31.

Dyer, M.; Frieze, A.; and Kannan, R. 1991. A random polynomial-time algorithm for approximating the volume of convex bodies. *Journal of the ACM* 38(1):1–17.

Efron, B. 1986. Why isn't everyone a bayesian. *American Statistical Association*.

Elekes, G. 1986. A geometric inequality and the complexity of computing volume. *Discr. Comput. Geom.* 1:289–292.

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

Frank, R.; Massy, W.; and Wind, Y. 1972. *Market Segmentation*. New Jersey: Prentice-Hall.

Ha, V., and Haddawy, P. 1997. Problem-focused incremental elicitation of multi-attribute utility models. In *Proceedings of the Thirteenth Conference on Uncertainty in Artificial Intelligence*, 215–222.

Haddawy, P.; Ha, V.; Restificar, A.; Geisler, B.; and Miyamoto, J. 2001. Preference elicitation via theory refinement. Manuscript submitted to Machine Learning Journal, Special Issue on Fusion of Domain Knowledge with Data for Decision Support.

Hazen, G. 1986. Partial information, dominance, and potential optimality in multi-attribute utility theory. *Operations research* 34(2):296–310.

Herlocker, J.; Konstan, J.; Borchers, A.; and Riedl, J. 1999. An algorithmic framework for performing collaborative filtering. In *Proceedings of the 1999 Conference on Research and Development in Information Retrieval*.

Hill, W.; Stead, L.; Rosenstein, M.; and Furnas, G. 1995. Recommending and evaluting choices in a virtual community of use. In *Proceedings of the ACM CHI'95 Conference on Human Factors in Computing Systems*, 194–201.

Holtzman, S. 1989. *Intelligent Decision Systems*. Reading,Mass: Addison-Wesley.

Howard, R., and Matheson, J. 1984. *The Principles of Decision Analysis*. Menlo Park, CA: Strategic Decision Group.

Jardine, N., and Sibson, R. 1971. The construction of hierarchic and non-hierarchic classifications. *Computer Journal* 11:117–184. New York: John Wiley.

Jerrum, M., and Sinclair, A. 1996. The markov chain monte carlo method: an approach to approximate counting and integration. In Hochbaum, D., ed., *Approximations for NP-hard Problems*. Boston, MA: PWS Publishing. 482–520.

Kahneman, D., and Tversky, A. 1979. Prospect theory: An analysis of of decisions under risk. *Econometrica* 47:276–287.

Karzanov, A., and Kachiyan, L. 1991. On the conductance of order markov chains. *Order* 8(1):7–15.

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

Kemeny, J. 1959. Mathematics without numbers. *Daedalus* 88:571–591.

Kendall, M. G. 1962. *Rank Correlation Method*. New York: Hafner Publishing Company. Third Edition.

Koksalan, M., and Sagala, P. 1995. Interactive aproaches for discrete alternative multiple criteria decision making with monotone utility functions. *Management Science* 41(7):1158–1171.

Konstan, J.; Miller, B.; Maltz, D.; Herlocker, J.; Gordon, L.; and Reidl, J. 1997. Applying collaborative filtering to usenet news. *Communications of the ACM* 40(4):77–87.

Korhonen, P.; Wallenius, J.; and Zionts, S. 1984. Solving the discrete multiple criteria problem using convex cones. *Management Science* 30(11):1336–1345.

Kreps, D. 1988. *Notes on the Theory of Choice*. Underground Classics in Economics. Boulder: Westview Press.

Lovasz, L.; Kannan, R.; and Simonovits, M. 1997. Random walks and an $o*(n^5)$ volume algorithm for convex bodies. *Random Structures and Algorithms* 11:1–50.

Matheiss, T. H., and Rubin, D. S. 1980. A survey and comparisons of methods for finding all vertices of convex polyhedral sets. *Mathematics of Operations Research* 5:167–185.

McCarthy, J., and Hayes, P. 1969. Some philosophical problems from the standpoint of artificial intelligence. In Mitchie, D., ed., *Machine Intelligence 4*. Edinburgh Univesity Press. 463–502.

Miyamoto, J. M., and Eraker, S. A. 1989. Parametric models of the utility of survival: Tests of axioms in a generic utility framework. *Organizational Behavior and Human Decision Processes* 44:166–202.

Nguyen, H., and Haddawy, P. 1999. The decision-theoretic interactive video advisor. In *The Fifteenth Conference on Uncertainty in Artificial Intelligence*. To appear.

North, D. W. 1990. A tutorial introduction to decision theory. In Shafer, G., and Pearl, J., eds., *Readings in Uncertain Reasoning*. San Mateo, CA: Morgan Kaufmann. 68–78.

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

Prasad, S.; Karwan, M.; and Zionts, S. 1997. Use of convex cones in interactive multiple objective decision making. *Management Science* 43(5):723–734.

Pyke, R. 1965. Spacings. *Journal of the Royal Statistical Society B* 27:395–436. Discussion: 437-49.

Raiffa, H. 1968. *Decision Analysis.* Reading,MA: Addison-Wesley.

Resnick, P.; Iacovou, N.; Suchak, M.; Bergstrom, P.; and Reidl, J. 1994. Grouplens: An open architecture for collaborative filtering of netnews. In *Proceedings of the 1994 Computer Supported Cooperative Work Conference*, 175–186.

Salton, G., and McGill, M. 1983. *Introduction to Modern Information Retrieval.* McGraw-Hill.

Savage, L. 1954. *The Foundations of Statistics.* New York: John Wiley & Sons. (Second revised edition published 1972).

Seidel, R. 1997. Convex hull computations. In Goodman, J. E., and O'Rourke, J., eds., *Handbook of Discrete and Computational Geometry.* CRC Press LLC. chapter 19, 361–376.

Shachter, R. 1986. Evaluting influence diagrams. *Operations Research* 34(6):871–882.

Shafer, G., and Pearl, J. 1990. *Readings in Uncertain Reasoning.* San Mateo, CA: Morgan Kaufmann.

Shafer, G. 1986. Savage revisited. *Statistical Science* 1(4).

Shardandand, U., and Maes, P. 1995. Social information filtering: Algorithms for automating "word of mouth". In *Proceedings CHI-95.*

Sinclair, A. 1993. *Algorithms for Random Generation and Counting: A Markov Chain Approach.* Boston-Basel-Berlin: Birkhauser.

Spearman, C. 1904. The proof and measurement for association between two things. *American Journal of Psychology* 15:72–101.

Spearman, C. 1906. Footrule for measuring correlation. *British Journal of Psychology* 2:89–108.

Tan, S., and Pearl, J. 1994. Qualitative decision theory. In *Proceedings of the Twelfth National Conference on Artificial Intelligence*, 928–932.

Toda, S. 1989. On the computational power of pp and +p. In *Proceedings of the 30th IEEE Symposium on Foundations of Computer Science*, 514–519.

Tversky, A., and Kahneman, D. 1967. Utility theory and additivity analysis of risky choices. *Journal of Experimental Psychology* 75:27–36.

Tversky, A., and Kahneman, D. 1986. Rational choice and the framing of decisions. *J. Business* 59:S251–S278.

Ulam, S. M. 1981. Future applications of mathematics in the natural sciences. *American Mathematical Heritages: Algebra and Applied Mathematics. Texas Tech. University, Mathematics Series.* 13:101–114.

Valiant, L. 1979. The complexity of computing the permanent. *Theoretical Computer Science* 8:410–421.

von Neumann, J., and Morgenstern, O. 1944. *Theory of Games and Economic Behavior*. Princeton Univesity Press.

von Winterfeldt, D., and Edwards, W. 1986. *Decision Analysis and Behavioral Research*. Cambridge: Cambridge University Press.

Wald, A. 1950. *Statistical Decision Functions*. New York: Wiley.

Zionts, S., and Wallenius, J. 1976. An interactive programming method for solving the multiple criteria proble. *Management Science* 22(6):652–633.

APPENDIX A
RANDOM GENERATION OF LINEAR EXTENSIONS OF A PARTIAL ORDER

We describe below an algorithm, due to Bubley and Dyer (Bubley & Dyer 1998), that almost uniformly randomly generates linear extensions of a partial order. The algorithm has running time of $O(n^3 \log n\epsilon^{-1})$, where n is the number of the elements of the partial order, and ϵ is the desired accuracy, which means that the generated random linear extension has a probability distribution that is within a total variation distance [1] of ϵ from the uniform distribution. The running time required to obtain a certain precision ϵ is often called the *mixing time* of the Markov chain. A Markov chain with a mixing time polynomial with respect to the input size (which is the number of elements of the partial order in this case) and ϵ^{-1} is called *rapidly mixing*.

Suppose that the partial order $\prec$ has n elements, and $N = \{1, 2, \ldots, n\}$. We encode the orderings of these elements with the permutations of the elements of N, and the set of linear extensions of $\prec$ by a subset $\mathcal{LE}(\prec)$ of the set of all permutations of the elements of N.

For a given concave probability distribution f on $\{1, 2, \ldots, n-1\}$, define a Markov chain $\mathcal{M}_f = \{S_t\}_{t \geq 0}$ on $\mathcal{LE}(\prec)$ as follows. At any time point $t \geq 0$, toss a fair coin. If the coin lands head, then let $S_{t+1} = S_t$. If the coin lands tail, then choose an index $i \in \{1, 2, \ldots, n-1\}$ according to the distribution f. If the permutation obtained from S_t by switching the i-th and $(i+1)$-st elements of S_t is also a linear extension of $\prec$, i.e., an element of $\mathcal{LE}(\prec)$, then let S_{t+1} be this new permutation. Otherwise, let $S_{t+1} = S_t$.

It is easily seen that $\mathcal{M}_f$ is ergodic with uniform stationary distribution. When f is the uniform distribution on $\{1, 2, \ldots, n-1\}$, $\mathcal{M}_f$ is the Karzanov-Kachiyan chain with mixing time $O(n^5 \log n + n^4 log\epsilon^{-1})$ (Karzanov & Kachiyan 1991). Bubley and Dyer showed that if f is defined as $f(i) = i(n-i)/K$, where $K = (n^3 - n)/6$, then $\mathcal{M}_f$ has mixing time of $O(n^3 \log n\epsilon^{-1})$.

[1]The total variation distance between two discrete distributions P, Q over a finite sample space S resembles the Spearman's footrule, and is defined as $\mathrm{d}_{TV}(P,Q) = \frac{1}{}$

$\sum \quad |P(s) - Q(s)|$

CURRICULUM VITAE OF VU ANH HA

EDUCATION

- Ph.D. in Computer Science, August 2001. University of Wisconsin-Milwaukee, Milwaukee, WI. Minor: Probability and Statistics
- M.S. in Computer Science, July 1995. Eotvos University, Budapest, Hungary.
- B.S. in Computer Science, July 1991. Kossuth University, Debrecen, Hungary.

HONORS & AWARDS

- University of Wisconsin-Milwaukee Dissertator Fellowship (1999-2000).
- University of Wisconsin-Milwaukee Graduate Fellowship (1998-1999).
- Hungarian Government Scholarship for graduate study at Eotvos University, 1991-1995 and for undergraduate study at Kossuth University, 1988-1991.
- Gold Medal at the 27th International Mathematics Olympiad in Warsaw, Poland, 1986.