DECISION ANALYSIS WITH MULTIPLE CONFLICTING OBJECTIVES PREFERENCES AND VALUE TRADEOFFS

(Preface, Chapters 1 & 2)

Ralph L. Keeney and Howard Raiffa

May 1975 WP-75-53

COMMENT

This working paper is the manuscript for a book titled Decision Analysis with Multiple Conflicting Objectives: Preferences and Value Tradeoffs being published by John Wiley and Sons, New York. It is being distributed now in very limited number prior to formal publication both (1) to facilitate the use of these results within the IIASA projects, and (2) to elicit comments on their content.

The work reported here began over five years ago when Ralph L. Keeney was affiliated with the Massachusetts Institute of Technology and Howard Raiffa was at Harvard University. The finalization of this work has taken place at IIASA where our interactions with various members of the applied projects has helped to make the presentation more useful to potential practitioners.

Efforts are now beginning to utilize the theories and procedures outlined in this book on the problems being addressed by the applied projects of IIASA. We plan to report on these developments in the various IIASA publications in the near future.

PREFACE

If we wanted our title solely to convey the subject matter of our book, it would be some horrendously complicated concoction like: "On Cardinal Utility Analysis with Multiple Conflicting Objectives: The Case of Individual Decision Making Under Uncertainty from the Prescriptive Point of View--with Special Emphasis on Applications but with a Little Theory Thrown-In for Spice."

Our present title, Decision Analysis with Multiple Conflicting Objectives: Preferences and Value Tradeoffs is longer than we think a title should ideally be, but it unfortunately is too short to prevent unjustified sales. Even in such a simple case, it is not so easy to balance among the conflicting objectives: convey the subject matter, minimize the length, and promote justified sales but prevent unjustified ones.

To an ever-growing circle of people "Decision Analysis" has carved out for itself a niche in the literature of operations research, systems analysis, management sciences, decision and control, cybernetics, and so on. Decision analysis looks at the paradigm in which an individual decision maker (or decision unit) contemplates a choice of action in an uncertain environment. The approach employs systematic analysis, with some number pushing, which is designed to help the decision maker clarify in his own mind which course of action he should choose. In this sense, the 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 super-rational being with an all powering intellect. It is rather a prescriptive approach designed for normally intelligent people who want to think hard and systematically about some important real problems.

The theory of Decision Analysis is designed to help the individual make a choice amongst a set of prespecified alternatives. Of course, decision analysts do admit that an insightful generation of alternatives is of paramount importance and they also take note of the often overlooked fact that good analysis of a set of existing alternatives may be suggestive of ways to augment the set of alternatives. But this is a sidepoint that is not suitable for development in a preface. What is of importance here is that the usual analysis (after suitable modelling has been done) involves two distinctive features: an uncertainty analysis and a preference (or value or utility) analysis. There has been a great deal that has been written on the uncertainty phase: on statistical validation of a model, on uses of historical and experimental data for inference, on the codification of judgmental estimates by the decision maker and by expert groups, etc. In comparison with this voluminous literature on the uncertainty side rather little has been written about the value or preference side of the picture. The ensuing 000 pages are designed to help improve the balance.

At present, this gross imbalance is also unfortunately very much in evidence in applications. Several person-years of effort will be utilized developing, modifying, and verifying an elaborate simulation model which outputs the possible levels of several indicators of interest resulting from any particular policy. Perhaps the output is synthesized in terms of a few graphs or tables and a summary report is written for the decision maker. This decision maker then struggles for perhaps a week with the implications of the alternatives and then chooses an alternative. The score: person-years on the modelling and uncertainty side of the problem, a week on the preference side. We feel the shifting of a little effort--perhaps only a few person-months--to the preference aspects could lead to significantly improved decision making in many situations. In this book, we suggest how one might constructively use more effort on the preference aspects of analysis.

An illustrative example can help set the stage. A decision making unit must make a policy choice in a complicated environment. Imagine that the problem is so complicated, that a computer-based simulation model is designed such that for each policy choice under review, a scenario can be generated which indicates how the future might unfold in time. Now suppose that the analyst effectively summarizes the relative desirability of any future scenario not by a single number but, let us say, by a dozen well-chosen numbers: some reflecting costs, others reflecting benefits. Since these output performance numbers may simultaneously deal with economic, environmental, social, and health concerns, these summarizing indices will, in general, be in incommensurable units. To complicate matters, let us suppose that stochastic elements are involved in the simulation so that for a single policy choice being investigated repeated simulation runs result in different sets of summary performance measures. The joint probability distribution of these performance measures as made manifest through repeated realizations of the simulation will, in general, indicate that these 12 measures are probabilistically dependent. Now assume you are the poor decision maker sitting in front of an output display device deluged with a mountain of conflicting information. You are confused. What should you do? How can you sort out the issues and start thinking systematically about your choice problem: which policy should you adopt in the real setting? Well, you might want to pause for a time and read this book. We believe we are addressing your problem and have something constructive to say about it that is not merely platitudinous.

Of one thing we are convinced: the decision maker cannot simply plug these incommensurate output performance measures into an objective formula that someone has proposed ex ante without any reference to the real-world meaning of the various measures. Rather, our prescriptions lead us in an opposite direction: we advocate that the responsible decision maker force himself to think hard about various value tradeoffs and about his attitudes towards risky choices and we suggest ways that this process can be systematically examined by dividing his complicated choice problem into a host of simpler choice problems.

The methodology will in a step-by-step fashion force the cooperating decision maker to articulate a rank ordering of all potential outcome vectors--in the illustrative example, an ordering of all 12-tuples. This rank ordering can be thought of as constituting a set of indifference curves plus an orientation in 12-space. But this is not enough since repeated simulations of the same policy will produce, because of stochastic elements, different 12-tuples. Our problem is a familiar one by now, and the utility theory of von Neumann-Morgenstern comes to the rescue. This theory tells us that in order to satisfy certain compelling behavioral desiderata, the decision maker must assign to each 12-tuple a single number, referred to as the utility of that 12-tuple, and this assignment must be such that:

a) the more preferred the 12-tuple the higher the associated utility and

b) these utilities must be scaled in a way that justifies the maximization of expected utilities.

This means that in order to evaluate the relative desirability of a given policy alternative one must (i) generate for each simulation run a set of output 12-tuples, (ii) associate to each 12-tuple a utility, and (iii) average the sequence of utilities generated by repeated runs for the same policy. Finally, one should choose the policy which maximizes the expected utility. Built into the assignment of utilities are all the aspects of risk aversion or proneness that one should be entitled to include. That this can be done and how it is done is the subject matter of Utility Theory which we review in Chapter 4.

Having stated our general approach, can it actually be done in practice? We argue Yes and we substantiate our case by citing many examples illustrating how it has already been done in practice. It's not easy to do; but what are the alternatives?

Outline of the Book

For conceptual purposes, the material presented can be partitioned into four main categories: (1) the structuring of multiple-objective problems: chapters 1 and 2; (2) the theory of quantifying preferences over multiple objectives: chapters 3 through 6; (3) the applications of that theory: chapters 7 and 8; and (4) special topics: chapters 9 and 10. Let us only briefly elaborate here since a more detailed outline is found in section 1.6.

Chapter 1 introduces the subject matter of concern more systematically than has been done above. Our basic problem is phrased in terms of the analysis of decision trees rather than in terms of a stochastic simulation model, but the distinction for our purposes does not matter. In chapter 2, we acknowledge that in a given context the set of objectives and attributes are not given for a problem. Some suggestions are made for generating and structuring appropriate sets of objectives.

The theory, chapters 3 through 6, presents techniques for quantifying preferences over multiple objectives. In order to obtain a von Neumann-Morgenstern utility function in such cases, one must address two separate issues: value tradeoffs among objectives and attitudes toward risk. Chapter 3 looks at value tradeoffs under conditions of certainty. Chapter 4 restricts itself to a single objective and introduces concepts and techniques that are needed in quantifying and assessing risk attitudes. This chapter essentially reviews single-attribute (i.e. unidimensional) utility theory. Chapters 5 and 6 consider both of these issues simultaneously; they present multiattribute (i.e. multidimensional) utility theory. Due to its length we have arbitrarily divided this material into two segments: two attributes (chapter 5) and more than two attributes (chapter 6).

Multiattribute utility is already sufficiently developed to make worthwhile contributions to some important complex problems. Chapters 7 and 8 dealing with applications present support for this claim; many problems are discussed where preferences have been quantified using multiattribute utility. These include: structuring corporate objectives, examining operational policies of fire departments, allocating school-system funds, evaluating time-sharing systems, siting nuclear power facilities, treating such medical problems as cleft lip and palate, and so forth. In each case, we describe the problem context in which the preference assessments took place. We want to communicate some of the art as well as the theory and procedures of using multiattribute utility analysis.

Chapter 8 uses the theory and procedures developed in earlier chapters in a major case study: the development of airport facilities for Mexico City.

Chapters 9 and 10 on special topics examine respectively preferences over time and aggregation of individual preferences. Each of these important problems can be cast and naturally studied in a multiattribute framework. As shown, many of the results of chapters 3 through 6 are relevant to the time and group problems. These two problems are often added complicating features in multiple-objective problems.

Our Intended Audience

Decision making is of such a pervasive interest that it is hard for us to exclude any group. Certainly this book should be of relevance to all sorts of analysts, policy makers, policy advisors, economists, designers, engineers, and managers. Meaningful and important applications can be found in business, in public policy, in engineering design, in resource management, in public health and medicine, in educational management, and on and on.

It's a big book and not all of it has to be read. There are parts, especially chapter 6 and the latter part of 9, where the mathematics will be discouragingly complicated except for the mathematical pros. It would be helpful if the non-mathematical reader were already familiar with the rudiments of decision analysis as explicated by Raiffa [1968] or by books at a similar level such as Schlaifer [1969] and Brown et al. [1974].

Depending on interests, the reader may wish to read only a selection of the chapters. Chapters 1 and 2 on structuring the multiple-objective problem can be read with no prerequisite. Similarly, if one is willing to accept the abstract formulation of the problem, the theory chapters 3 through 6 are essentially self-contained. Even within this group, the reader with some mathematical background could begin with either value tradeoffs (chapter 3), unidimensional utility theory (chapter 4), or multiattribute utility theory (chapters 5 and 6). For a full understanding of the applications in chapters 7 and 8, a knowledge of the main theoretical results of the book is required. However, a reader interested in the domain of applicability of multiattribute utility and a feeling for how one uses it in a specific context could pick them up reading only chapters 7 and 8. Before reading chapters 9 or 10, it would be advisable in most cases to at least glance through chapters 3 through 6. However, a reader who feels at ease with the level of mathematics (not that it is so high) in these chapters could begin with either 9 or 10 and only refer back to the basic theory chapters when back references indicate it may be worthwhile.

To our knowledge, there are no other books which overlap much in content with this one. However, most of the theoretical results have appeared in professional journals. Many of these are due to researchers other than ourselves. We have attempted to appropriately reference the original contributions so that a reader can easily trace the development of any particular topic. A large bibliography of these works is included following chapter 10.

Acknowledgements

In June 1969, we began this joint effort which managed to occupy approximately six years. With two authors, that is around 4000 man-days of effort to produce approximately 200,000 words. Somehow it seemed to each of us that we were producing more than 50 words a day. Comparing our June 1969 outline of the proposed monograph with the current table of contents clearly indicates that much of the included material was not available in 1969. This is particularly true of the applications--essentially all represent efforts in the 1970's.

During the course of writing this book, a number of people have helped us in many ways. Several individuals read various draft versions of chapters and suggested many useful comments. We would particularly like to thank Craig Kirkwood, Tjalling Koopmans, John Schmitz, and Mike Spence. After writing a preliminary version on preferences over time, chapter 9 was edited and the advanced sections were added by our friend Richard F. Meyer, who is currently doing research on the forefront of theory and applications in this area. The final product has clearly benefited from his efforts. Much to our delight, David E. Bell agreed to read the entire last draft of the manuscript for technical content. His many worthwhile suggestions relegated that version to the next-to-last draft. However, because he is a good friend, to make us feel better

about the additional revisions, he agreed to absolve us of any responsibility for technical errors in the text.

From September 1971 - June 1974 support for Ralph L. Keeney's contribution to this book came from the Office of Naval Research Contract N00014-67-A-0204-0056 with the M.I.T. Operations Research Center. This final year of work by both of us has been supported by the International Institute for Applied Systems Analysis in Laxenburg, Austria. The many interactions with our colleagues at IIASA have helped to make the presentation more useful to potential practitioners.

Finally, we intend over the next few years to compile a bibliography of applications of multiattribute value and utility analyses along the lines suggested in this book. Since many applications will likely appear in technical reports, etc. rather than the open literature, we would appreciate it if authors of such material would send each of us a copy. If there ever is an opportunity to have an updated edition of this book, we would plan on adding material on these 'more recent' applications.

Ralph L. Keeney

Howard Raiffa

Laxenburg, Austria
May 1975

CONTENT

CHAPTER 1

THE PROBLEM

In an uncertain world the responsible decision maker must balance judgments about uncertainties with his or her preferences for possible consequences or outcomes. It's not an easy task to do and even though we all have a lot of practice, we are not very very good at it. In this book we suggest formal techniques that we think can be of assistance in this decision process. We will concentrate on formalizing the preference or value side of the problem rather than developing procedures for the assessments of uncertainties. This is not to be interpreted that we do not think modeling of the uncertainties is a critically important task. However, we feel that many capable scholars have and continue to address the modeling aspects of the class of problems we have in mind: Our efforts on the *value* side of the problem are meant to complement these. So, let us assume that the assessments of uncertainties are given, and let's worry about how we, as decision makers, can make sense out of our conflicting values, objectives, or goals, and arrive at a wise decision. As one of our associates likes to put it, "the aim of the analysis is to get your head straightened out!"

We will be concerned with suggesting--or *prescribing* if you will--how a decision maker (perhaps *you*) should think systematically about identifying and structuring his or her objectives, about making vexing value trade-offs, and about balancing various risks. A few thumbnail sketches of problems will set the stage.

1.1 SKETCHES OF MOTIVATING EXAMPLES

1.1.1 Electrical Power vs. Air Quality *

A mayor must decide whether he or she should approve a major new electric power generating station. There is a perceived need for more electricity but the addition would lead to a worsening of the city's air quality, particularly in terms of the air pollutants: sulfur dioxide, particulates, and nitrogen oxides. The mayor should be concerned with the consequences his actions will have on

a. health effects of residents (on morbidity as well as mortality),
b. economic effects on residents,
c. psychological effects on residents,
d. economic effects to the city, to the state,
e. effects on businesses,
f. political implications.

Each of these broad categories, and others as well, must be clarified and made more operationally meaningful before

---

*This example is discussed in detail in Section 7.1. That discussion makes use of the theoretical concepts introduced in the intervening chapters.

measurements and evaluations can be made and before a delicate balancing of the possible impacts can be systematically undertaken. Even if the consequences of each possible action of the mayor could be foreseen with certainty--which is far from the true state of affairs--he would be faced with a complex value problem.

1.1.2 Location of An Airport*

What should Secretary Bracamontes, head of the Ministry of Public Works, recommend to President Echeverria regarding the development of future airport facilities in Mexico City? Should Mexico modernize its present facilities at Texcoco or build a new airport at Zumpango, north of the city? The decision is not a static one (Texcoco or Zumpango now!) but rather a dynamic one which considers phased developments over a number of years. There are numerous uncertainties, including the possibilities of technological breakthroughs (e.g., noise suppressants, new construction methods for building runways on shallow lakes or marshlands, increased maneuverability of commercial aircraft); of changes in demand for international travel; of future safety requirements imposed by international carriers; and so on. But even if Secretary Bracamontes had his own reliable clairvoyant, his choice problem is still a complex one. He must balance such objectives as

---

*Chapter 8 is devoted entirely to this example.

a. minimize the costs to the Federal Government,

b. raise the capacity of airport facilities,

c. improve the safety of the system,

d. reduce noise levels,

e. reduce access time to users,

f. minimize displacement of people for expansion,

g. improve regional development (roads, etc.),

h. achieve political aims.

These objectives are too vague at this stage to be operational. However, in making them more specific, the analyst must be careful not to distort inadvertently the sense of the whole.

1.1.3 Treatment of Heroin Addiction

Heroin addiction has reached pandemic proportions in New York City and something simply must be done about it. But what? The problem has been studied and restudied but yet the experts differ widely in their proposed strategies. The reason is in part that the problem is so complex that experts have honest differences about the implications of any specific treatment modality. In more technical parlance they differ in their assertions of what a reasonable model of the phenomena should include, and on what reasonable rates of flow from one category to another within the model should be. Therefore their probabilistic predictions of the future vary widely. Once again, if these experts all had crystal balls, disagreements about uncertainties would disappear, but the controversy would still rage. Now however, it would be focused on values only rather than also on uncertainties. The Mayor of New York would like to

a. reduce the size of the addict pool--this is more complicated than it sounds since there are different types of addicts and one must make tradeoffs between sizes of these categories,

b. reduce costs to the city and to its residents,

c. reduce crimes against property and persons,

d. improve the "quality of life"--whatever that may mean--of addicts, including morbidity and mortality considerations,

e. improve the quality of life of non-addicts, make NYC a more pleasant place to live; reverse the disastrous trends of in-and out migration of families and businesses,

f. discombobulate organized crime,

g. live up to high ideals of civil rights and civil liberties,

h. decrease alienation of youth,

.

.

.

z. get elected to higher political office (...perhaps the Presidency?).

Sure, the problem is too complicated, but still one must act and one must informally, if not formally, combine assessments of uncertainties with value preferences*. In this book we shall concentrate on the value side of this type of problem.

1.1.4 Medical Diagnostics and Treatment

Doctor William Schwartz**, Chief of Medicine at Tufts Medical School, makes the rounds of the wards with his student advisees and he drives them mad because he insists on sharing his thought processes with them:"Well, for Z we can do this or this or this, and we must worry about the implications of our actions if she has disease state A or B or C. I think the chances are 0.2 that she has A, 0.4 that .... If we do this and that happens, then we'll learn so and so, which will revise my probabilities of A, B, C by .... But if that happens we must weigh the information we get with the possibility of side effects, discomfort, and costs to Z." And on and on. Very few doctors articulate their thought processes with such clarity. However, they all must, to some extent, constantly combine probabilities with value judgments. And some of these value judgments are not easy to think about. Not

*See Moore [1973] for a formal attempt to examine various policy options concerning heroin use in New York City.

**See Schwartz, Gorry, Kassirer, and Essig [1973].

only are there the usual costs to the patient; cost to the insurance carriers; payments to the doctor; utilization of scarce resources such as doctors, nurses, surgical facilities, and hospital beds, etc., but also one must worry about pain, suffering, anxiety, duration of incapacitation to the patient, ... and, yes, even death. Then there are societal externalities that get mixed up in the value problem: contagion effects, the information gained from one patient that can be of use in the treatment of other patients, development of resistent bacterial strains, and so on. These societal considerations often pose a conflict for the doctor: what's right for his particular patient may not be right for the society. All of this has to get sorted out somehow and decisions have to be made. Can the value side of the problem be systematically addressed? We'll argue affirmatively in this book, but this is not to say that there is an "objectively correct solution". Subjective values will have to be inserted. Our aim will be to develop a framework for assessing and quantifying these subjective values and systematically including them in the decision making process.

1.1.5 Business Problems

Most routine business problems do not involve complicated value issues. Profit, or better yet, the net present value of a profit stream, may be the index to maximize. True, one might have some difficulties clarifying what is fixed cost and what is marginal, but by-and-large these details are conceptually simple. However, top management does not get personally involved in most routine problems. The problems that do filter up to the top often defy a simple dollar-and-cents solution. Ethics, tradition, identity, aesthetics, and personal values in contrast to corporate values, are not uncommon factors to be considered. The more one studies problems of top management, the more one realizes that these so-called uncommon problems are not so uncommon, and the slogan "Maximize profits!" has its operational limitations. We will see, however, that in business contexts it is often natural to try to scale non-monetary intangibles into dollar values. Our concern will be: When is it legitimate to do this and how can it be done?

As top management is all too aware, many of its strategic decisions involve multiple conflicting objectives and, hence, it is simply not true that "qualitatively speaking, business decisions are simple because the objective function is crystal clear".

1.2 PARADIGM OF DECISION ANALYSIS

The simple paradigm of decision analysis* that we will

---

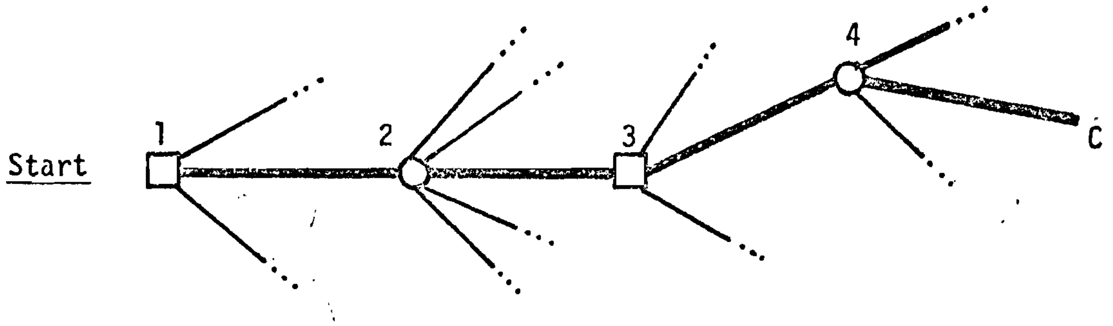
*See for example any of Brown, Kahr, Peterson [1974], Howard [1968], Raiffa [1968], Schlaifer [1969], Tribus [1969], or Winkler [1972].

employ in this book can be decomposed into a five-step process.

Pre-Analysis: We assume that there is a unitary decision maker who is undecided about what course of action he or she should take in a given choice problem. The problem has been identified and the viable action action alternatives are given.

Structural Analysis: The decision maker structures the qualitative anatomy of his problem. What action choices can he take now? What choices can be deferred to later? How can later choices be made conditional on information learned along the way? What experiments could be performed? What information can be gathered purposefully and what can be learned willy-nilly? This melange is put into an orderly package by means of a decision tree as shown in Fig. 1.1. The decision tree has certain nodes where the choice of a branch is under the control of the decision maker (i.e., the nodes depicted with squares in Fig. 1.1) and other nodes which are not under his full control (i.e., the nodes depicted with circles in Fig. 1.1). We shall refer to these two types as decision nodes and chance nodes.

Uncertainty Analysis: The decision maker assigns probabilities to the branches emanating from chance nodes.



Schematic Form of a Decision Tree

[Nodes 1 and 3 are decision nodes; nodes 2 and 4 are chance nodes.] .

Figure 1.1

These assignments are made using an artful mixture of various techniques and procedures based on past empirical data, on assumptions fed into and results taken from various stochastic, dynamic models, on expert testimony (duly calibrated, hopefully, to take into account personal idiosyncracies and biases resulting from conflict of interest positions), and on the subjective judgments of the decision maker. The assignments should be policed for internal consistencies.

Lest there be some confusion resulting from the special schematic decision tree of Fig. 1.1, we note here that we do include the possibility that certain chance nodes can have a set of outcomes represented by a continuum in a singular or higher dimensional space.

Utility or Value Analysis: The decision maker assigns utility values to consequences associated with paths through the tree. In Fig. 1.1 one possible path (from Start to the point labeled C) is shown. In a concrete problem, associated with this path would be various economic and psychological costs and benefits to the decision maker as well as to others whom the decision maker wishes to consider in the characterization of his decision problem. The gestalt is conceptually captured by associating with each path of the tree a consequence which completely describes the implications of that path. The decision maker is then required in this phase of the analysis to register his "likings" for all the possible consequences in terms of cardinal utility numbers*. This measurement reflects not only the decision maker's ordinal rankings for different consequences (e.g., C' is preferred to C" which is preferred to C"') but it must also indicate his relative preferences for lotteries over these consequences. For example, in Fig. 1.2, we consider a choice problem between act a' and a" which gets translated into a choice between lottery ℓ' and ℓ''. The decision maker must assign numbers to consequences (such as u_i' to C_i' and u_j'' to C_j'') in such a manner that he feels that

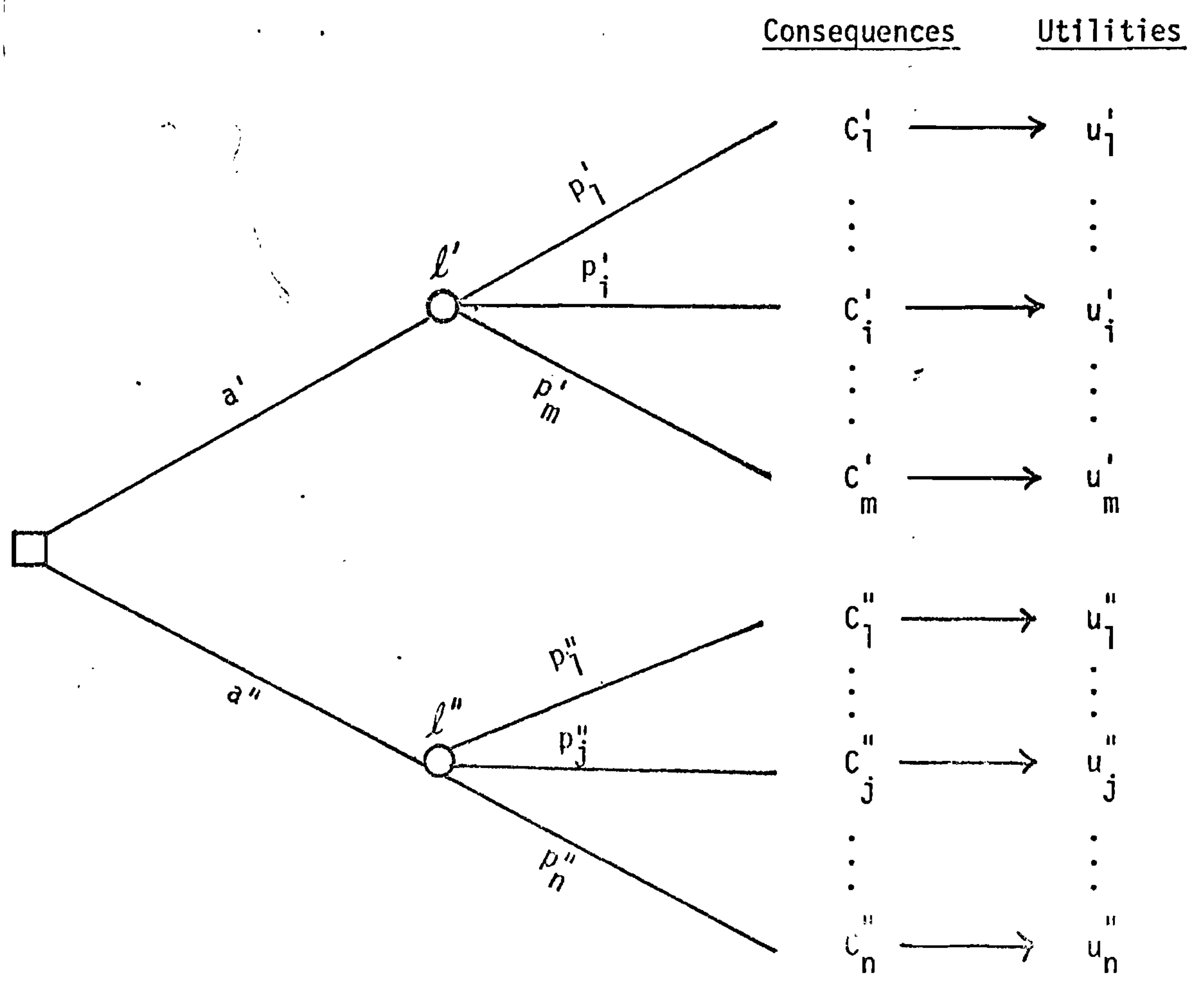
$$(a' \text{ is preferred to } a'') \Longleftrightarrow \left(\sum_{i=1}^{m} p_i' u_i' > \sum_{j=1}^{n} p_j'' u_j''\right) .$$

In other words the assignment of utility numbers to consequences must be such that the maximization of <u>expected utility</u> becomes, tautologically, the appropriate criterion for the decision maker's optimal action.

<u>Optimization Analysis:</u> After the decision maker structures his problem, assigns probabilities, and assigns utilities, he calculates his optimal strategy--that strategy which maximizes expected utility. This strategy indicates what he should do at the start of the decision tree and what

---

*Throughout this book, we assume that the reader has some familiarity with cardinal utility theory. However, in Chapter 4, we do review aspects of the theory which will be needed.



A Choice Problem Between Two Lotteries

Figure 1.2

choice he should take at every decision node he can possibly get to along the way. There are various techniques an analyst can employ to obtain this optimal strategy but the simplest is the usual dynamic programming algorithm of averaging-out-and-folding-back, with which we assume the reader is already familiar*.

1.3 COMMENTS ABOUT THE PARADIGM

Now is this a reasonable paradigm for the class of problems we sketched at the outset: problems of air-quality control, of location of an airport, of treatment modalities for heroin addiction, of medical diagnostics and treatment, of strategic business problems?

1.3.1 Unitary vs. Group Decision Making

First of all throughout most of this book - all but Chapter 10 - we assume that there is a unitary decision maker. Should we not be more concerned with group decision making? Aren't most public decisions and many business decisions an intricate composite of different choices made by many individuals? Let's take an example.

New York City is concerned with the poor quality of air being breathed by its residents. Should the city government impose more stringent limits on the sulfur

---

*See for example, Raiffa [1968], pages 21-27 and 71-74.

content of fuels burned in the city for space heating and power generation? Lots of people are involved in settling this problem: the mayor, city council, Environmental Protection Agency, lobbyists for power companies, political parties, the citizenry, and so on. Any after-the-fact description purporting to explain what has happened in any past period certainly must involve many individuals: Descriptively it is a group, interactive, decision problem.

But wait! What we are trying to do here is not to describe what has been done but to prescribe what should be done*. Let's first clarify for whom we are prescribing. Who is the client for our proposed analysis? Well suppose it is the head of some appropriate agency. He alone surely does not dictate what will eventually happen but he might be called upon to make a proposal to the mayor, for instance. Suppose he's confused about whether he should offer proposal A or B or C. Well the agency head has a decision problem, has he not? He might want to analyze systematically what he should do. He must consider what other actors in this "game" might do and perhaps he might want to view the actions of the mayor and the city council as part of the uncertainties confronting him. One individual's decisions may be another individual's uncertainties.

The point that we wish to emphasize is that decisions,

---

*Clearly there is much overlap of interest between the prescriptive and descriptive viewpoints. Over the past twenty-five years, the contributions of many people addressing descriptive aspects of decision making has had a significant impact on prescriptive decision analysis. Four excellent reviews of this work are Edwards [1954,1961], Slovic and Lichtenstein [1971], and Fischer and Edwards [1973].

as we use the term, do not have to be grandiose end-determinations. There are more modest decisions: should an individual vote for passage of a bill, propose an amendment, apply political pressure, and so on. If such an individual has choices to make, we can view him as the decision maker. It is in this sense that we can assert that there are many decision problems in the public sector where the decision maker can be viewed as a well-specified, identifiable, unitary entity. Now some of these decision makers, some of the time, might want to analyze their particular problem in a systematic manner. In this book we're concerned with effectively adapting the decision paradigm outlined in the preceding section to help such a decision maker.

1.3.2 Personal Conviction, Advocacy, and Reconciliation

Throughout this book we approach problems from the point of view of an as-yet-undecided decision maker who wants to decide and convince himself of the appropriate course of action he should take. He recognizes that some of his snap judgments may turn out to be wrong in the sense that he might change his mind after deeper reflection. He also recognizes that when a problem is decomposed into parts, he might initially give answers to a series of questions that turn out to be internally inconsistent. When this occurs we shall assume that the decision maker will want to scrutinize his answers carefully and perhaps change some of his earlier responses so that the total pattern of modified responses is consistent and seems reasonable to him. Only if he can structure his preliminary responses in a coherent fashion, will we be able to use deductive analysis to carry him to the next step of commitment. The spirit is one of Socratic discovery, of unfolding what one really believes, of convincing onself and deciding.

We authors have found that in many of our consulting contacts, decision makers embark on formal decision analyses with their minds already made up at the start. You can view the formal analysis as just a sort of window dressing. We don't want to preach against such activities; rather we merely want to emphasize that in this book we want to address that class of problem situations where the unitary decision maker has not as-yet "made up" his mind. But, in passing, let us also remark that there is often a legitimate purpose for doing careful analyses even if the decision maker has already decided what to do prior to the analysis. First, there is the problem of psychological comfort: he might want the security of having a formal analysis to corroborate his unaided intuition. Secondly, he might want to use the formal analysis to help the communication process. Thirdly, there is the question of advocacy: he might have to justify his conclusions to others or to convince others of the reasonableness of his proposed action. In addition, there is always the possibility that these post-decision analyses will uncover new insights that result in a change of the chosen alternative, one which is perceived as better from the decision maker's viewpoint.

Indeed an analysis done solely to convince oneself might be quite different from one done for advocacy purposes. A personal analysis might very well incorporate very sensitive information, such as assessments of potential future actions of political associates, an economic value placed on the life of a human being, value tradeoffs between the benefits to various identifiable groups, and so on. On the other hand, an advocacy document must often be intentionally vague on such issues.When an analysis is put on public display one can hardly expect one's adversaries to give up without a fight. They will carefully scrutinize the reasoning and seek out the soft spots. This unfortunately means that it is often impolitic to base a decision on a formal analysis which includes subjective feelings if the analysis will be disclosed to a critical public audience. This is not the place for us to get involved in questions about moral obligation on the part of government officials to be open and honest or to share their real analyses with other government officials, agencies, and concerned citizenry. To repeat once again, we are primarily concerned in this book with techniques to help a confused decision maker make up his mind.

There is yet another reason why one might do a formal analysis of a decision problem even though one's mind is already made up. Although what we have in mind might be considered a variation of an advocacy role we prefer to look at it more constructively in terms of a reconciliation process. As an example, suppose a mayor must decide what to do and two agencies strongly recommend that he do different things. The rhetoric is sharp and divisive; the protagonists, eloquent and able; and the situation suitably complex so that there is apparant merit on each side. How can the decision maker weigh the arguments and make a responsible decision?

A formal analysis which attempts to decompose the overall problem into component parts can often help this reconciliation process. Perhaps the parties can agree on what they agree about and what they disagree about. Perhaps they can further decompose areas of disagreement in a manner to highlight fundamental sources of differences of opinion. Would the collection of more information help to sort out the merits of the two positions? Could they agree on what additional objective (or even subjective) evidence could help them decide? Or is it not a matter of assessment of uncertainties but of differing value judgments? Perhaps here is the place where the mayor could exert his own overriding value structure.

We don't want to appear excessively naive by implying that formal analysis which decomposes a complex problem into smaller more manageable component parts is the key to the reconciliation process. We are well aware that, in some circumstances, the more confusion that abounds the easier it will be to establish a compromise. But still, in principle at least, we think that in some circumstances (how's that for a hedge?), familiarization could facilitate reconciliation. And furthermore, we shall report in Chapter 8 an example of just such an undertaking in which both of us were involved as consultants. We were only partially successful.

1.3.3 Pre-Analysis and the Iterative Nature of an Analysis

As we indicated previously, we assume that the decision maker's problem has already been identified and viable action alternatives are prespecified. This is not to say that, in practice, the preliminaries are not crucially important. By some creative insight, one must not only recognize that a problem exists, but one must have an intuition about what types of problems are worth attempting to analyze in a systematic, scientific manner.

Complex problems, especially in societal contexts, tend to have spillover effects in all directions. Thus, bounding a problem is critically important. We all know the dangers of sub-optimzation but if problems are not bounded in some way, they remain hopelessly intractable. The process of identifying and bounding a problem area is intimately connected with the generation of alternative decision choices to be considered. When we make the assumption, as we do in this book, that the alternative decision strategies are prespecified, we seriously misrepresent the art of formal analysis. In practice, the process is an iterative one. The analyst might bound his problem one way only to find that he's posed an impossible morass; so he backs up and redefines his problem area: he bounds it differently and generates new restricted alternatives to consider. Or in the course of analysis, he recognizes that the conclusions he draws are sensitive to one given facet of the problem that has not been delicately enough modeled; If this happens, he may redesign the structure of the model. It has also been our experience that a careful analysis of the posed problem often helps to trigger a line of thought that generates action alternatives which might have been overlooked otherwise. Yes, we do recognize the iterative nature of the overall process of analysis but for our purposes, with all due apologies, we will assume henceforth that the pre-analysis stage has been completed.

It is our impression that even experienced analysts often fail to exploit sufficiently the usage of adaptive and process-oriented action alternatives. It is not only important for the analyst to know what must be done now and what he can defer to the future, but also it is critically important that he recognizes the possibility that future actions could be made dependent on information learned along the way. A dynamic strategy for action should be adaptive and exploit the gradual, time-dependent unfolding of uncertainties. The decision-tree framework of analysis is especially suitable to promote thinking about adaptive, time-dependent, action alternatives. However, it does not help us in thinking about process alternatives. Let us explain.

"You analysts want to decide on everything," a nameless voice exhorts. "Why decide at all? Let the contending factors address the issues in an open, democratic process." Well often that advice is right. Establishing a process may be that creative new alternative we alluded to earlier. Still someone might be in a position where he must decide whether decision strategy A or B or decision process C or D should be adopted. And that is a decision problem. Furthermore, if say process C is selected then amongst the host of decision makers who will influence the actual denouement, there may be one confused, analytically-minded soul who wants to get his mind straightened out by means of the decision framework we are espousing.

We do not deny the point that it is often desirable to institute an advocacy process for resolving complex issues in the public domain. However, we do not think that this assertion necessarily diminishes the usefulness of the decision analytic framework. It may, of course, influence the nature of the problems to be analysed or the identity of the decision maker who employs these tools. As a last point on the subject of process, we remark that the decision analytic framework can in some applications be employed to help structure the process of debate and action.

1.3.4 Subjective Values and Formal Analyses*

It is almost a categorical truism that decision problems in the public, societal domain are complex--too complex. They almost universally involve multiple conflicting objectives, involve nebulous types of non-repeatable uncertainties, involve costs and benefits accruing to various individuals, businesses, groups, and other organizations--some of these being nonidentifiable at the time of the decision--and involve effects which linger over time and reverberate throughout the whole societal super-structure. It would be nice if somehow we could pour this whole mess into a giant computer system and program the superintellect to generate an "objectively correct" response. It just can't be done! You can only go so far without the introduction of subjective attitudes--no matter how hard one squeezes the available objective data it won't come close to providing courses of action for complex problems. Indeed, a purely 'objective' analysis might fall so far short of providing guidelines for decision making that the output of the analysis may not

---

*This and the following subsection liberally adapt material from Keeney and Raiffa [1972].

pass the threshold of relevancy. It is our opinion that complex societal problems--and for that matter, complex <u>business</u> problems also--demand the consideration of subjective values and tradeoffs. The question, as we see it, is not whether subjective elements should be considered, but rather whether they should be articulated and incorporated into a formal, systematic analysis. The choice is between <u>formal analysis</u> and <u>informal synthesis</u> and this metadilemma does not have an obvious solution.

How often we have heard the general expression that formal analysis is inappropriate for complex problems, since these problems require subjective evaluations. Of course they do, but the fact is that formal decision analysis stands ready to receive such subjective evaluations as inputs for the decision algorithm. The trouble with formal analysis is <u>not</u> that subjective evaluations cannot be accommodated into the framework, but rather that there is a demand for <u>too many</u> subjective inputs; and although decision makers argue for inclusion of subjective evaluations they tend to be most reluctant to put these evaluations down in black and white on paper.

There is a widely held feeling that one should beware of those analysts that try to <u>quantify</u> the <u>unquantifiable</u>. But let us remember that it is also a grievous sin for us not to learn how to quantify the quantifiable. The question is: <u>What is quantifiable?</u> An art expert might be hard pushed to give an objective formula for ranking the quality of paintings, nevertheless he might be able to rank-order these paintings, saying in effect that if given a choice between two paintings he would prefer the one that has a higher place in his ordering. And where we have rank-orders, numbers can't be far behind. Our artist might even be willing to put a price tag on each painting; thereby quantifying one aspect of his subjective judgment. This sort of quantification is not done by means of an objective formula but by subjective introspection. Is it legitimate to work with such numbers? We do it all the time. As analysts we must learn how to incorporate soft, squishy considerations (such as aesthetics, psychic factors, and just plain fun) into our analyses. If we don't learn how to do this, the hard will drive out the soft and efficiency--very narrowly interpreted--will prevail.

On the other hand, the quantification of these subjective factors cannot be done frivolously. They should be generated by making the best use of the accumulated experience and expertise available. And on problems of public concern, such as power plant siting, this quantification should undergo the scrutiny of independent 'experts' as well as the concerned citizenry.

1.3.5 Strategic vs. Repetitive Decisions

There is a feeling that formal analysis is appropriate for repetitive operational decisions--like: "where should we send the sanitation trucks today?" or "what procedures should be used for operating airport runways in order to minimize travel delays?" or "what should we charge for breakfast cereal WOW?". But the feeling goes that analysis is nigh-on impossible for those one-of-a-kind, strategic decisions, like: "Should we dispense methadone to heroin addicts?" or "Should we spend 200 million dollars for research on nuclear breeder reactors?" or "Should the Mexican Government build a new airport miles from Mexico City or modernize the old?" or "Should Corporation X internationalize its marketing operations?". No one claims it is easy to analyze complicated strategic problems, but we believe that many of these strategic policy-type questions are amenable to systematic attack.

1.3.6 Implementation, Post-Analysis, and Other Considerations

Other than the very few brief remarks we are about to make in this paragraph, we will say nothing about another critical aspect of an integrated analysis--the implementation phase. By the implementation phase we mean to include all those indispensable activities that go on in order to execute the chosen strategy which results from a given analysis. This includes the communication of instructions, the delineating of responsibilities, the establishment of incentives and rewards, the punishment of willful deviations, the monitoring of the system, the systematic collection of data, the creation or adaption of a management information system, the dissemination of reports, the further refinement of the model, identification of new key variables, creation of new alternatives that were overlooked, and so on. In practice, it is artificial to completely divorce the identification and analysis of a problem from the problems of implementation. Clearly what's called for, once again, is the ability to iterate. If a suggested solution cannot be realistically implemented, then the analysis must be redone with some attention paid to constraints imposed by the implementation phase.

As long as we are still talking about things we are <u>not</u> going to do, let us also mention a few other questions we are <u>not</u> going to address: How do good analyses get done? How can you choose good analysts? Should you use outside consultants or an inside group? Where in the organization hierarchy should an analytical capability be created? How does the introduction of an analytical team shake-up an existing bureaucracy? On all of this our contribution is Silence--except for the gratuitous platitude: The decision of whether or not to do formal analysis cannot be divorced from the question of organizational structure, of the personal incentives of the people involved, and of the quality of the analysts.

We hope that our non-existent treatment of the crucial considerations of the analytical process raised in this section is not interpreted as belittling their importance. Indeed we won't be insulted if readers claim that we have only scrutinzed a part of the entire problem because we are doing this with some awareness.

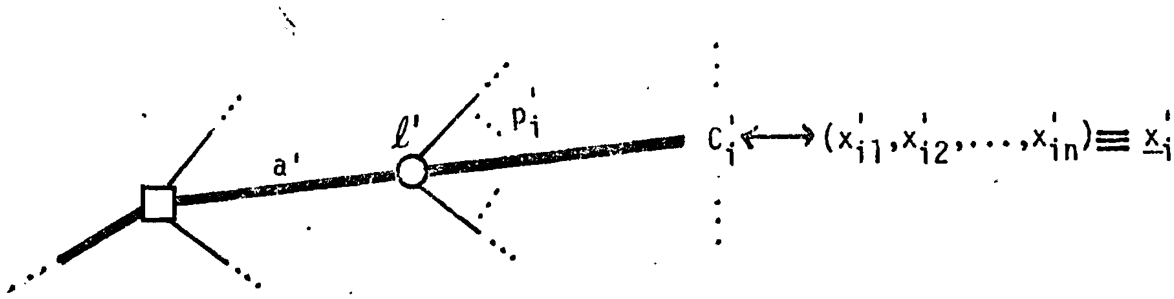
1.4 COMPLEX VALUE PROBLEMS

1.4.1 Simple versus Complex Value Problems

Consider a decision maker who has already decided on the identification and bounding of his problem and has generated the set of alternative actions he wishes to consider. Let's assume that he has structured his problem in the form of a decision tree, and by one device or another has assigned probabilities to all the branches of chance nodes. We enter into the phase of the problem where he is contemplating the encoding of his preferences for consequences. Let's turn back to Fig. 1.1 and look at one path through such a tree and consider its consequence C, depicted at the terminus of the path. In some problems it is possible in a purely objective manner to assign a single number to each consequence C that adequately describes the full implications of that path. For example, in a business problem the single numerical value might be a monetary value which fully reflects all the financial considerations of the problem and there may be no other considerations to worry about. In a medical context, a possible single summary number might be a cure rate for a given disease. In such problems consequences are adequately described in terms of an objective, single numerically scaled attribute--"numeraire", for short. Let's suppose the value associated with consequence C' is X(C') = x' and with C" is X(C") = x". Here x' and x" are real numbers. Also assume that preferences are such that C' is preferred to C" when and only when x' > x". (This last assumption is made for convenience and can be trivially generalized.) Problems of this genus will be called <u>simple value</u> problems in contrast to <u>complex value</u> problems. In complex value problems, consequences at the ends of the tree can <u>not</u> be adequately described in objective terms by means of a single numeraire (e.g., money). Our main concern in this book is with complex value problems.

Simple value problems would be conceptually trivial to solve if there were no uncertainties involved--if there were no chance moves in the tree. This would then boil down to a straightforward maximization problem with a well specified payoff function. There is another way of saying all this might be helpful. Imagine a decision problem abstracted in the form of a decision tree. If a decision maker had the services of a perfect predictor (i.e., a clairvoyant, or as a colleague of ours, John Lintner, puts it, "if he had a phone line to the Lord"), would his problem be conceptually simple? It would be, if every consequence were already described in terms of a single numeraire. He would just choose that strategy leading to the highest x-payoff.

In Fig. 1.3. we schematically show a section of a



Section of a Tree Resulting in a Complex Consequence

Figure 1.3

decision tree with one path that ends in consequence C_i'. Now let's suppose that C_i' can be adequately described in objective terms only by means of n numbers: x_{i1}', $x_{i2}',\ldots$, x_{in}'. We can think of the number x_{ij}' as the performance measure of consequence C_i' on the jth attribute scale. When the decision maker contemplates taking action a' in Fig. 1.3, he is led to consider the lottery ℓ' which, with probability p_i', results in a consequence described by an n-tuple $\underline{x}_i' \equiv (x_{i1}',\ldots, x_{in}')$, where the i-subscript ranges over the number of branches of the chance node. In slightly more technical parlance, lottery ℓ' can be interpreted as a discrete probability distribution with outcomes in an n-dimensional space. The decision maker must clarify in his own mind which one of these n-dimensional distributions he would rather choose. No easy task, this. How can he think in a systematic manner about this?

Notice that if the decision maker has a clairvoyant his problem would not become trivially simple. It would be easier to be sure, since there would be no uncertainties, but he still would be faced with a complex value problem of the type: given possible ending consequences C_1, $C_2,\ldots$, C_q where C_i is described in terms of $\underline{x}_i = (x_{i1},\ldots,x_{in})$, which consequence should he prefer? This choice problem involves complex value tradeoffs.

Let's return to the uncertainty case as depicted in Fig. 1.3. In purely formal terms, the problem can be answered by the introduction of a utility function u which would associate to each n-tuple a single real number. Let $u(x'_{i1},\ldots,x'_{in})$ be denoted by u'_i. In this case the relative desirability of lottery ℓ' would be given by $\Sigma_i\ p'_i u'_i$, the expected utility of lottery ℓ'. In terms of expected utilities we can now work backwards through the tree in order to pick out the optimal strategy. Pretty easy. The rub is, of course, it isn't so easy to find an appropriate utility function u. Some would say it's impossible to do this in a responsible manner. Our task in this book is to indicate techniques that one might employ in helping oneself discover an appropriate u function. We will discuss in the sequel some basic principles for decomposing the overall complex value problem into more manageable and "thinkable" component parts. Some of these decomposition principles we feel are so basic that they might profitably be employed by some to partially structure their value problem even though they might be reluctant to go the whole hog--to go all the way to the determination of an overall utility function. How far one should go in formalizing one's value problem depends on so many factors: on its importance, on the need to convince others, on one's training,..., and on the availability of techniques that can be employed in the thought process.

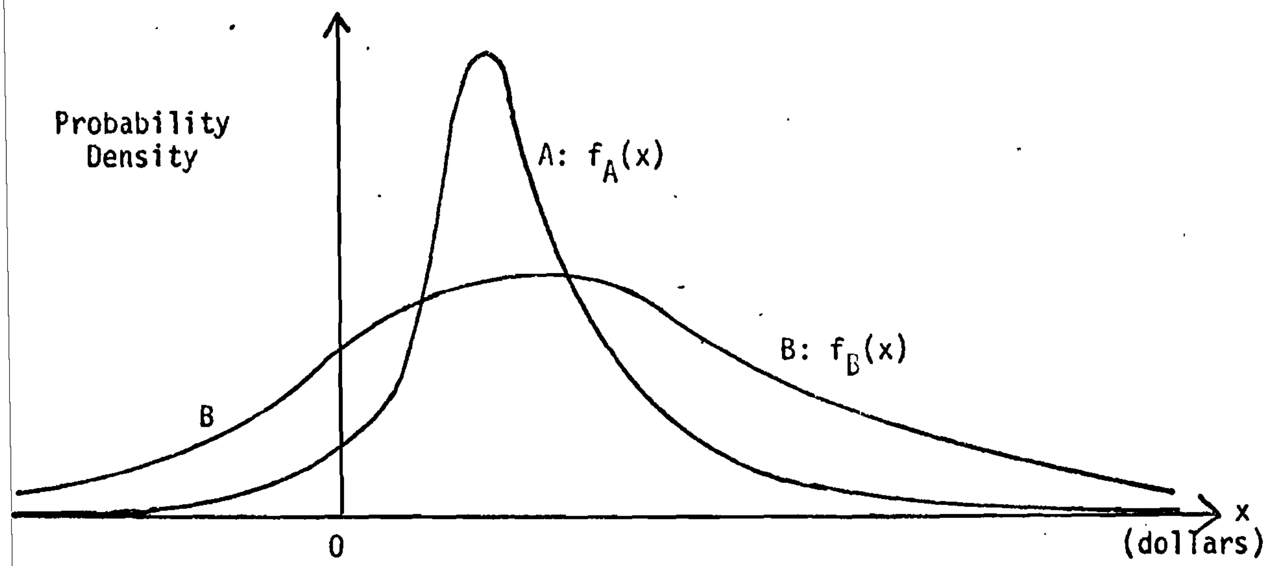
1.4.2 Is Utility Analysis Necessary?

Those who have worked on problems in decision analysis can readily testify that it's hard enough to get responsible utility functions for a single numerically scaled attribute, like assets, and one must admit furthermore that such techniques are only rarely used in practice. Should anybody take seriously, then, an endeavor which tries to do the same thing for higher dimensional space? If you haven't completely succeeded in one-dimensional space why go to 10-space? Let's leave aside the response that there are lots of nice mathematical theorems to prove and it's a fertile field of new theoretical development. Can the theory to be developed have any practical value? We think so, and let us say why.

First consider the unidimensional case. Suppose that the decision maker must decide between actions A and B which result in the probabilistic, monetary payoff distributions shown in Fig. 1.4. It's not immediately clear which distribution should be chosen and a formal analysis could be made by introducing a utility function u and then comparing

$$\bar{u}_A = \int u(x) f_A(x) dx \quad \text{and} \quad \bar{u}_B = \int u(x) f_B(x) dx.$$

But in practice this bit of formality is usually sidestepped. Instead the decision maker looks at the distributions f_A and f_B, which, in the unidimensional case, as contrasted to the multidimensional case, can be visually interpreted. He then subjectively reacts to the whole distribution and comes to a choice unaided by formal utility analysis. We, the authors, personally would prefer to introduce the



Comparison of Two Distributions of Payoffs

Figure 1.4

formality of a utility function if we were personally responsible for the decision, because we have trained ourselves to think hard about what we want our utility function to be and thus we would feel more comfortable with the derived results than we would be if we reacted directly. But experience has also shown us that our attitude is not commonly shared, even amongst business executives who have been adequately exposed to the concepts of utility analysis. In the unidimensional case they can circumvent the formal approach by acting intuitively to easily comprehended alternatives.

Now let's contrast the above unidimensional case with a choice involving many attributes. Actions A and B lead to complicated distributions not over a single x but over n-tuples $(x_1,\ldots,x_n)$. No longer is it possible to draw the distributions in a simple manner and the mind boggles at the complexity. No wonder that in practice decision makers introduce pragmatic simplifications, such as "Let's just look at the most important attribute and forget the others," or, "Let's not worry about uncertainties but take some value of central tendency for each attribute and set up aspiration levels on each of these attributes." Decisions get made on the basis of ad hoc, heuristic simplifcations. We authors believe that many, though not necessarily all, of these decision makers would be better served if they systematically probed their value structure and created for themselves a derived utility function. How this can be done, will be the subject matter of this book.

1.4.3 The Use of Hypothetical Questions in Assessments

A fundamental principle of decision analysis is to separate the preference inputs from the modeling and probabilistic inputs that enter the decision analysis. Therefore, we must ask hypothetical questions to obtain the preferences of the decision maker. The approach is to ask simple questions involving simple probability distributions which are intended to focus on the basic preferential attitudes of the decision maker. Then, the answers to these simple hypothetical questions are consistently put together to provide (hopefully) the information necessary to arrive at a specific utility function. Our feeling is that it is easier for the decision maker to understand his own preferences and articulate them in a form useful for constructing his utility function by answering questions in these simple contexts rather than in complex situations. In checking the consistency of any such utility function, we would suggest a comparison of the implications of the utility function with the decision maker's responses to "more realistic" probability distributions as a first step toward ascertaining whether the use of hypothetical questions contributed to a systematic error in the utility function.

Critics of decision analysis often attack the use of hypothetical questions in the assessment procedure. However, for any problem, every question concerning preferences addressed to the decision maker other than "Which of your real options do you prefer?" is by definition hypothetical. It appears that if it is desired to have any analysis whatsoever of the problem, hypothetical questions will necessarily have to be asked concerning parameters in any model, probabilities of various outcomes, preferences, etc. Thus, if analysis is deemed worthwhile, an important point is the degree of hypothetical questioning and not whether any hypothetical questions should be used.*

Of course, the particular phraseology of the hypothetical questions should be in a vernacular that's comfortable to the decision maker. (See, for example, Grayson [1960]). The trick is to be as realistic as possible but still to pose hypothetical questions that are easily understood and precise. Compromises, of course, have to be made and an analyst often

---

*It has been suggested that by observing how a decision maker <u>does</u> make decisions, his preference structure can be derived. If these 'revealed preferences' are to be used for normative purposes one must assume the decision maker has made 'optimal' decisions in the past. Another assumption is that one can separate the decision maker's perceptions (i.e., probabilities) in previous problems from his preferences (utilities). It seems to us that these two assumptions lead one to conclude the 'revealed preferences' alone simply do not provide enough information to specify a decision maker's preference structure, especially when interdependent uncertainties and multiple objectives are both involved.

has to go to artful extremes when his respondent has a low threshold for hypotheticality. For some problems one might begin with more complex, more realistic questions involving many of the critical issues of the problem and work toward simpler questions focusing on single critical issues. In the process, it may be possible to sensitize the decision maker to these individual issues and, hence, increase his respectiveness to thinking hard about the 'hypothetical' questions involving them. This in turn might help clarify his thinking.

1.5 CLASSES OF EXAMPLES AND METHODOLOGICAL NIGHTMARES

We have a two-fold purpose in this section. First we would like to cite a few broad categories of methodological problems that fall in our domain of concern. In contrast to the motivating examples we have already mentioned (like air pollution, power generation, heroin addiction, airport location, and so on), we now look at categories of problems that are organized by methodological type--problems such as cost-benefit and cost-effectiveness analysis, analyses involving time-streams of payoffs (e.g., discounting), and analyses of awesome consequences such as deaths. Our second purpose is to mention briefly a host of issues that we feel are crucially important and relevant to our domain of discourse but which we do not do justice to in this book. We refer to some of these as methodological nightmares.

As our point of departure for this section, let us consider an abstraction of a real problem that results in a decision tree where each consequence C is described in terms of n attributes $X_1,\ldots,X_n$.

1.5.1 Private Decisions or Individual Cost-Benefit Analyses

Mr. Smith is the decision maker and his actions will only affect himself and not others. When he totes up the ledger resulting from any action he might take, he might be concerned with various costs and obligations that will accrue to him (assume that for the time being that these costs and obligations are immediate) and with various benefits in terms of money, prestige, power, sense of community responsibility, and so on. In a particular example of this type we might have the following identifications:

$X_1 \equiv$ out-of-pocket costs to Smith

$X_2 \equiv$ measurement of time commitment

$X_3 \equiv$ monetary rewards

$X_4 \equiv$ combined index of psychological satisfaction (other than financial).

Now there are a lot of questions that will immediately come to your mind that we do not want to address until the next chapter. Some of them are: How should one generate these objectives? What about overlaps? What about measurement problems (e.g., with X_4)? What about completeness? What about uniqueness of the set of attributes? What could be done with the evaluations if they were made? And so on.

You might want to anticipate some of the discussions in the next chapter by thinking of various categories of individual choice problems. What would be a reasonable set of attributes to consider in the choice of (1) a job, (2) a house, (3) a car, (4) a spouse, (5) a birth-control technique (if any), (6) a college (see Hammond [1965]), (7) a summer vacation?

1.5.2 The Case of the Altruistic Dictator - A Social Welfare Problem

In contrast to the preceding example, let's imagine that Ms. Tate must decide what she, as an agency head, should do. In this case she is concerned with the way in which her actions will affect the costs and benefits to diverse individuals, business firms, and other identifiable organizations. She is also not completely altruistic because she must worry about the implications to her agency and to herself in particular. Her decision might be complicated by the fact that she might not know how a segment of the concerned citizenry really feels about a given societal modification. True, she (or others) can ask them (or a sample of them) but it's not always easy to do.

In one small part of his doctoral dissertation, Jan Acton [1970] conducted a door-to-door sample survey in which he asked heads of households what they would be willing to pay for an emergency coronary care unit in their community. Well, there were all kinds of problems with this. Most people just weren't willing to take time to try to understand what the issues were. Even if they took the time, it's not clear they knew how to think in a reasonable manner about such a complex issue. But even if they took the time, and could think straight about their own interests, then what about honesty. After all, why is this guy asking me this question?

Then there is also the problem of present-versus-future tastes. Our benevolent despot might be of the firm conviction that her subjects really don't know what's good for themselves. Those poor souls don't know that if they only listened long enough to classical music they'd eventually like it. Sure those misguided dupes voted against the bond issue for improved schools, but if they only knew what a good school system is really like, then they would have voted for the bond issue. The populace is not interested in pollution now since they are more interested in the wherewithal for daily survival, but in time things will improve and they will be concerned with air quality.

The methodological issues these points raise are devilish to work with. Still, decisions must be made.

1.5.3 Cost-Effectiveness and Cost-Benefit Analyses

Consider a given decision problem where one possible consequence resulting from a given action (or strategy) can be adequately described in terms of a cost C, and r benefit measures $B_1, B_2, \ldots, B_r$. In this case it will be easier to think of the description in the form $(c, b_1, \ldots, b_r)$ where a small letter designates a specific amount of the respective measure rather than in the less suggestive but more neutral notation $(x_1, \ldots, x_{r+1})$. It is important to understand that these benefit measures may be in different units of measurements so that one cannot simply add them up. For example, B_3 may be in man-hours of work saved and B_7 may be measured in the architectural quality of a given building.

Leaving aside uncertainty for a moment, we can imagine that any agency head has a specified amount c^* that he may spend on projects. His objective, in loose terms, is to accept all those projects which, in totality, do not exceed his cost constraint c^* and, subject to that, will yield a desirable portfolio of joint benefits. This problem is difficult to make more precise since the various benefits are in incommensurable units and not much can be done with coalescing these various separate entities.

In <u>cost-effectiveness</u> analysis no attempt is made to combine the various benefit measures into one single, composite, benefit measure. In a cost-effectiveness analysis, one might investigate a problem of the form, "Characterize various sets of projects which yield at least b_1^*, $b_2^*, \ldots,$ b_r^* on the respective benefit measures." Here the so-called <u>aspiration levels</u> $b_1^*, \ldots, b_r^*$ are usually preassigned. Are there feasible sets of projects that will meet these combined aspirations? If not, one changes some of the b_i^*'s. If yes, one investigates whether he can squeeze out a bit more by raising some of the aspiration levels. Of course, this leaves out of the formal analysis two important relevant considerations:

1) How should one select the aspiration levels in the first place? What should the tradeoffs be

amongst them?

2) How can all this be generalized to bring in the everpresent problem of uncertainties?

In <u>cost-benefit</u> analysis, in contrast to cost-effectiveness analysis, one takes the heroic step of collapsing the benefits $B_1, \ldots, B_r$ into a single composite measure, B_0 say. One usual technique is to introduce a set of conversion factors $w_1, w_2, \ldots, w_r$ and then one defines

$$b_0 \equiv w_1 b_1 + \ldots + w_i b_i + \ldots + w_r b_r .$$

Of course the units of measurement of the w_i's are such that the individual summands $w_1 b_1, w_2 b_2, \ldots, w_r b_r$ are all in commensurable units. The trick, in practice, is to find suitable conversion factors. Often this is done by some objective market mechanism or one subjectively imputes dollar prices of monetary worth to various measures (e.g., a dollar value of \$ 500 to keep each child off the street during the summer months).

If we go to cost-benefit analysis, and if, once again, we leave aside uncertainties for a moment, then the kth project can be evaluated by the pair $(c^{(k)}, b_0^{(k)})$ where $c^{(k)}$ and $b_0^{(k)}$ represent the cost and the composite benefit measures.

One can imagine having a list of possible projects with cost and composite benefit measures for each. Now <u>if</u> the problem is "Select a subset of projects to maximize the sum of benefits subject to the constraint that the total cost does not exceed a preassigned c^*," then the analysis calls for the ranking of the projects according to benefit-cost ratios R_k (i.e., $b_0^{(k)}/c^{(k)}$ for the kth project) and accepting projects in order until the cut-off c^* is reached. Let's ignore the problem of indivisibilities, what to do with a fractional project.

It is much easier to come to definitive answers using cost-benefit analyses than cost-effectiveness analyses. And therefore, it is not surprising that many studies go this route. Of course, one must be careful to observe the legitimacy and the reasonableness of the transformations that collapse $b_1, \ldots, b_r$ into b_0 and then collapse c and b_0 into R. All too often, in practice, important benefits are not included in the listing because it is not clear how a market mechanism can be conjured up to "price out" this particular benefit. We're thinking here of such benefits as aesthetics, psychological well-being, security, and so on.

From our point of view there are several difficulties with both cost-effectiveness and cost-benefit studies. Both suffer from an inability to cope with uncertainties in an operationally reasonable or theoretically sound manner. That's not to say that ingenious efforts have not been made. But, by and large, we believe the utility approach we take in this book is a more systematic way of handling uncertainties. Admittedly we have to pay the price of increased complexity. Also, as we will emphasize in the sequel (see Section 3.8), it is <u>not always appropriate</u> to collapse an r-tuple of benefits $(b_1, \ldots, b_r)$ by means of a simple linear weighting rule

$$w_1 b_1 + \ldots + w_i b_i + \ldots + w_r b_r ,$$

or even by a generalized-linear rule

$$w_1 g_1(b_1) + \ldots + w_i g_i(b_i) + \ldots + w_r g_r(b_r)$$

using suitably chosen non-linear functions (transformations) $g_1, \ldots, g_r$. The legitimacy of these procedures will be systematically analyzed throughout this book.

1.5.4 <u>Temporal Considerations: Present vs. Future</u>

Our society is often accused of selling its future generations short. In an attempt to ameliorate our immediate woes we often act in a manner that exacerbates our future problems. Analysts must constantly make tradeoffs between what is right for the present generation and what is right for future generations. Some think that we're worse off today than we were in the past and that this trend is likely to continue in the future. Others feel that future generations are going to be better off than we are today and it's reasonable to borrow from the future to improve the present. What obligations do we have to future generations? Should the future be given more weight just because there will be more people in the future than in the present? It seems that as our time perspective unfolds, our spatial concerns grow too:

today and tomorrow, it's our family; in the decades ahead, it's our country; in the centuries ahead, it's the world's population; and in the millenia, it's the planet Earth.

On a more mundane level, government agencies are concerned with finding an appropriate rate of discount. Should we do research on the development of a new nuclear breeder reactor? Well, a lot depends on whether we use a 5%, 10%, or 15% rate of discount. Or is any discount rate appropriate? Let's look at what these problems involve.

Consider a given decision problem where one possible consequence resulting from a given action can be (just adequately) described in terms of a stream of costs c_1, c_2, ..., c_t, ..., one for each time period t, and of r different streams of benefits:

Benefit stream of type 1: $b_{11}, b_{12}, \ldots, b_{1t}, \ldots$

$\vdots$

Benefit stream of type i: $b_{i1}, b_{i2}, \ldots, b_{it}, \ldots$

$\vdots$

Benefit stream of type r: $b_{r1}, b_{r2}, \ldots, b_{rt}, \ldots$.

We are not complicating things here needlessly. This is the prototypical problem found in most cost-effectiveness and cost-benefit analyses of societal problems.

In order to make the mass of numbers

$$\begin{bmatrix} c_1 \ , & c_2 \ , & \ldots, & c_t \ , & \ldots \\ b_{11}, & b_{12}, & \ldots, & b_{1t}, & \ldots \\ \vdots & \vdots & \vdots & \vdots & \\ b_{r1}, & b_{r2}, & \ldots, & b_{rt}, & \ldots \end{bmatrix}$$

associated with any consequence more manageable, and therefore more "thinkable", various reduction procedures are employed. For example, in cost-benefit analyses it is customary, as we indicated in the previous subsection, to combine benefits of various types into a composite benefit. In this more complicated example, one would then proceed this way <u>for each period</u>. Thus we can let $B^o_{.1}$ be the combined benefit of the column of benefits in the first period, $B^o_{.2}$ for the second period, and so on*. This reduction leads to a simpler summary of the consequence, namely

$$\begin{bmatrix} c_1 \ , & c_2 \ , & \ldots, & c_t \ , & \ldots \\ b^o_{.1}, & b^o_{.2}, & \ldots, & b^o_{.t}, & \ldots \end{bmatrix} \quad .$$

In this display we merely have time streams of costs and of composite benefits. The usual procedure at this point

---

*In the notation $B^o_{.t}$ think of the zero as an aggregating or collapsing indicator and the subscript as indicating that in this case the collapsing is done over types of benefits for time period t. Shortly we'll meet the notation $B^o_{i.}$.

is to coalesce each of the time streams into so-called <u>present value</u>. Since costs in the future are less painful than costs in the present (e.g., we could put money today in the bank and get back more in the future), discounting is usually employed and the present value of the cost-stream is

$$\Sigma_t \frac{c_t}{(1 + \lambda)^t}$$

where λ is the effective period-to-period interest rate. Many government agencies use a λ value of .10 and argue that it has something to do with the time-value of capital funds in the private sector. In a similar manner one can also discount the composite-benefit stream.

Another alternative would be for each project to collapse, for each i, the benefit stream of type i: B_{i1}, ..., B_{it}, ..., to get a present value $B^o_{i.}$ of the ith stream. One would then compare the present value of the cost stream with the r values $(b^o_{1.},\ldots,b^o_{i.},\ldots,b^o_{r.})$. And this reduced form now presents us with a problem of the type discussed in Section 1.5.3.

Are these reduction procedures reasonable? Are there alternatives? If discounting is used how should one think of a reasonable discount level to use? Should the discount factors be constant from period to period? What about uncertainties? Should one discount expected values? Is it reasonable to raise the discount rate to account for uncertainties? Should one discount streams of physical quantities (as is commonly done with money values) or should one first transform these physical measurements into psychological values or utilities before the discounting takes place?

We are not going to answer all of these questions because many answers will be of the type, "It depends, ...," but we will in this book provide a conceptual framework that can be applied equally well to value problems of temporal tradeoffs.

1.5.5 The Value of A Life

There are a number of problems, in surprisingly different contexts, where descriptions of consequences may involve dire happenings, like human deaths and suffering. It's not very comfortable thinking about such problems, and therefore we often act in such matters without sufficient reflection. Who likes to play God? Well, if we all abdicate our responsibilities to think hard about such matters as "the value of a life" and allow decisions to be made by happenstance, then we may inadvertently contribute to the lot of human anguish.

The problems we cite at the beginning of this chapter (e.g., electrical power generation vs. pollution, location of an airport, treatment of heroin addiction, medical diagnostics) all involve in one way or another considerations that involve life- and-death matters.

There are other classes of problems, more complicated from an ethical point of view, that we have decided to ignore; problems such as abortion, population control, euthanasia, genetic engineering. Not that these problems cannot be thought about in the framework we will develop, but we haven't sorted out our thoughts on these topics clearly enough to subject them to the perusal of an audience in today's highly, emotionally charged arena.

Let's simplify our discussion a bit and consider just the case where a decision maker must choose among several life-saving programs. For a cost of so-and-so he can achieve a certain probability distribution of saved lives. In a public setting it's important to think of alternative uses of funds. If we save lives by spending more money to keep people alive on kidney machines, are the alternatives "more milk for the malnourished" or "better dental care for the needy" or "more money for military research?"

We have a cherished symbolism about the Sanctity of a Single Life. But perhaps our morality has gone astray when it comes to numbers. Emotionally we get choked up about a little girl getting killed--especially if we can see her picture--but we do not feel emotionally touched by thousands of people being wiped out by a tidal wave or an earthquake. Somehow we need to learn that our grief should rise monotonically with the magnitude of a catastrophe. Numbers are important.

Charles Fried [1970] has pointed out that as a society, we are romantic sentimentalists. We're willing to spend a lot more money on rescue than on prevention, more to save trapped miners and marooned astronauts than to save many more statistical anonymous lives. If we conjure up a face, we can emphathize with the victim.

If a public official acts to save lives, he gets more kudos if he can point to ten specific identifiable persons who have been saved, than if he can prove conclusively that one thousand lives have been saved but he can't identify who these people are.

The problem of identifiability and partial identifiability comes up all the time in circumstances less dramatic than in matters of life-or-death. In counting up the benefits of Program A it's really helpful to know that John Smith and Mary Doe have been helped. If Program B benefits many more people than Program A, but if these people can't be brought together or identified easily, then, descriptively speaking, Program A will beat out Program B in a competition for survival. As a society we have to learn how to respect such numbers more.

1.5.6 Group Decisions

In many situations, it is not an individual, but rather a group of individuals who collectively have the responsibility for making a choice among alternatives. Such a characterization is referred to as a group decision problem. With each group decision, there is the crucial metadecision of selecting a process-oriented strategy by which the group decision is to be made. A general strategy for this may be first to obtain each individual's preferences for the alternatives, and then to combine these in some reasonable manner to achieve the group's preferences. With this framework, the essence of the group metadecision is how to integrate the individual's preferences.

It should be clear by now that we authors feel that often the methodology and procedures discussed in this book would be helpful to the separate individuals in specifying their preferences, whether ordinal or cardinal, for the alternatives. We also believe that in some cases, the procedures of multiattribute preference theory discussed here might be useful in providing a process by which group decision can be responsibly made. Thus, the implications of the concepts and methodology for use in group decision processes and suggestions for implementation are included in the book.

1.6 ORGANIZATION OF REMAINING CHAPTERS

To help explain the organization of this book we can consider the following abstraction. Assume that associated with each action of the (unitary) decision maker--the individual who really wants to make up his mind--there will be a resulting consequence. We shall partition this class of problem by means of the following double dichotomy:

a) First, is it a problem under certainty or uncertainty? If it falls in the uncertainty category, then we shall assume that to each action there is a well-specified probability distribution over the possible resulting consequences. To the subjectivist--often referred to as a "Bayesian"--this is not any loss of generality for he, if called upon, can always generate (at least conceptually) such a probability distribution. For the objectivist, the existence of a well-specified probability distribution does, admittedly, restrict the generality of our abstraction.

b) Second, is it a single or multiple attribute problem? That is, can the typical consequence be adequately described, in terms of a single attribute (e.g., money, degree of pain, or number of lives saved), or is more than one descriptor needed?

The most general case we will consider--and the case that is of primary interest to us--is when the consequence of an action is both uncertain and multidimensional. Let's label it $\underline{\tilde{x}}$ where the superscript tilde ($\tilde{\cdot}$) represents uncertainty--some might prefer to view the tilde as the sign for random variable--and the underscore ($\underline{\cdot}$) represents a vector in contrast to a scalar.

We shall distinguish four cases as exhibited in Fig. 1.5. When the consequence is both certain and unidimensional the analysis is clear--at least conceptually:

	single attribute	multiple attribute
certainty	x	$\underline{x}$
uncertainty	$\tilde{x}$	$\tilde{\underline{x}}$

Double Dichotomy of Decision Problems

Figure 1.5

one merely chooses that feasible alternative that maximizes the given single objective measure. Of course, in practice, if the alternatives are numerous and constraints are given in terms of a set of mathematical restrictions, one might be hard pressed to find the optimum. The entire arsenal of mathematical programming techniques might have to be employed. But still, the problem is conceptually straightforward, and, as such, we will not deal with that case in this book.

Chapter 3--we'll come back to Chapter 2 shortly--deals with the case of certainty when there is more than one descriptor. This can be thought of as complex value analysis under certainty. Much of the flavor of this book comes through in this analysis. Basically the problem boils down to the following: how can one systematically think about ranking a set of consequences when each consequence is described in terms of performance values on many attributes. The problem of subjective tradeoffs must be met in earnest in these discussions. We don't suggest a magic objective formula to grapple with these tradeoffs but rather we suggest several concrete procedures that one might employ to help probe and articulate one's basic values or tastes.

In Chapter 4 we generalize to the uncertainty case but at the same time specialize to the case where there is only one descriptor. The uncertain consequence associated with an action can now be labeled by $\tilde{x}$ rather than by $\underset{\sim}{x}$. In this chapter we present a general review of what is now known as the theory of risky choice, or cardinal utility theory, or von Neumann-Morgenstern utility theory. An elementary version of this material can be found in Chapter 4 of Raiffa [1968], but the discussion in the present book is more analytic and surveys some of the considerable progress that has been made in the last few years. In order to describe succinctly the problem examined in this chapter, suppose that to each action there is a probability distribution of an as-yet-unknown monetary reward. You, as decision maker, are called upon to rank order such probability distributions and, as such, you must implicitly characterize your attitudes towards gambling situations. What kind of a risk taker are you?

In Chapter 5 and 6 the consequences are both uncertain and multidimensional, and the techniques developed in the two preceding chapters for the two special cases (certainty-complex and uncertainty-simple) come into play here but collectively they do not quite satisfy our needs. Additional techniques are developed to handle the difficulties introduced by the interactions between uncertainty and multidimensionality. We have, a bit arbitrarily, divided the subject matter into two chapters because of the overall length. Chapter 5 deals primarily with utility functions over two attributes whereas Chapter 6 copes with more complicated multiple attribute structures.

Now let us back up and briefly describe the contents of Chapter 2. This chapter starts by establishing some basic vocabulary: goals, objectives, attributes, evaluators, measures of performance, subjective scales, and so on. Some of these terms will be part of our technical vocabulary and we must establish a common understanding of their meanings--at least as we shall use the terms. We then turn our attention to perhaps the most creative part of our subject matter but, unfortunately, a part that is difficult to describe systematically, namely: how should one generate the objectives and attributes in concrete problems. After all, these objectives are not in practice delivered to the decision maker on the proverbial platter but he or she must, literally, create them. The best way we know how to deal with this phase of our subject matter is to describe some concrete cases and illustrate how one might have thought about pertinent objectives. We would like the set of objectives to be complete but yet we do not want to encumber ourselves with a lot of trivial considerations that do not mount up to anything significant as far as the making of decisions is concerned. You will see that the generation of a suitable set of attributes is not unique and as such one must understand what considerations should be involved in a choice between alternatives. But one can't decide on what constitutes a desirable set of objectives without understanding what could be done with these objectives after they have been thought up. This involves some understanding of how various attributes can be evaluated , of how redundancies can be handled, of how parts of the problem can be isolated from other parts, of how values get intertwined with probabilistic assessments, of how inconsistencies of measurement inputs can be detected, of how such inconsistencies may be rectified, and of how calculations can be made in order to select a wise course of action. In short, when choosing a set of attributes to consider, one must worry about what comes next, and therefore one must have some appreciation of the contents of Chapter 3 to 6. But yet in Chapter 3 to 6 we assume for the most part that a set of attributes has already been determined. It is not until Chapters 7 and 8 that the separate parts get integrated.

In Chapter 7 we look at a series of concrete problems and discuss how one might generate suitable sets of attributes describing the possible consequences, but now, unlike Chapter 2, we can also discuss whether these attribute sets can be manipulated in a tractable fashion. In particular we shall consider such problems as

(1) Should New York City lower the legal limit on the sulfur content of fuel oils burned within the city,

(2) How should budget allocations be made among diverse activities of an educational program,

(3) Which response strategies available to an urban fire department result in the best overall deployment of service,

(4) How can one evaluate the quality of service of a computer system,

(5) Can the process of siting and licensing of nuclear power facilities be significantly improved,

(6) What is the best procedure for a team of medical doctors to treat a patient who has a serious medical problem,

(7) What policies should management adopt to 'best' achieve the objectives of a corporation.

The emphasis of Chapter 7 is to indicate how the ideas of previous chapters have been used on various aspects of some complex problems and to suggest the relevance of these same concepts and techniques to other strategic issues.

It is in Chapter 8 that we discuss a case from start to finish. The problem concerns selecting a strategy for developing the major airport facilities of Mexico City over the period to the year 2000. This study serves two purposes. First, it further illustrates the applicability of many of the techniques and procedures developed in earlier chapters to a very important "typical" problem -

typical of those one-of-a-kind strategic problems with so many atypical features. Secondly, the Mexico City Airport study indicates the integration of and inter-connections among different aspects of the analyses: defining and structuring the problem, modelling possible impacts of variant alternatives, specifying the value judgments of the Mexican Ministry of Public Works, etc. The larger setting within which the analysis occurred is also discussed.

Chapter 9 and 10 contain two very important methodological problems which can be naturally cast and studied within a multiattribute framework. These are respectively 'preferences over time' and 'group preferences and the social welfare problem', both of which were outlined in Section 1.5. The analytical results of Chapters 3 through 6 are relevant to either situation if the appropriate assumptions are satisfied. Thus, concerning temporal preferences, we may obtain a utility function for consequences of the form $(x_1, x_2, \ldots)$, where x_i indicates the consequence in time period i. In the group interpretation, it would be desirable to measure overall group preferences for consequences $(u_1, u_2, \ldots, u_n)$, where u_j indicates the preference of group number $j, j=1,2,\ldots,n$. In both chapters 9 and 10, we present brief surveys of previous work on the respective problems, an interpretation of multiattribute utility in the contexts of concern, and a discussion of procedures for implementing the multiattribute results within these contexts.

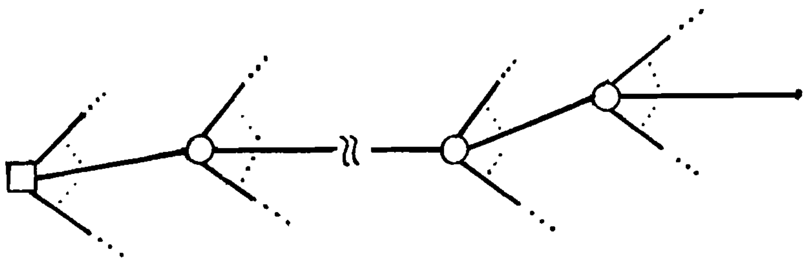
CHAPTER 2

THE STRUCTURING OF OBJECTIVES

Let us start with the decision paradigm mentioned in the previous chapter, where we abstract a decision problem into the form of a decision tree as shown in Fig. 2.1. At each tip of the tree there is some consequence, C, that characterizes the full cognitive impact of that position point in time and space. The decision maker is called upon not only to rank the consequences at the tips of the tree but also to evaluate the strengths of his preferences and his attitudes towards risk in terms of a utility function defined on these consequences. This is not an easy task. As a step in this evaluation procedure we imagine that the decision maker first describes each consequence C in terms of an ordered set of, say r, numerical (or some simple generalization thereof) evaluators or descriptors. These r evaluators are designed presumably to make the abstract consequence C a bit more concrete. Instead of making a paired comparison between C and C' in the abstract, it may be easier to think of the comparison between

$$X_1(C),\ldots,X_i(C),\ldots,X_r(C) \quad \text{and} \quad X_1(C'),\ldots,X_i(C'),\ldots,X_r(C') \quad ,$$

where $X_i(C)$ refers, for example, to the "level" (to be defined



Schematic Decision Tree

Figure 2.1

more accurately, later) of the consequence C as evaluated by X_i. If this is a worthwhile step to take, these r evaluators must in some sense be an adequate representation of the consequences they purport to describe.

2.1 OBJECTIVES AND ATTRIBUTES

There are no universal definitions of the terms objective, goal, attribute, measure of effectiveness, standard, etc., so we will begin this section by indicating in an informal manner how these terms will be used in this book. Our approach will be to illustrate our terminology in problems similar to the motivating examples in Chapter 1.

2.1.1 Some Illustrations

A. Air Pollution: Because of excessive levels of pollution in a given city, the authorities might be interested in-- or have an area of concern in-- "the threatened well-being of the residents of the city." A broad overall objective corresponding to this area of concern is to "improve the well-being of the residents." Such a broad objective provides little if any insight into which of a number of alternative programs may be best or even worthwhile to pursue. It does, however, provide a useful starting point for specifying detailed objectives in more operational terms.

For example, two more detailed objectives, or lower-level objectives as we will refer to them in this area of concern, might be "reduce the emissions of pollutants from sources within the city" and "improve the citizen's attitude toward their air quality." The first of these subobjectives might further be broken into three lower-level objectives: "reduce sulfur dioxide emissions," "reduce emission of nitrogen oxides," and "reduce the particulate emissions." For each of these lowest level objectives we might want to associate an attribute which will be used to indicate the degree to which alternative policies meet this objective.

Achievement in terms of reducing sulfur dioxide emissions might be indicated in terms of the attribute, "tons of sulfur dioxide emitted per year." This attribute is measured with a scalar quantity, and thus is referred to as a scalar attribute. Similarly, scalar attributes for our other two lower-level emission objectives might be in "tons of nitrogen oxides emitted per year" and "tons of particulate emitted per year." Together these three scalar quantities could be represented as a vector measuring the degree to which the next level objective, "reduce the emissions of pollutants from sources within the city" is met. Thus, the composite of the three scalar attributes is referred to as a vector attribute.

The objective "improve the citizen's attitude toward their air quality" may be measured by an attribute "percent of residents alarmed by the city's air pollution." In each of these cases, the attribute provides a scale for measuring the degree to which its respective objective is met.

B. The Postal Service: Suppose the overall objective of the Postal Service is "to provide efficient, dependable service to the users of the system and to the government." There are many possibilities for subobjectives, or lower-level objectives. These include "minimize total transit times for parcels and letters," "maximize the percentage of mail delivered (i.e., avoid losses)," "minimize the total cost of handling the mail," and "provide services to the government." The cost objective may be broken into "minimize direct mailing costs to users," and "minimize the cost to government," the government being ultimately responsible for all postal service expenses.

For the first objective--minimize total transit time for parcels and letters--a rather obvious attribute is "the time in days from sender to receiver." However, it may be more appropriate to decompose "mail" into categories where the k^{th} category refers to a particular destination at a particular time of year. Let us denote by X_k the attribute "the time in days that a randomly selected letter of category k is in transit from sender to receiver." For a given alternative this attribute will have a frequency distribution. In some examples we might want to summarize this distribution in terms of a single summary number (e.g., the mean, or an adjusted mean, or some other more sophisticated index which reflects the nature of the tails of the distribution). If we follow this tack, the k^{th} category will be summarized by some single number x_k, and if the categories k range from 1 to K, then the objective "minimize transit time for parcels and letters" will be evaluated by the vector attribute $(x_1,\ldots,x_k,\ldots,x_K)$.

The problem of finding an attribute, most likely a vector attribute, to indicate the degree to which alternatives meet the objective "provide service to government" may be very difficult. Aspects of this include facilitating communication among all citizens, informing citizens of their government's activities, and providing employment for thousands of people. Even if we do effectively spell out a set of lower-level objectives in this case, it will be difficult to identify useful attributes for each. Such problems are addressed throughout this chapter.

2.1.2 Terminology: Objectives, Attributes, Goals, etc.

It is very likely that objectives, as we have chosen to use the term, will conflict with each other in the sense that the improved achievement in terms of one objective can only be accomplished at the expense of achievement of another objective. For example, must businesses and public services have objectives like "minimize cost" and "optimize the quality of service." Since better service can often only be attained for a price, these objectives conflict. It may be possible in some cases to simultaneously increase achievement on both objectives relative to the current situation. That is, a better strategy--in terms of all objectives--may exist. However, at some point one will be faced with the proposition that further achievement on one objective can only be accomplished at the expense of achievement on the other.

In general, although not necessarily always, an objective indicates the 'direction' in which one should strive to do better. Recall the Postal Service <u>objective</u> "minimize total transit time for a given category of mail," which was measured in terms of the <u>attribute</u> 'days'. Since it is unlikely that transit times would be reduced to zero, one could always strive to do better. Let us contrast this objective and its associated attribute with a so-called <u>goal</u>. For this problem, a goal may be "deliver at least ninety percent of the parcels and letters within two days." A goal is different from an objective in that it is either achieved or not.

Goals are useful for clearly identifying a level of achievement to strive toward. President Kennedy's stated goal in 1961 was to reach the moon by 1970. This goal would either be achieved or not. It is much easier to inspire people, including oneself, to climb a mountain when it has a summit than when there is none. However, for our subject matter we feel that objectives are more relevant than goals for evaluating alternatives in strategic decision problems. This is not to say that the use of goals is not a useful tactical device for implementing an action program. In the sequel we shall confine our language to objectives and attributes and minimize the usage of the term "goal."

2.2 GENERATION OF OBJECTIVES AND ATTRIBUTES

In practice there is considerable interplay in the creative process of generating objectives and selecting attributes for these objectives. Before pursuing the inter-relationships in depth, it is necessary to first consider objectives and attributes separately.

2.2.1 Some Techniques for Generating Objectives

Let us suggest some guidelines that may be helpful in generating objectives for a specific problem. As a starting point, assume one objective has been specified, such as the overall objective "improve the well-being of the residents" in the air-pollution problem. Clearly, in this case it would be desirable to be more specific about such a broad objective. Answering the question, "What is meant by "well-being of the residents"? would better specify the objectives. For instance, one might include health and economic conditions as part of well-being. Each of these may in turn be broken down further.

MacCrimmon [1969] suggests the following approaches for generating objectives: (a) examination of the relevant literature (b) analytical study and (c) casual empiricism. "Examination of the literature" should be clear. If others have faced problems similar to yours, they perhaps have documented some objectives which are relevant to your problem. "Analytical study" suggests that by building a model of the system under consideration and identifying relevant input and output variables, suitable objectives will become obvious. This might be useful for picking up objectives which were originally omitted either by oversight or intention. Some objectives originally regarded as insignificant might seem important after considering the results of various studies with the model. The third suggestion, "casual empiricism," includes observing people to see how in fact they are presently making decisions relevant to the problem. How do they rationalize their actions? What do they talk about? For instance, in selecting objectives for choosing among alternate housing developments, one might observe how people choose among currently available options. This may provide some indication of relevant objectives.

Surveys may be useful in selecting objectives for public decision making. Individuals, who will be affected by a certain decision, can be asked what objectives should be included in a study. Such a process might generate many "low-level" objectives. In such a case, we would want to utilize these lower-level objectives to specify broader objectives. For instance, if one objective were "to not feel nauseated by the smog," this might be translated into a broader objective by answering the question "Why is it important that one not feel nauseated?" Feeling nauseated indicated some adversity effecting people's health, so a broader objective might be to "improve the health of the specified popluation."

In many instances, it may be useful to have a group of knowledgeable experts identify the objectives in a problem area. The board of directors in business firms often plays this role of setting objectives. In recent years, especially in technological and scientific problem areas, both government and private industry have begun to use the "panel of experts," a group of people with expertise in the area of interest, to generate the objectives.

2.2.2 Illustrations*

A. Scientific Objectives of NASA

An ingenious approach was utilized in specifying objectives for the National Aeronautics and Space Administration to use in evaluating the scientific merit of alternative plans for space exploration. The scientific objectives were first grouped into five main sub-areas: (1) Earth and its environment; (2) Extraterrestrial Life; (3) The Solar System; (4) The Universe; and (5) Space as a Laboratory. Then lists of what were called action phrases, target features, and target subjects were developed. The idea can best be explained by referring to Table 2.1 which is reprinted from Dole, et al. [1968a]. One would try each of the combinations of an

---

*Precisely speaking, the two studies briefly described in this section do not specify objectives as we have chosen to define them. In our terms they identified areas of concern from which one could generate objectives. For this section, we have retained the terminology of the cited works.

Table 2.1

Generating Scientific Objectives for a Space Program

Action Phrase	Target Feature	Target Subject
Characteristic circulation patterns in	the photosphere of	the sun
Measure tidal deformations of	the surface of	the moon
Establish the structure of	the interior of	Jupiter
Measure relativistic time dilations in	______	the space environment

(Action Phrase, Target Feature, Target Subject) and then ask, "Is this one of the scientific objectives of a space program?" If the grouped words were an objective, it was included in the list. If the words were meaningless, clearly they were omitted. Thus, for instance, "Establish the structure of the interior of the sun" was an objective, whereas "Measure tidal deformations of the space environment" was not an objective. This procedure generated one thousand and thirty lower-level objectives. The complete results are in Dole, et al. [1968b].

B. The Louisville Study*

In Louisville, Kentucky, a group of citizens representing diverse segments of the community, who worked closely

*For details, see Schimpeler et al. [1969].

with the mayor, identified areas of concern and selected objectives for public policy. This Mayor's Citizens Advisory Committee identified ten major areas of concern which were further specified into thirty-five lower-level aims representing interests of the city of Louisville. These "goals" are indicated in Table 2.2.

Table 2.2

Areas of Concern to Residents of Louisville, Kentucky

MAJOR OBJECTIVES*	LOWER-LEVEL OBJECTIVES
A. Public Safety Program Development	1. Insure safe public facilities 2. Provide for adequate public safety regulations and their enforcement 3. Provide for the removal of contaminants
B. Public Utility and Transportation Development	1. Minimize maintenance costs of public utilities 2. Insure maximum effectiveness of public utilities by design and locational consideration 3. Develop a balanced, effective and integrated transportation system which provides for the accessibility requirements of each land use.
C. Economic Development Programs	1. Develop public improvement programs within available financial resources 2. Maintain highest equitable property values 3. Insure effective utilization of mineral, vegetation, air and water resources 4. Establish strong economic base through commerce that will bring money into the community

*Or more precisely: Areas of Concern

	5. Establish trade development that provides maximum convenience to consumers 6. Insure the optimal utilization of all land 7. Achieve increased disposable income for all people
D. Cultural Development	1. Preserve historic sites and areas of natural beauty 2. Promote adequate public libraries, museums and cultural activities 3. Protect meaningful local tradition and encourage civic pride
E. Health Program Development	Establish the mechanism for adequate preventive and remedial health programs and facilities
F. Education Program Development	Develop education facilities and opportunities for citizens at every level
G. Welfare Program Development	1. Eliminate injustice based on discrimination 2. Develop needed public welfare programs 3. Encourage development of religious opportunities 4. Develop an aesthetically pleasing environment
H. Recreation Program Development	1. Establish open space programs 2. Provide adequate recreational facilities utilizing parks, rivers and lakes
J. Political Framework	1. Improve the framework (channels, systematic use) for citizens participation in government functions 2. Establish equitable taxation policies (bases, mixes, rates) 3. Achieve efficient governmental administration representative of all citizens 4. Develop adequate government staffs and personnel programs (high job standards, reasonable salary ranges, effective delegation of authority) 5. Establish sound governmental fiscal programs 6. Develop an effective, long-range, metropolitan-wide planning process 7. Establish effective control mechanisms

K. Housing Development	1. Encourage rehabilitation and conservation neighborhood programs 2. Provide adequate low-cost housing 3. Develop neighborhood units 4. Promote a wide variety of housing as required within the community

Table 2.2 provides an excellent point to further articulate objectives. Consider area of concern C6, "Insure the optimal utilization of all land." This identifies land utilization as important, and almost by definition, everyone would want optimal utilization. However, this likely means very different things to different people. What exactly is meant by optimal utilization? This difficult problem should perhaps be addressed by the Mayor's Citizens Advisory Committee or another such group with the assistance of the City's property tax authorities. The identification of such open problems is one of the contributions made by a formal specification of objectives.

Once a "first-cut" list of objectives is published, it also can be used by all interested parties and individuals as a base for constructive criticism and improvement. This type of iteration should help generate more objectives for a given problem but equally important, it has the beneficial effect of getting concerned individuals to think actively about a complex problem of relevance to themselves.

2.2.3 Specification of Attributes

To describe completely the consequences of any of the possible courses of action in a complex decision problem would require volumes. In the air quality example a consequence would explain who got sick when; how badly they felt; when they recovered; the economic impact on each individual due to pollution; and all related psychological, physical, and economic impacts. This would certainly be complete. However, information in this form is not useful for decision making purposes. What is needed are summary statistics to reduce this morass to a useful, manageable form.

To be useful to the decision maker, an attribute should be both comprehensive and measurable. An attribute is <u>comprehensive</u> if, by knowing the level of an attribute in a particular situation, the decision maker has a clear understanding of the degree to which the associated objective is achieved. An attribute is <u>measurable</u> if for each alternative it is reasonable both (a) to obtain a probability distribution over the possible levels of the attribute--or in extreme cases to assign a point value--and (b) to assess the decision maker's preferences for different possible levels of the attribute--for example, in terms of a utility function or in some circumstances a rank ordering*. Furthermore we would like both these

---

*We are implicitly assuming that all other attributes are held fixed at some specified levels. It could happen that preferences for different levels of an attribute might shift when the other attribute levels are changes. This is discussed fully in later chapters.

tasks to be accomplishable without taking an inordinate amount of time, cost, or effort. So, to some extent, comprehensiveness refers to the appropriateness of the attribute on theoretical grounds--i.e., does it give us the information we would like to have, regardless of whether we can get it--and measurability refers to the practical considerations--i.e., can we get the necessary assessments?

A comprehensive attribute should be relevant to the particular alternative courses of action under consideration and not subject to other extraneous considerations. For instance, suppose one objective of a proposed law to require the wearing of seat belts by all travelers in all vehicles at all times is to reduce vehicle casualties. In this case, the attribute "number of casualties in automobiles per year" would not be comprehensive, because it is difficult to differentiate the effects on the level of this attribute due to wearing seat belts from the effects due to other factors, such as the number of accidents.

As another example, suppose the overall objective of a government "stop smoking campaign" is "to improve the health of the nation." Then the attribute "number of deaths due to smoking" is not comprehensive in that it offers no information about those who are sick or disabled by the pollution. Whenever one considers attributes involving numbers of sick, injured, etc., the problem of precision must be addressed. For example, in a transportation problem where one objective is to decrease injuries, the attribute "number of injuries" is not precise because the definition of injury is not clear. This is aside from the question of whether all injuries, using any specified definition, should be considered as equally important. Because of the imprecision, different people might assign different levels to the "number of injuries" even though they had access to the same information.

In many cases, the choice of an attribute will not be difficult given that the objective is clear. If a businessman's objective is to maximize profits, then profits measured in dollars would be a logical choice of an attribute. Knowing the profits for a particular endeavor would indicate the degree to which the objective "maximize profits" is achieved. If a freight shipping firm wanted to deliver all shipments on time, a reasonable attribute might be the delay time in the arrival of the shipment. In a medical context, a major objective might be to keep a patient alive in which case the attribute "probability of death" would be appropriate. One could assign a number to the delay time of shipments, to profits, and to the probability of death; whereas the respective objectives, per se, cannot be quantified.

2.2.4 Subjective Attribute Scales

Many of the attributes one intuitively thinks of using are objective (as opposed to subjective) in nature.* By this,

* use footnote on following page

we mean there already exists a commonly understood scale for that attribute and its levels are objectively measurable. However, there are objectives for which no objective index exists, and in such cases, a subjective index must be constructed. The scale for the subjective index is specific to the problem at hand.

Consider the businessman who wishes to "maximize profits" and "increase prestige." As mentioned previously, an obvious attribute for the first objective is the objective index: "profits, measured in dollars." However, since there is no objective scale for prestige, one is obliged to establish a subjective index for this objective. A first step could establish a ten point ordered scale going from, say, "desultory low" to the "pinnacle of world-renowned esteem." One would then subjectively assign consequences--ranked from worst to best--to several identification points along this scale. In some circumstances one might have to assess probability distributions and establish a cardinal utility measure over this scale. The literature in psychometrics is replete with examples which establish such scales but the motivation for that literature is quite distinct from ours. Nevertheless, in this book we can, and do, build up

---

*Note that we use the terms objective and subjective to describe two types of attributes, both of which are used to indicate the degree to which objectives are met. However, we shall not facetiously define a "subjective," or worry about achieving it.

from that methodological base.

We cite here only one example of a subjectively assessed scale. Huber, Sahrey, and Ford [1969] asked a number of experienced, professional personnel of a large hospital to subjectively evaluate twelve hypothetical hospital wards on a scale from zero to one hundred. They asserted that their results strongly indicate that professionals can develop and reliably use subjective evaluation models. In our work if we were to use such a scale in conjunction with other scales in a multiattribute problem, we would be obliged to structure this scale _internally_ in such a manner that it would mesh _externally_ with other scales. This leads to the problem of conjoint measurement which we will address in Chapter 3.

There are, of course, difficulties in using subjectively defined attribute scales and depending on context it may be important to go to creative, fanciful extremes in order to get an objective base. In Section 2.5, we discuss the notion of _proxy_ attributes which alleviates some of these difficulties.

2.3 THE HIERARCHICAL NATURE OF OBJECTIVES

Suppose one has thought hard about the objectives in a given problem and has produced a list which encompasses all the areas of concern. No doubt the different objectives will vary widely in their scope, explicitness and detail, and be inconsistent. The question is, "How can one bring some structure to this list of objectives?" Often these objectives can be structured in a meaningful way by the use of a hierarchy. Almost everyone who has seriously thought about the objectives in a complex problem such as, for example Manheim and Hall [1967], MacCrimmon [1969], Raiffa [1969], Miller [1970], Gearing et al. [1974], the NASA study (Dole et al.[1968a]), the Mayor's Citizens Advisory Committee of Louisville (Schimpeler, et al. [1969]), has come up with some sort of hierarchy of objectives.

2.3.1 Constructing the Hierarchy

From an original list of objectives, how does one construct a hierarchy? And how does one recognize if, in fact, "holes" are present in such a hierarchy? The concepts of specification and means-ends discussed by Manheim and Hall [1967] help here. Specification means subdividing an objective into lower-level objectives of more detail, thus clarifying the intended meaning of the more general objective. These lower-level objectives can also be thought of as the means to the end--the end being the higher-level objective. Thus, by identifying the ends to very precise objectives (the means), we can build the hierarchy up to higher levels.

When one goes up the hierarchy, there is the natural stopping point at the all-inclusive objective. This objective is extremely broad and indicates the reason for being interested in the problem, but it is often too vague for any operational purpose. For example, as seen in Figure 2.2, the overall objective used by Manheim and Hall for evaluating passenger transportation facilities for serving the Northeast Corridor in 1980 was "the good life." However, when we go down a hierarchy, there is no obvious point where one stops specifying the objectives. One's judgment must be used to decide where to stop the formalization by considering the advantages and disadvantages of further specification. If this were not done and the hierarchy were carried to absurd lengths, one would end up with an astronomical set of objectives. In planning passenger transportation in the Northeast Corridor, one could carry things to the point where each affected individual (maybe fifty million of them) had a sub-hierarchy representing only themselves in the overall hierarchy of system objectives. Of course, no one would advocate such an approach, but the point of all this is that one must be pragmatic about the level of detail or specification one is prepared to assess.

2.3.2 How Far to Formalize?

How far should one extend the objectives hierarchy? The answer depends a great deal on what will be done next with the hierarchy. Are we going to identify attributes for each of the objectives? This is related to the qualitative versus quantitative growth of the hierarchy soon to be discussed and to the notion of direct preference measurements. Are we willing to use subjective

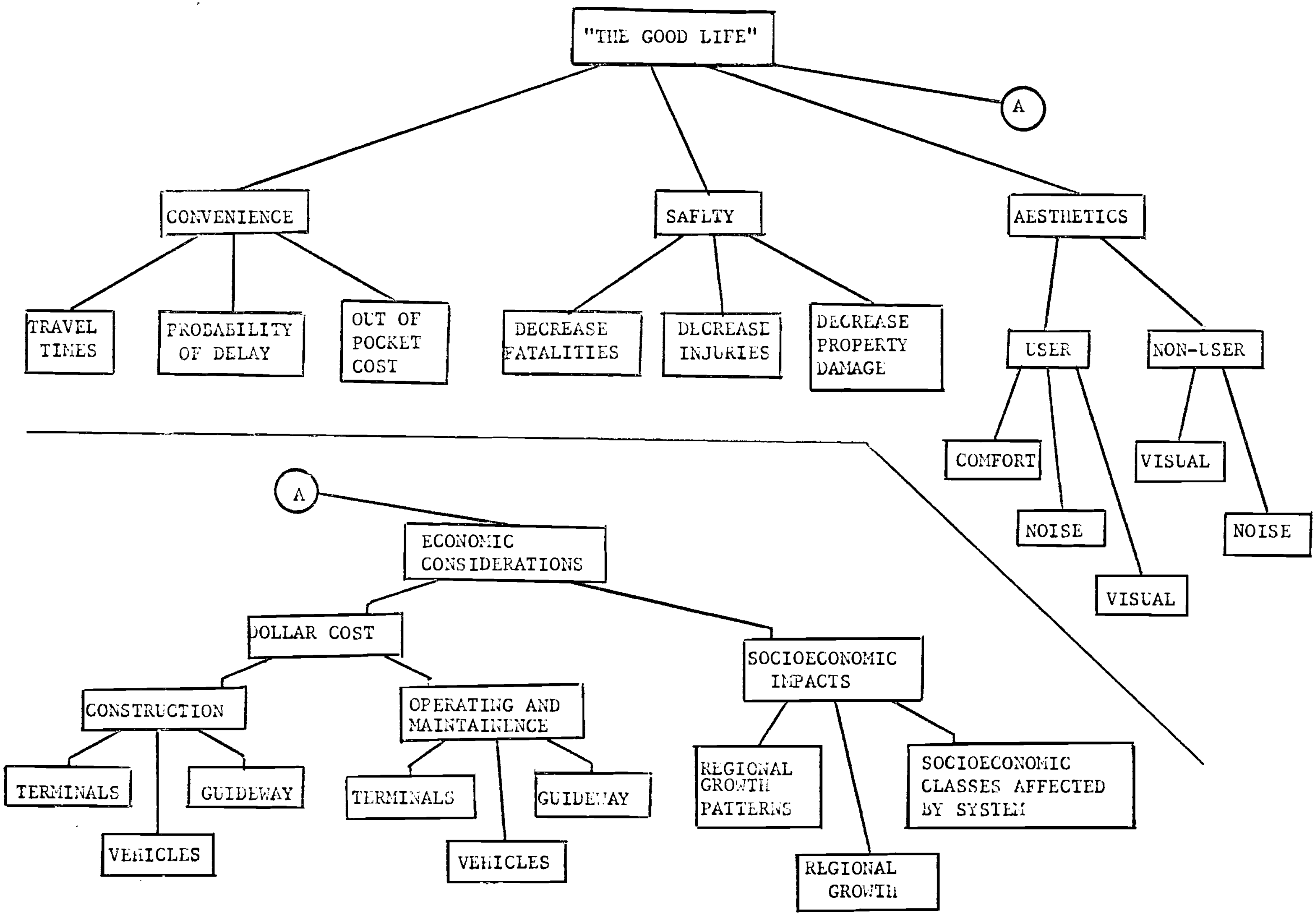


Figure 2.2. A heirarchy of objectives for evaluating passenger transportation facilities for the northeast corridor in 1980.

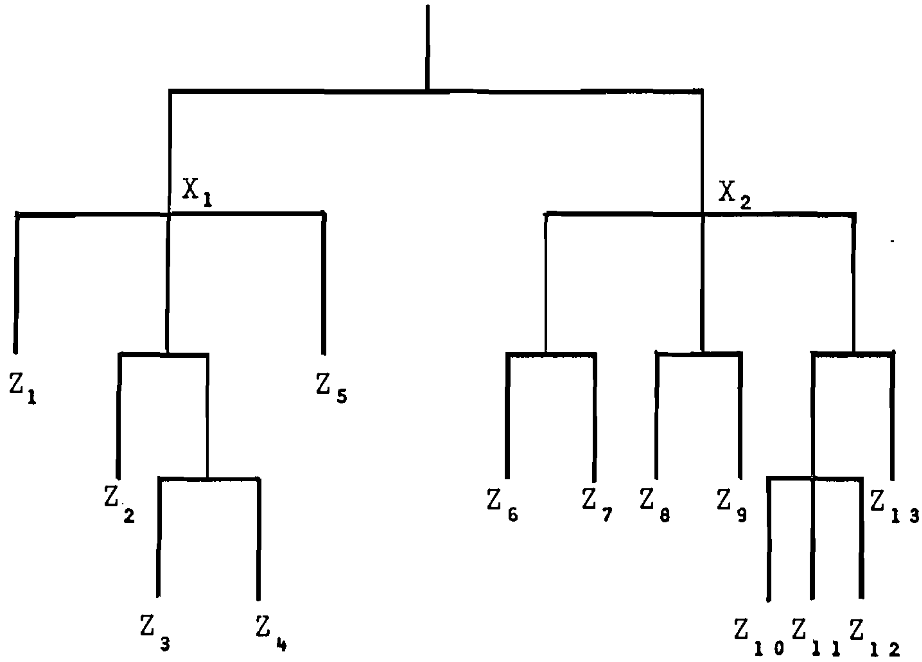
indices of effectiveness or do we prefer objective ones? This question depends partially on who the decision maker is and on who is performing the analysis and for what purpose.

When dividing an objective into subobjectives, at any level, care must be taken to insure that all facets of the higher objective are accounted for in one of the subobjectives. However, one must guard against a proliferation of the hierarchy in the lateral direction as well as the vertical. For instance, if we ended up with hundreds of lower-level objectives, which are specifiers of a higher-level objective, some of these might be so insignificant relative to others that they could be excluded from the formal analysis without leading the decision maker astray. Still care must be exerted in discarding objectives lest the remainder become seriously non-comprehensive.

Ellis [1970] introduces a "test of importance" to deal with this problem. Before any objective is included in the hierarchy, the decision maker is asked whether he feels the best course of action could be altered if that objective were excluded. An affirmative response would obviously imply that the objective should be included. A negative response would be taken as sufficient reason for exclusion. Naturally, one must avoid excluding a large set of attributes each of which fails the "test of importance" but which collectively is important. As the analysis proceeds and the decision maker gains further insight into his problem it is worthwhile to repeat the test of importance with the excluded objectives. If the decision maker has a change in mind, then some objectives and their associated attributes must be added to the problem and certain parts of the analysis repeated.

2.3.3 Qualitative Proliferation of the Objectives Hierarchy

In this book our ultimate aim in a specific applied context is not merely to generate a good objectives hierarchy for the problem. We are concerned with using this hierarchy as a step along the way in a decision analytical framework. In the next chapter, we shall begin to talk about preference tradeoffs between attributes and quantifying our preferences. Numbers will loom large in the ensuing analysis. Let us consider for example, the abstracted schematic version of the hierarchy shown in Fig. 2.3. In this hierarchy there are 13 lower-level objectives and let their associated attributes be $Z_1, Z_2, \ldots, Z_{13}$. Thus a given consequence of the decision problem could be described by a 13-tuple $(z_1, z_2, \ldots, z_{13})$. One might choose to formalize a utility function in this 13-dimensional space and thus assign values such as $u(z_1, z_2, \ldots, z_{13})$. But this is not necessary in order to proceed. As an alternative, one might quantify preferences at a much higher level of aggregation. For example, it may be better to work directly with the attributes X_1 and X_2 where X_1 is subjectively assessed composite of Z_1 to Z_5 and X_2 of Z_6 to Z_{13} (see Fig. 2.3). Instead of engaging the utility analysis at the level of $(z_1, z_2, \ldots, z_{13})$,



An Abstract Objectives Hierarchy

Figure 2.3

utility assignments for entities of the form (x_1, x_2) could be used. Of course, in this case for a given consequence C, the values of $X_1(C) = x_1$ and $X_2(C) = x_2$ might have to be subjectively assessed.

We can use the hierarchy in a manner that is convenient to ourselves and embark upon a further analysis by introducing utility assignments at various levels of the hierarchy. However, if we were going to quantify our preferences at the X_1, X_2 level, why proliferate the hierarchy down to the Z_1 to Z_{13} level? The answer simply is that the qualitative structuring of the objectives associated with X_1 and X_2 might help us to think more clearly about X_1 and X_2. In other words, the vertical depth of the proliferation of the hierarchy does not necessarily force us to quantify our preferences down to this level of detail. The hierarchy after a given level may merely serve as a qualitative check list for things to consider.

Extending the hierarchy for qualitative purposes can be illustrated using one of the major objectives of the air pollution problem cited in Chapter 7. For the objective "achieve the best political solution," it was decided to use a subjective index to indicate the degree of achievement. However, to stimulate thinking about the assessment of this subjective index, it may be desirable to specify the major objective further. For instance, one could identify some subobjectives of this major objective such as to "improve relations with the City Council," to "gain the support of certain political groups," to "maintain good terms with the landlords" who must buy fuel to heat their buildings, and to "transmit the notion that the City Administration is concerned about the welfare of its residents and the environment of the area." If we were to assess utilities directly for the major objective, preferences and likelihoods relating directly to the lower-level objectives need not be assessed, and therefore we do not need to identify measures of effectiveness for them. Thus, many of the considerations one might think about in extending an objectives hierarchy for quantitative reasons are not relevant to the case where certain parts of the hierarchy are to be used for qualitative reasons only.

2.3.4 Subjective vs. Objective Measures Revisited

The further one sub-divides an objectives hierarchy, the easier it usually will be to identify attribute scales which can be objectively assessed. When the hierarchy is less expanded, one often has to resort to subjective measures of effectiveness. To illustrate this point, consider another one of the objectives in the air pollution problem of Chapter 7, specifically, "to improve the physical health of the New York City residents." Other than a subjective index, no single measure could be found to indicate the degree to which this objective is met. The difficulty was that both mortality effects and morbidity effects of various kinds were important. Thus, the subobjectives "decrease mortality" and "decrease morbidity" were specified, and objective clinical measures of effectiveness were identified for each.

As a second example consider the design of a new transportation system and let us concentrate on one objective within the hierarchy "maximize passenger comfort." There is no readily available engineering index which can capture the essence of this feature. But if one were to specify comfort in terms of types of comfort (e.g., smoothness or ride, quality of light, maneuverable space, background noise, etc.) one could assign engineering, physical measurements to most of the subobjectives which were introduced to give specificity to the objective "passenger comfort."

2.3.5 Who is the Decision Maker? The Need to Convince Others. Reconciliation of Viewpoints

Let us again suppose that a qualitative objectives hierarchy is as shown in Fig. 2.3. If the decision maker is his own analyst and he does not have to convince anyone of the correctness of his action, it may be convenient for him to assign subjectively assessed values for the X_1 and X_2 attributes and to synthesize in his mind, in a purely informal manner, the consideration of any further detail (such as the further specification of the Z's).

However, when the single decision maker and his analyst work farther from each other, the problem becomes more involved. In this situation, the analyst will presumably need to present his results and recommendations to the decision maker who will then choose an alternative course of action. Thus to better support his work, the analyst will likely need to specify formally the objectives hierarchy in greater detail. He will want to use objective indices rather than subjective indices whenever possible in the interest of "objectivism." He might be forced down to the Z-level rather than remain at the subjective X-level.

If the single decision maker has to convince others of the correctness of his decision as well as to get his own mind straightened out, he may be well advised to go as far as he can with jointly held objective conceptions and this may force him to push the hierarchical analysis down to the objective Z-attributes. But this also cuts another way. The more involved the analysis, the harder it may be to explain it to others and therefore it may be easier to work at the X-level than the Z-level.

Let's now look at the problem from the point of view of an analyst serving multiple clients. He might develop the hierarchy down to the Z-level and obtain objective, engineering measurements for the Z-attributes--measurements that might be accepted by all his clients. Of course, the trouble will come at the next stage of the analysis when the various attribute n-tuples--in this case $(z_1, z_2, \ldots, z_{13})$--have to be rank ordered and scaled (perhaps with utilities) by the various decision makers. But at least the analyst could postpone that consideration while he tries to synthesize the commonly held objective features of the problem.

Suppose now that two or more decision makers project the hierarchy down to the X-level and suppose that they disagree on their overall rankings (or utilities) for consequences. In a reconciliation process it may be desirable to understand <u>why</u> they disagree. One way of proceeding is to decompose the problem further--in this case to further specify the meanings of the X-attributes in terms of the Z-levels. Then, for example one could probe the contending values of Z_2 say, holding the other Z-values fixed. In the sequel we shall introduce various qualitative independence assumptions concerning preferences for multiple attributes and the individuals might conceivably hold qualitatively similar viewpoints that could help probe their differences. Of course, in some circumstances reconciliation could not be achieved by such rational decompositions. Indeed there are lots of cases where reconciliation is only achieved by creative obfuscation. We like to think that the complementary set of circumstances is not a null set. In Chapter 8 we shall discuss these issues further in terms of a concrete case.

2.3.6 Non-Uniqueness of the Objectives Hierarchy

As alluded to earlier, the objectives hierarchy for a particular problem is not unique. It can be varied simply by changing the degree to which the hierarchy is formalized. However, even if the degree of formalization remains unchanged--in the sense that the number of lowest-level objectives is the same--the objectives hierarchy can be significantly varied. Whether one arrangement is better than another is mainly a matter of the particular points the decision maker and the analyst wish to make. Two alternative analyses of employment possibilities, which are reviewed in Section 7.7, provide a fascinating example of such considerations. With different hierarchies, different tradeoffs facing the decision maker can be more easily identified and illustrated.

There is another case where the specific display of the hierarchy may be exploited. This involves cases where some of the lower levels of the hierarchy can be pruned off for consideration of certain alternative courses of action because the further distinction does not matter. As an example, one could imagine that for a heroin problem like the one outlined in the first chapter, one might at times wish to distinguish between the effects on different sexes and age groups. If the lowest level makes the differentiation between effects on males and females, and if for particular alternatives the decision maker is not concerned about these separate effects, the two attributes associated with these objectives can effectively be coalesced into one.

2.3.7 An Illustrative Example: Choice of a Transportation System

To illustrate some of the ideas discussed in this section, reconsider the objectives hierarchy for the Northeast Corridor transportation system given in Fig. 2.2.

As can be seen, the overall objective is to acquire "the good life." Clearly we would not expect to find a single attribute for this overall objective. This was divided into four objectives: "provide maximum convenience," "provide maximum safety," "provide an aesthetically pleasing transportation system," and "minimize system costs and promote regional economic development." For completeness, these four objectives should include all the aspirations of the individuals responsible for the decision which must be made.

The next step involves applying the test of importance to each of these to determine if in fact they need to be included in the formal analysis. Since, in this case, it is fairly obvious that each of these objectives should be kept in the analysis, we won't emphasize the approach at this point.

Let us now take the objective "provide maximum convenience" and attempt to find an attribute which expresses the degree to which this objective is met. Convenience implies that service should be fast, dependable, and economical, at the very least, and no apparent single attribute satisfying the criteria of Section 2.2 includes all of these facets of convenience. Hence, we might choose to subdivide this objective further.

Now that we have made the decision to specify "convenience" to a greater degree, it becomes necessary to consider what might be a suitable set of subobjectives. In this case, one might come up with the following:

1. minimize travel time,
2. minimize departure delays,
3. minimize arrival delays,
4. minimize fare costs,
5. provide easy access to the system.

Since it is desirable to have as few as possible final attributes, we try to generate the minimum number of subobjectives each time this process is carried out. Of course, care must be taken to insure that the list includes all relevant considerations. In this situation, let us consider the possibility of combining some of the five objectives listed above. We might reasonably think that easy access to the system means we can get to the system quickly, and then combine objective 1 and 5 into minimize door-to-door travel time. Whether this would be appropriate in a specific problem would depend on the situation at hand. The point is that one should look for ways of combining objectives in this manner. For argument's sake, let us agree that objectives 1 and 5 are so combined.

Let us agree also that no other combinations are apparent, and so the next step is to apply the test of importance to each of the remaining four subobjectives. Take "minimize door-to-door travel time." We essentially ask "Is this objective important enough to possibly influence the final decision?" It seems entirely reasonable that this objective would be important. Hence, it should be kept in the hierarchy of objectives. The same conclusion can be reached for "minimize fare costs."

The story with "minimize departure delays" and "minimize arrival delays" may be different. For example, one could reason that leaving on schedule and arriving on schedule is not much different from leaving an hour late and arriving on schedule. This is not to say that it is not inconvenient to wait for late departures, but that departure delays might not be particularly serious in themselves. Much of the importance of delayed departure results from its causal effect on total travel time, and total travel time is already included in our analysis. Finally, we consider the question of whether arrival delays--in addition to their impact on total travel time--are important enough to have an influence on the alternative courses of action chosen. A negative response means this attribute has failed to make the test of importance, and it need not be considered explicitly in any ensuing analysis of this problem.

So, as a result of this, we have ended up with two rather than five subobjectives of "convenience":

1. minimize door-to-door travel time,
2. minimize fare costs.

Now we try to find a meaningful attribute for each of these. In this case, the attributes "door-to-door travel time in minutes" and "fare cost in dollars" would be likely candidates. Of course, this brings up the problem of to whom and from where do these times and costs apply.

Unfortunately, even when we have resolved the problems just mentioned, the process isn't complete. The procedure we have just been through must be repeated for the three remaining lowest-level objectives--those concerning safety, aesthetics, and economic and regional impact.

2.4 SETS OF OBJECTIVES AND ATTRIBUTES

The previous two sections concerned building the objectives hierarchy and selecting an attribute for each of the lowest-level objectives. These two topics were considered in isolation. Now we must ask ourselves the broader question: Is the set of objectives and their associated attributes appropriate for the problem? In this regard, we shall define five properties--or should we say "objectives"--that are desirable for selecting a set of attributes.

2.4.1 Desirable Properties of a Set of Attributes

It is important in any decision problem that the set of attributes is complete, so that it covers all the important aspects of the problem; operational, so that it can be meaningfully used in the analysis; decomposable, so that aspects of the evaluation process can be simplified by breaking it down into parts; non-redundant, so that double counting of impacts can be avoided; and minimal, so that the dimensionality of the problem is kept as small as possible. Let us be specific about these properties.

a. Completeness: A set of attributes is complete if it is adequate for indicating the degree to which the overall objective is met. This condition should be satisfied when the lowest-level objectives in a hierarchy include all areas of concern in the problem at hand and when the individual attributes associated with each of the lowest-level objectives in this hierarchy satisfy the comprehensiveness criterion specified in Section 2.2.

There is another way to view the property of completeness. We have associated with each lowest-level objective, a single scalar attribute which takes on real values. Suppose an overall objective in the hierarchy has been subdivided into two subobjectives and scalar attribute X_1 has been associated with the first of these and scalar attribute X_2

with the second. We can think of measuring the overall objective with some vector attribute Y, which is some composite of attributes X_1 and X_2. A specific value of attribute Y will be a two-tuple (x_1, x_2), where x_i is a specified value of X_i. Now, to say the set of attributes X_1 and X_2 is complete is equivalent to saying that the vector attribute Y is comprehensive. Generalizing, a set of n attributes is complete if by knowing the value of the n-dimensional vector attribute associated with the overall objective, the decision maker has a clear picture about the extent to which the overall objective is met.

An illuminating example of a "good decision analysis gone astry" because the attribute set was not complete was written for one of us as an undergraduate thesis a few years back. The problem concerned the alternative course of actions to be followed after graduation: these included joining the military service, going to graduate school, or accepting a civilian position with a firm. The attributes included financial aspects, future flexibility, etc., but the author did not feel comfortable with the implication of his own formal study. The fault was that the analysis contained no considerations for the romantic life of the individual, and this factor was important enough to change the overall implications. He did not deem it suitable at first to bring sex into his attribute hierarchy. Clearly, with many people such aspects should be considered before signing up for work on the North Slope of Alaska or in a nuclear submarine for a five-year stay. But in subsequent iterations he learned how to become more honest himself and he finally reached a point where the formal analysis felt right to him and he acted accordingly. He referred to this experience as a cheap and orderly way to psychoanalyze oneself.

b. Operational: A set of attributes must be operational. This implies many different things depending somewhat on the intended use of the analysis. Basically, since the idea of decision analysis is to help a decision maker or decision makers choose a best course of action, the attributes must be useful for the purpose. The attributes must be meaningful to the decision maker, so that he can understand the implications of the alternatives. They should also facilitate explanations to others--especially in those cases where the main purpose of the study is to make and advocate a particular position. Consider the Mayor of a large city, who is appraising alternatives for handling solid wastes. It may not be possible for him in a publicly discussed study, to include an attribute like "annual number of tons of untreated solid waste dumped into the ocean" even though this amount might be extremely important. Given the analysis were to be released, inclusion of the attribute might make the Mayor too politically vulnerable. The analyst and decision maker must be aware of the many nontechnical problems which may render a set of attributes as nonoperational. Some of these issues are discussed in Keeney and Raiffa [1972].

c. Decomposable: A formal decision analysis requires that one quantify both the decision maker's preferences for consequences and his judgments about uncertain events. For a problem with n attributes, this means assessing an n-attribute utility function as well as joint probability distributions for the relevant uncertainties. Because of the complexity involved, these tasks will be extremely difficult, if not impossible, for decision problems in which the dimensionality n is even modestly high like five or so--unless the set of attributes is decomposable. By this we mean that the aforementioned tasks can be broken down into parts of smaller dimensionability. For instance, if the problem involves five attributes, it might be possible to break the assessments into two parts, one involving two attributes and one involving three. This idea, in the case of preferences, is one of the central themes of this book and is discussed in detail in chapters three through six.

d. Non-Redundancy: We do not want redundancies in our final set of attributes. The attributes should be defined to avoid double counting of consequences. For example, if one were evaluating a portfolio with investments in companies A and B, the attributes "income from company A" and "income from investments" are clearly redundant since income from company A is counted in both attributes. One should use just "income from investments" or "income from company A" and "income from company B" to avoid the redundancy. A more subtle example is discussed in McKean [1958] in conjunction with the allocation of water resources. Two attributes he considered were "increase in farm income" and "increase in livestock yield." These may be redundant in that the latter may be important only by virtue of its impact on the former.

This second illustration points out a common way that redundancies creep into a set of attributes. The problem is that the means-ends relationships of the objectives are not clearly indicated and attributes are included which are associated with both means and ends objectives.

Another way redundancies enter sets of attributes is by having some attributes represent variables which are inputs to a system and others represent variables which are outputs. An example of such a problem concerns the evaluation of space vehicles. An input might be "weigth" and an output might be "thrust" required to break out of the earth's gravitational field. Again, the former may only be important because of its implications on the latter.

e. Minimum Size: Subject to the four criteria for sets of attributes just discussed, it is desirable to keep the set as small as possible. Each time an objective is subdivided, possibilities for excluding important concerns occur. In addition, the difficulties in obtaining joint probability distributions and quantifying multiattribute preferences increase greatly as the number of attributes increases.

In some problems, it may be possible to combine attributes and hence reduce the dimensionality. For example, in the two company portfolio problem, the decision maker may not be concerned with whether his income comes from company A or company B in which case the single attribute "income from investments" would be appropriate.

The minimum size of a set of attributes is obviously one. One grandiose objective, suitably chosen, could be complete, and if we did not require that the set of attributes be operational, we could always pick such an objective*. However, as should be clear, in most complex decision problems this would not make the problem more tractable. Here, as in most problems of the real-world, we often want to fulfill conflicting objectives and since this is an ideal we cannot achieve, we must engage in vexing tradeoffs--which incidentally is the theme of this book.

2.4.2 Non-Uniqueness of a Set of Attributes

A set of attributes is not unique for a specific problem

---

*In Section 4.11 we discuss an example where a single attribute is both comprehensive and objectively measurable but nevertheless the attribute had to be partitioned into several lower level attributes in order for these to become operationally meaningful to the decision maker.

nor is it unique even for a specific objectives hierarchy. To illustrate this, consider the objective of an airline "to provide frequent service between Los Angeles and San Francisco." To measure this objective, one might use the number of flights per day, the maximum time between scheduled flights, or the average time between scheduled flights. In fact, the first and third suggested attributes are deterministically related. If n is the number of flights in a day and t is the average time in hours between flights, then $t = 24/n$.

As a second example, suppose X represents the crimes solved in one area and Y represents the crimes solved in another area. Then, if we were interested in the impact on crime in both of these areas, we could include X and Y in our total set of attributes. However, the average number of crimes solved, $(X + Y)/2$, and the difference in crime solved in the two areas, $X - Y$, could be used equally well. It should be clear that a knowledge of the effects of a program on these two attributes is equivalent to a knowledge of the effects on X and Y. The choice of which is a better set to use depends on the future uses of the analysis, and in particular on assessments of probabilities and utilities.

2.4.3 An Illustrative Example: Medical Treatment

Here we will try to tie together many of the properties discussed in the preceding subsections. These properties are intertwined in many ways as we hope to show. Also, quite naturally it turns out that the degree to which a certain set of attributes meets one meta-objective might only be improved at the expense of the degree to which it meets other meta-objectives.

Consider a simplification of the medical problem sketched out in the first chapter. A doctor about to perform a critical operation on a patient may have the overall objective to "do the best for the patient." We will avoid the question here about whose objective, the doctor's or the patient's, for the time being. Anyway, suppose this objective is divided into "minimize costs" and "avoid death." Then, as we have discussed, the attributes of total cost in dollars and the probability of death might be used for these objectives respectively. So if we define the overall objective as Y and costs in dollars as X_1 and probability of death as X_2, we have $Y = X_1 \times X_2$. The question is whether Y is complete. Since we have considered at length the desirable properties of attributes for lowest-level objectives, let us assume the attributes X_1 and X_2 satisfy these criteria. The question of whether Y is complete now reduces to whether or not the objectives "minimize costs" and "avoid death" cover all important aspects of the problem. As indicated in the beginning of Section 2.2, whether one concludes that all important aspects of a problem are included in a set of objectives is mainly a matter of resourcefulness in selecting additional objectives and judgment.

In our example, after some thought, it might be concluded that amount of pain and suffering that the patient might undergo would be important enough to influence decisions and hence should be represented by some objective. This might be formalized by including an objective to "minimize pain." With this, we would have three subobjectives under the overall objective. The original two were not complete.

A next step would be to assign a measure of effectiveness to the objective "minimize pain." As suggested earlier, this would likely be very difficult due to our inability to measure pain. It might be possible though to set up a subjective index appropriate for this purpose*. However, care must be taken to insure that this index is meaningful to the patient and/or the doctor. Otherwise, it would not be operational.

As a consequence, we may be forced to search for another attribute to indicate the degree of pain which is operational and possesses the other desired properties to the degrees possible. In this case, the "number of days which the patient must stay in bed" might be useful as such an attribute.While this clearly does not directly indicate the degree of pain, it is related in some manner to the amount of pain

---

*An interesting effort in this same spirit is the development of a severity of burn index by Gustafson, Feller, Crane, and Halloway [1971]. The work is briefly described in Kneppreth et.al. [1974].

suffered by the patient. Such attributes, which are called proxy attributes, are discussed in detail in the next section.

Suppose the patient and the doctor could meaningfully use a subjective index for "minimize pain" and suppose this, along with days in bed, cost, and probability of death, were suggested as a set of four attributes for the problem. In such a case, you might argue that days in bed may be eliminated from the list because it is redundant with the pain index. This would also reduce the number of attributes by one, which is desirable of course. Someone else may suggest eliminating the pain index in favor of number of days in bed for the same reason. Which of these suggestions is better would have to be weighed by the decision maker, and his choice should depend on the degree to which the remaining three attributes satisfy the various desirable properties for a set of attributes.

Going a bit further, one might decide that the particular circumstances of this problem make it such that the total cost is very closely related to the number of days in bed. This may also be directly related to the pain. So, it might be possible to eliminate both cost and the subjective pain index from the original list of four attributes and still end up with a complete, operational set of attributes--a set of two, namely 'the number of days in bed' and 'the probability of death'. This would have no redundancies and have the property that it is of minimal reasonable size.

The discussion of the preceding few paragraphs should clearly bring out the point that sets of attributes are not unique. We have suggested several combinations which might serve for a particular medical problem.

2.5 PROXY ATTRIBUTES AND DIRECT PREFERENCE MEASUREMENTS

In this section we are concerned with the age-old problem confronting analysts which one might raise with a statement like "...but what if we have specified an adequate objectives hierarchy and we just cannot find reasonable attributes for some of the lower-level objectives? We cannot go on subdividing objectives forever as you might suggest. And if we did this long enough, each of the objectives would fail to satisfy the test of importance; consequently they would be eliminated in further analysis, and we would have no attributes for some aspects of the hierarchy."

After reading this chapter to here, the question raised above may represent the thoughts of many. It is a very important question and invariably comes into play in complex decision problems. What can be done if no attributes reasonably meet the criteria discussed in Section 2.2? In many cases, one can use proxy attributes and direct preference measurements. These two concepts provide us with methods for surmounting the difficulties just raised. Their use, however, opens up additional ways that flaws can enter the analysis; but without them we can often only continue working on "half a problem." Let us discuss what we mean by these two concepts and when and how they should be used.

2.5.1 What are Proxy Attributes?

A proxy attribute is an attribute that reflects the degree to which an associated objective is met, but it does not directly measure this. Thus, the proxy attributes can be thought of as *indirectly* measuring the achievement on a stated objective. One could argue that essentially all attributes are proxy attributes because nothing can be absolutely measured. There are just varying degrees to which an objective is directly measured. Rather than get into a philosophical discussion which would not be very fruitful, let us illustrate some points with an example.

Some mathematical symbolism might help here. Suppose that in a given context we have a rather natural set of lower-level objectives measured by attributes $X_1, \ldots, X_i, \ldots, X_n$. Let us further assume that it would be relatively easy for the decision maker to state his preferences for attribute evaluations of the form $\underline{x} = (x_1, \ldots, x_n)$. But now let us assume that it is impossible because of measurement reasons to use the set of X-attributes. For example, in a decision concerning environmental standards one might be concerned with a set X of health attributes associated with different levels of pollution. One might simply not know very much about the linkage between a constellation of pollution levels--let us call these $\underline{y} = (y_1, \ldots, y_j, \ldots, y_r)$ where y_j might, for example, be the annual tonnage of particulate matter that is injected into the air over New York

City--and the ultimate health levels $\underline{x} = (x_1,\ldots,x_n)$. Now, conceptually speaking, for each $\underline{y}$ one could assign a probability distribution for the uncertain $\tilde{\underline{x}}$ associated with that $\underline{y}$. If $u_X(\underline{x})$ designated the utility for the composite health levels $\underline{x}$, then one could calculate an induced utility function u_Y over $\underline{y}$ levels by taking

$$u_Y(\underline{y}) = E_{X|\underline{y}}\, u_X(\tilde{\underline{x}}) \quad ,$$

where the operator $E_{X|\underline{y}}$ expects out the uncertain quantity $\tilde{\underline{x}}$ (a random variable) using the conditional probability distribution over $\tilde{\underline{x}}$ given $\underline{y}$. In schematic form this is depicted in Fig. 2.4. The branch $\underline{y}$ leads to a chance fork of $\underline{x}$-possibilities--really a continuum of $\underline{x}$ possibilities in n-space. One then assigns a utility value $u_X(\underline{x})$ for each end position and averages these utilities over the $\tilde{\underline{x}}$-fan using the conditonal probability distribution for $\tilde{\underline{x}}$ given $\underline{y}$.At position B in Fig. 2.4 one then obtains the induced utility value $u_Y(\underline{y})$. This is repeated for each $\underline{y}$. Now one can proceed in the usual way, backwards, by putting a probability distribution over $\tilde{\underline{y}}$ and averaging-out back to position A, and so on.

A situation where this procedure may be particularly desirable is when decisions are made to "improve life" in terms of the X attributes, but where the entire impact of the decision can be specified by its impact on the Y attributes. Use of the induced utility function u_Y could then greatly reduce the total effort involved, since one major

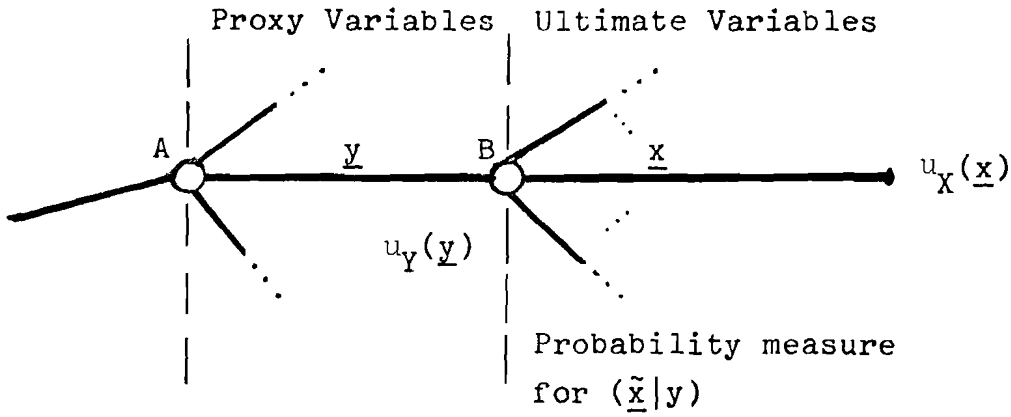


Figure 2.4

An Induced Utility Function u_Y

part of the model--once it has been done--can be neglected except for prudent periodical reviews.

In the case where there are several X attributes and only one Y attribute, we are effectively evaluating a multiattribute problem with the much simpler unidimensional framework. An air quality example where Y designates a single variable influenced by the decision alternatives is one plausible situation where this may occur.

But now let us suppose that we cannot responsibly assign a distribution to the chance fork B. In this case we can then subjectively assess directly our preferences or utilities for $\underline{y}$-configurations. Thus in using proxy variables $\underline{y}$ instead of the "ultimate"variables $\underline{x}$ we suppress, in Fig. 2.4, the chance fork emanating from B and use our mind as an informal synthesizer for directly assessing the $u_Y(\cdot)$ function.

Different decision makers using the same proxy variables $\underline{y}$ might differ in their u_Y assignments because they might differ on (a) the u_X assignments, (b) the probability distribution of $(\tilde{\underline{x}}|\underline{y})$, or (c) discrepancies arising from the informal synthesis of utilities and probabilities.

2.5.2 Example: Emergency Ambulance Service

The overall objective of an emergency ambulance system might be stated as "deliver patients to the hospital in the best possible conditions given the circumstances." Since there is no obvious attribute for this objective, suppose it is subdivided into "minimize the likelihood of death on arrival at the hospital" and "minimize the likelihood of arrival in critical condition." The proportion of patients dead on arrival and the proportion arriving in critical condition might be reasonable attributes for these objectives. However, the question of what is a critical condition would be difficult. Furthermore, a patient might receive the best care and treatment possible and still die enroute to the hospital. In such a case, the result should not be attributed to the competency of the ambulance service. But how would one differentiate this case from another where poor service contributed to the death of the patient? The point is that it might not be possible to identify suitable attributes which directly indicate the extent to which the objectives are achieved.

Faced with the problem of analyzing emergency ambulance systems, both Savas [1969] and Stevenson [1972] have chosen to use the proxy attribute "response time." This was defined as the time between receipt of a call for an ambulance and arrival of an ambulance at the scene. The "delivery time," the time between receipt of the call and arrival of the patient, is another important proxy attribute used in ambulance studies. The premise is that shorter response times and shorter delivery times will contribute to achieving the overall objective of an emergency ambulance system. And, because of this relationship, they may be used as attributes which reflect the degree to which this objective is achieved.*

2.5.3 The Mind as an Informal Synthesizer

When we use proxy attributes, the decision maker must process some additional information in his mind in choosing the best alternative. He must informally decide on the degree to which the objectives are met by the different levels of achievement as indicated by the proxy attributes.

The point is clarified by expanding on the ambulance example. Consider Fig. 2.5 which represents a simplified model of an emergency ambulance system. Our input variables are

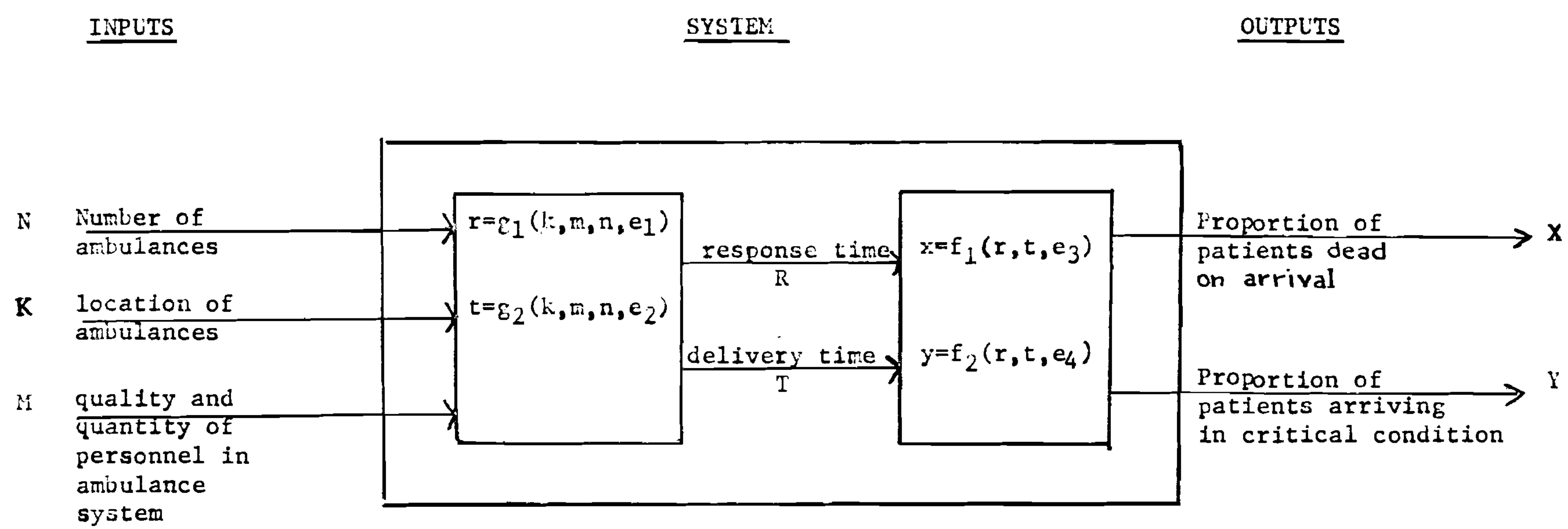
$N \equiv$ the number of ambulances,

$K \equiv$ the location of ambulances, and

$M \equiv$ quality and quantity of personnel in the system.

---

*Response time has been used as a proxy attribute in analyzing other emergency services. For example, Larson [1972] uses police response time in evaluating various allocation strategies in urban police departments, and Carter and Ingall [1970] use the response times of the various pieces of equipment answering calls for service in comparing operational policies available to the New York City Fire Department. See Section 7.3 for an attempt to aggregate the response times of these various pieces of equipment into an overall index of the quality of response to fires.



The e_i represent causal factors not included in the model and random disturbances.

Fig. 2.5 Proxy Attributes in a Simplified Model of an Emergency Ambulance System.

Although this may be vague, we would only complicate the discussion by being more specific. What we would like to do is measure the extent to which the objectives are met in terms of attributes X and Y, which represent the proportion of the patients arriving at the hospital dead and in critical condition, respectively. These can be thought of as the output of the system. The decisions control the inputs, and achievement is measured by the outputs.

However, we just argued that it might not be practical to use X and Y for evaluating the decisions, and as an alternative, we suggested using response time R and delivery time T for this evaluation. If our model gave us everything we wanted, we could get probability density functions for X and Y conditional on each possible decision. But it does not give us this, so we must settle for probability density functions over R and T. Now X and Y have some probabilistic relationship to R and T which we will designate* by

$$x = f_1(r,t,e_3)$$

and

$$y = f_2(r,t,e_4) \quad ,$$

where the e_i represent causal factors, other than response time and delivery time, and random disturbances. Our model does not indicate what f_1 and f_2 are and this is the reason

---

*Small letters will represent specific amounts of variables and attributes. That is, a specific value of response time R will be r.

we cannot get the probability density functions for X and Y. Actually both R and T are functions of N, K, and M and the model gives us

$$r = g_1(k,m,n,e_1)$$

and

$$t = g_2(k,m,n,e_2)$$

where g_1 and g_2 are those functions.

So what does one lose by using R and T rather than X and Y to evaluate the various courses of action? Presumably, when we ask the decision maker to express his preferences for different amounts of R and T, he does this by considering the effects R and T have on X and Y. But this requires an understanding of $f_1(r,t,e_3)$ and $f_2(r,t,e_4)$ or at least an understanding of how different values of R and T contribute to the overall objective of getting patients to the hospital in the best possible condition considering the circumstances. So essentially, the introduction of proxy attributes requires that some of the modelling of the system be done in the decision maker's head. Often, this is what we would like to avoid, because there is too much information in complex problems to handle effectively this way. However, when it is unavoidable, careful thinking may permit the decision maker to express a useful set of relationships between proxy attributes and the original objectives. It is probably safe to say that, in general, when a smaller part of the model must be implicitly considered by the decision maker, the quantified preferences more accurately reflect his true preferences for the basic objectives. For this reason, Hatry [1970] cautions against the excessive use of proxy attributes even though they might be easier to handle analytically or might be easily accessible.

It should be mentioned that R and T might still be useful even if the set X, Y is not complete. For example, suppose that a third attribute Z representing "annual cost of the ambulance system" is needed. There may again be problems with using X and Y on a practical problem, but Z itself may be adequate for cost considerations. In such a case, R and T might again be collectively proxy for X and Y, the service considerations, and the set R, T, and Z may reasonably satisfy our criteria for a set of attributes.

Suppose that in the ambulance problem, we could not build an analytical or simulation model of any sort; that is, we could not relate the inputs to the outputs or to any sets of proxy attributes which we felt might be appropriate for the problem. In this case, the decision maker might have to implicitly consider the entire model in his head by relating the possible levels of inputs to achieving the stated objective. This means that the decision maker must assess his preferences over various levels of K,M, and N by considering their effect on X and Y. And so, these three variables can be thought of as another set of proxy attributes which one might need to "fall back on" in our analysis. This indicates two points. First, there is no unique set of proxy attributes, and secondly, the proxy attributes can come in degrees. That is, some sets of proxy attributes are more closely related to the basic objectives than other sets.

2.5.4 Common Proxy Attributes

Earlier in this section, we remarked that all attributes might be proxy attributes because nothing measures completely and precisely all that we are interested in. But clearly some are "less proxy" than others. Here we would like to point out a couple of attributes which are so conventionally used that often one does not think of them as proxy attributes.

The best examples of this are attributes total wealth, income, or profits which are associated with the very commonly stated objective "maximize profits." However, is the basic objective to accumulate dollars for their own sake, or for other things such as consumer goods, the power to implement ideas, etc., which dollars help one to achieve? Probably, in many cases, the latter are more important, so profits can be thought of as a proxy attribute.

Another similar example concerns the "share of the market" which many large firms use in evaluating their relative position. But this might often be a proxy attribute for such intangibles as prestige and power. Or "share of the market" may be a proxy for future profits which in turn may be a proxy for other more basic attributes.

The fact is that for many problems, it is imperative to introduce proxy attributes in order to handle operationally some very messy difficulties.

2.5.5 Direct Preference Measurements

With both proxy attributes and subjective indices, one needs to obtain, for each alternative, a probability distribution for the various possible levels of the attribute, to assess a utility function over these levels, and finally to calculate the expected utility over the attribute for each course of action. The result would be a single number (expected utility over attribute Y) for each course of action indicating the preferences for that course of action relative to the others as far as that particular objective was concerned. In some instances, it may be virtually impossible to assess these probability distributions and the conditional utility function. When this is the case, the decision maker may prefer, or perhaps be forced for lack of alternatives, to directly assign a utility index of performance on a particular attribute for each of the various courses of action under consideration.

To illustrate the idea with a simple example, let us take a business with two objectives: "maximize profit" and "maximize goodwill." We will let X and Y designate the respective attributes for these objectives. For X, the measure "profits in dollars" may be chosen, but there appears to be no clear objective index for Y. Three options for handling this are a subjective index, a proxy attribute , or direct preference measurements. With a subjective index, the procedure should now be reasonably clear. We attempt to establish a scale of goodwill meaningful in the context of our problem at hand. Then, for each alternative,probability distributions are assessed to describe the possible impact in terms of Y, and a utility function is assessed over the Y attribute. Expected conditional utilities--conditional on the X attribute being held constant--can then be calculated for each alternative and used in the ensuing analysis.* With a proxy attribute, the process still involves assessing probability distributions over Y--now a proxy attribute--for each alternative and a conditional utility function. Then again, conditional expected utilities are calculated for each alternative. With direct preference measures, the story is different. The decision maker must directly assign the conditional expected utilities for achieving the objective "maximize goodwill." This avoids the formalism of specifying an attribute for goodwill, of assessing conditional probability distributions, and of assessing the conditional utility function. However, it clearly requires hard and thoughtful input on the part of the decision maker.

Some direct preference measurements are used by Miller

---

*Throughout this subsection, we are implicitly assuming the X attribute is held fixed. In Chapter 5, concepts are introduced which indicate when it is reasonable to conditionally calculate expected utility over one attribute while the other attributes are fixed at convenient levels.

[1969] in structuring the decision process for choosing among various employment opportunities. He used three attributes to describe continuing aspects of the jobs which would make them desirable. These were personal interest in the technical content of the job, degree of variety implicit in the job, and the amount of training in management skills realizable from the job. Preferences for four different jobs were assessed directly along each of these three attributes. Another use of direct preference measurements is discussed in the dynamic analysis of the Mexico City airport study, described in Chapter 8.

2.5.6 Some Comments on Proxy Attributes and Direct Preference Measurement

When one finds it necessary to use proxy attributes, or direct preference measurement, it is important to find attributes with which the decision maker is familiar. For instance, fire department officials are accustomed to thinking in terms of response times. When we then ask such a person his preferences, he will presumably be able to relate the response times to achievement of the basic objectives in a meaningful way. Similarly, one might expect a politician to directly assign preferences for alternatives in terms of the attribute "political effects." Essentially, in both cases, we are asking the decision maker to distill his years of experience in providing these preferences. The more accustomed the decision maker is to thinking in terms of the attribute, the more easily he will be able to express preferences and the more likely he will understand the complex relationships between the attribute, the alternatives, and the basic objectives.

A second point is probably very obvious to most readers. That is, for every proxy attribute we suggest, we can easily find an associated "proxy objective." For instance, the objective "minimize the emergency ambulance response time" is a proxy objective. We point this out because confusion on this matter can easily result in a redundant set of attributes for a problem. If one builds an objectives hierarchy for the ambulance problem with "minimize response time," "minimize the proportion of arrivals dead at the hospital," and "minimize the proportion of arrivals in critical condition," etc., one is likely to end up with redundancies in the final set of attributes.

Finally, note that in most of our examples, improving performance in terms of the proxy attributes, contributes to meeting the basic objectives. For instance, a lower response time contributes to "getting the patient to the hospital in the best possible condition." In some problems, it may be more convenient to look at performance on proxy attributes which is improved by meeting the basic objectives. For instance, one objective of a municipal sanitation service might be stated "to keep the streets clean." An attribute which might directly measure this would be "pounds of dirt and garbage per hundred yards of street." Proxy attributes like "number of garbage pickups per week" and "time between street cleanings" indicate performance which contribute to accomplishing the basic objective. On the other hand, a proxy attribute like "the number of citizen complaints about dirty streets per week" also indirectly indicates the level of service provided. In this case, however, presumably better service in terms of the basic objective causes better performance as measured by the proxy attribute.

2.6 SUMMARY AND PERSPECTIVE ON THE PRACTICAL ASPECTS OF SPECIFYING OBJECTIVES AND ATTRIBUTES

To attempt any formal analysis of a complex decision problem requires an articulation of the decision maker's objectives and an identification of attributes useful for indicating the degree to which these objectives are achieved. Unfortunately these objectives and attributes are not simply handed to us in an envelope at the beginning of an analysis. The intertwined processes of articulating objectives and identifying attributes are basically creative in nature. Hence, it is not possible to establish a step-by-step procedure which leads one in the end to a meaningful set of objectives and attributes.

What we have attempted to accomplish in this chapter is to set down some guidelines which may be useful in carrying out the necessary thought processes. At one end of the spectrum--the input side--some suggestions were included to help the decision maker and/or analyst probe his mind when facing the problems of obtaining objectives. At the other end of the spectrum, a set of criteria were suggested for the quality of the output of the objective and attribute generation processes. This output--namely the set of attributes--is crucial in the ensuing analysis. Since it is not usually the case that nice objective attributes are available to measure all the objectives in a complex problem, three specific procedures for handling such problems, subjective indices, proxy attributes, and direct preference measurements, were introduced and illustrated.

Before concluding this chapter, it seems appropriate to try to put some of the ideas we have discussed into proper perspective. Perhaps the biggest shortcoming of going through many examples such as we have done in the chapter is that inevitably, the overall feeling for what you are trying to do does not come through as well as some specific points used for illustrations, although the former is more important than the latter. This is mainly due to the fact that much realism is lost in reducing the problem at hand into written form and again in trying to distill that to bring out specific points. Without this reduction of scope, our ideas would probably be lost in the multitudes of words necessary to adequately describe all the relevant aspects of the problem. In establishing a meaningful objectives hierarchy and associated set of attributes for a complex problem, one can bring to bear many factors we have not explicitly considered here. The process of specifying the objectives is not done in a vacuum. At the same time, one may have relevant information about what data is accessible, the quality and quantity of other available resources (e.g., computers), various types of constraints which are in force, (e.g., time, political), the range of alternative courses of action, etc. All of these might significantly affect the objectives hierarchy and choice of attributes.

The message should be clear. Although we have offered some guidelines which will hopefully facilitate the selection of an objectives hierarchy and associated attributes, we view our work as far from complete. It would be erroneous to assume any of our suggestions can replace serious thinking and resourcefulness.

DECISION ANALYSIS WITH MULTIPLE
CONFLICTING OBJECTIVES
PREFERENCES AND VALUE TRADEOFFS

(Chapters 3 & 4)

Ralph L. Keeney and Howard Raiffa

May 1975 WP-75-53

CHAPTER 3

TRADEOFFS UNDER CERTAINTY

Many complex decision problems involve multiple conflicting objectives. It is often the situation that no dominant alternative will exist which is better than all other alternatives in terms of all of these objectives. Perhaps some of the original alternatives can be eliminated from further consideration because they are dominated, but in general you simply cannot maximize several objectives simultaneously. You cannot maximize benefits and at the same time minimize costs; you cannot necessarily maximize yield and minimize risk; nor can you share a pie by giving the maximum amount to each child. The literature is replete with high sounding rhetoric where an advocate cries out for doing "best" for everybody, in every possible way, in the shortest time, with the least inconvenience, and with the maximum security for all. Ah, for the simplicity of the romanticist's dream world!

3.1 THE MULTIATTRIBUTE VALUE PROBLEM

Our problem is one of value tradeoffs. In this chapter we will see what can be done about systematically structuring such tradeoffs. In essence, the decision maker is faced with a problem of trading-off achievement in terms of one objective against achievement in terms of another objective.

If there is no uncertainty in the problem, in the sense that the multiattribute consequence of each alternative is known, the essence of the issue is "How much achievement on objective 1 is the decision maker willing to give up in order to improve achievement on objective 2 by some fixed amount?" If there is uncertainty in the problem, the tradeoff issue remains, but difficulties are compounded because it is not clear what the consequences of each of the alternatives will be.

The tradeoff issue often boils down to a personal value question, and, in those cases, it requires the subjective judgement of the decision maker. There may be no right or wrong answers to these value questions, and naturally enough, different individuals may have very different value structures. If the tradeoff issue requires deep reflection--and we believe it often does in complex problems--there are two possibilities for resolving the issue; the decision maker can informally weigh the tradeoffs in his mind or he can formalize explicitly his value structure and use this to evaluate the contending alternatives. Of course, there are a mixture of intermediary possibilities between these two extremes. In this chapter, we shall discuss some techniques to help a confused decision maker formalize his or her own value structure. These provide a framework of thought which can be used by the decision maker to assist him in articulating his preferences.

3.1.1 Statement of the Problem

Let a designate a feasible alternative and denote the set of all feasible alternatives by A. To each act a in A we will associate n indices of value: $X_1(\underline{a}),\ldots,X_n(\underline{a})$. We can think of the n evaluators $X_1,\ldots,X_n$ as mapping each a in A into a point in an n-dimensional consequence space, as shown in Figure 3.1.

Often we shall talk about some attribute X, such as the aesthetic appeal of a design, and about an evaluator X of this attribute. We unashamedly will use the same symbol X for the attribute in question and the evaluator of that attribute. The context will make it clear what we are talking about and sometimes it is just plain convenient not to draw distinctions between these two notions.

In this chapter, we take the point of view that the n attributes are given. But, of course, one has to keep in mind that in practice, we have to design and create these attributes that purport to describe the consequences of actions. The ideas of Chapter 2 may be useful for this task.

Observe that if $(x_1,x_2,\ldots,x_n)$ is a point in the consequence space, we will never compare the magnitudes of x_i and x_j, for $i \neq j$, since in most situations this would be meaningless because attributes X_i and X_j may be measured in totally different units.

Roughly--and this is really "roughly"--the decision maker's problem is to choose a in A so that he will be happiest with the payoff $X_1(\underline{a}),\ldots,X_n(\underline{a})$. Thus we need an

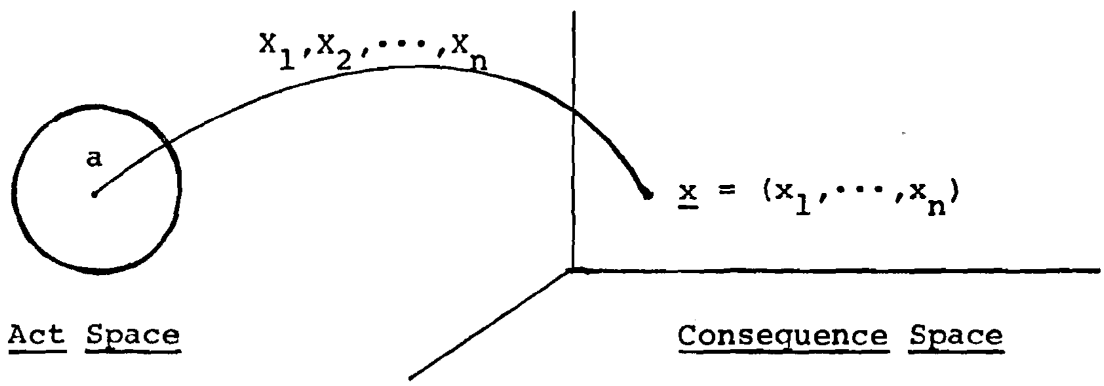


Figure 3.1

The Mapping of Acts into Consequences

index that combines $X_1(\underline{a}),\ldots,X_n(\underline{a})$ into a scalar index of preferability or value. Alternatively stated, it would be satisfactory to specify a scalar-valued function v defined on the consequence space with the property that

$$v(x_1,x_2,\ldots,x_n) \geq v(x_1',x_2',\ldots,x_n') \Longleftrightarrow (x_1,x_2,\ldots,x_n) \succsim (x_1',x_2',\ldots,x_n'),$$

where the symbol $\succsim$ reads "preferred or indifferent to". We refer to the function v as a <u>value function</u>. The same construct has many other names in the literature such as ordinal utility function, preference function, worth function, or utility function. Given v, the decision maker's problem can now be stated to choose $\underline{a}$ in A such that v is maximized. The value function v serves to compare various levels of the different attributes indirectly, through the effects the mangitudes $x_i, i=1,\ldots,n$, have on v.

3.1.2 Organization and Perspective of the Chapter

Our main consideration is how to structure and assess a value function v. It would be nice if we could find some function, call it f, with a simple form such that

$$v(x_1,x_2,\ldots,x_n) = f\left[v_1(x_1),v_2(x_2),\ldots,v_n(x_n)\right] \quad ,$$

where v_i designates a value function over the single attribute X_i. Some of the constructions of v in this chapter do exactly this.

However, before delving deeply into this problem, we shall first discuss some concepts which do not require complete formalization of the preference structure. In some cases, this may prove to give us enough information for a responsible decision. Then we respectively consider the structure of value functions where there are two, three, and more than three attributes. This is followed by a rather detailed illustration of the assessment of multiattribute value function.

It is important to point out that much of this chapter is expository in nature. Many of the concepts and results discussed are due to other individuals including Debreu [1960], Gorman [1968a,1968b], Krantz et al. [1971], Leontief [1947a,1947b], Luce and Tukey [1964], Pruzan and Jackson [1963] and Ting [1971]. When important results are stated, they will be designated as theorems for easy reference, but in many cases the formal proofs will be omitted since the proofs are accessible in the original works. We will, however, try to capture the ideas of these theorems with several 'informal proofs'. One price we pay for this is that assumptions such as continuity, differentiability, essentiality, and solvability, which are often utilized in the formal proofs, are sometimes ignored in our informal ones. Essentially we assume without much ado in this chapter exactly what is necessary to make our reasoning work and we concentrate on only the simplest nonpathological cases. In later chapters, where our work becomes less expository, we become a bit more formal and careful.

Section 3.9 attempts to provide the reader with a brief guide to the literature on multiattribute value functions.

In summary then, this chapter will look at the certainty case--the case where associated to each alternative there is a certain known consequence in n-space. In ensuing chapters we will look at the probabilistic case--the case where we only know the associated payoff in the consequence space in probabilistic terms. Techniques developed for the certainty case will prove useful also for the probabilistic case.

3.2 CHOICE PROCEDURES WHICH DO NOT FORMALIZE PREFERENCE STRUCTURES

Let acts a' and a" have consequences

$\underline{x}' = (x_1', \ldots, x_i', \ldots, x_n')$ and $\underline{x}'' = (x_1'', \ldots, x_i'', \ldots x_n'')$

where

$X_i(a') \equiv x_i'$ and $X_i(a'') \equiv x_i''$, for $i = 1, \ldots, n$.

Furthermore, let us assume throughout this section that preferences* increase in each X_i.

3.2.1 Dominance

We shall say that $\underline{x}'$ dominates $\underline{x}''$ whenever

---

*More formally, in terms of vernacular to be introduced later, we assume that each X_i is preferentially independent of the complementary set of attributes (see Section 3.5), and that preferences increase in each X_i.

a. $x_i' \geq x_i''$, all i , (3.1)

and

b. $x_i' > x_i''$, for some i . (3.2)

If $\underline{x}'$ dominates $\underline{x}''$, then the act a'' is a <u>noncontender</u> for "best", since a' is at least as good as a" for every evaluator (given by (3.1)), and strictly better for at least one (given by (3.2)).

In the case n = 2, we can plot the points $\underline{x}'$ and $\underline{x}''$ as in Fig. 3.2 and we see that $\underline{x}'$ dominates $\underline{x}''$ if and only if $\underline{x}'$ is "northeast" of $\underline{x}''$.

Observe that the notion of dominance exploits only the <u>ordinal</u> character of the numbers in the consequence space (i.e., given two numbers $x_i' = 6$ and $x_i'' = 3$ we are interested in the relationship that $x_i' > x_i''$) and not the <u>cardinal</u> character of these numbers (i.e., the fact that the difference between 10 and 6 is greater than the distance from 6 to 3 or that 6 is twice 3). Also observe that dominance does not require comparisons between x_i' and x_j'' for $i \neq j$.

3.2.2 The Efficient Frontier

For any (feasible) act $\underline{a} \varepsilon A$ there is an associated consequence $\underline{x}$ in n-space (i.e., the evaluation space) where $x_i \equiv X_i(a)$, all i. Let R be the set of consequences in n-space which are associated with acts in A--the set R is the so-called range-set of the vector X of evaluators $X_1, \ldots, X_n$ which are defined on the domain A.

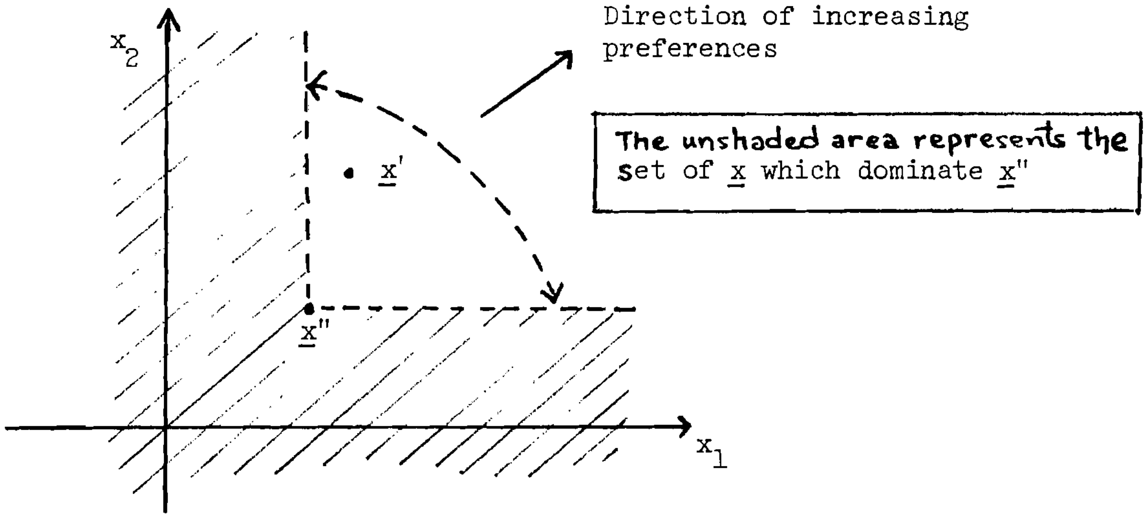


Figure 3.2

Dominance with Two Attributes

Fig. 3.3 depicts various range-sets, R, when n= 2. We shall have occasion in the sequel to discuss these qualitatively different cases.*

The set of consequences of R that are not dominated will be called the efficient frontier of R. It is also known as the "Pareto optimal set." In Fig. 3.3A, B, and C, the efficient frontiers are darkened. Thus in Fig. 3.3A the choice of $\underline{x}''$ can be ruled out because there is the consequence $\underline{x}'$ in the efficient frontier which dominates $\underline{x}''$. In Fig. 3.3C the consequence $\underline{x}^{(3)}$ is efficient (i.e., lies on the efficient frontier) even though it lies in a local valley, so to speak. In Fig. 3.3D the set R consists of discrete consequences and the efficient points are marked with an overlaying ●. The cases depicted in Fig. 3.3A and B are the easiest to handle analytically, since the sets of consequences are convex and the efficient frontiers continuous. Notice, however, that the concept of convexity introduces cardinal (as opposed to ordinal) notions.

---

*We don't want to be too fussy about mathematical details but somehow we must rule out pathological cases or else we will get into trouble. We shall assume that the region R is bounded and that it contains all of its boundary points. That is, we definitely want to rule out the case where there is a sequence of points $\underline{x}^{(1)}, \underline{x}^{(2)}, \ldots, \underline{x}^{(m)}, \ldots$ in R such that each point in the sequence dominates the preceding consequence, and where the sequence approaches some point $\underline{x}^*$, say, which does not belong to R.

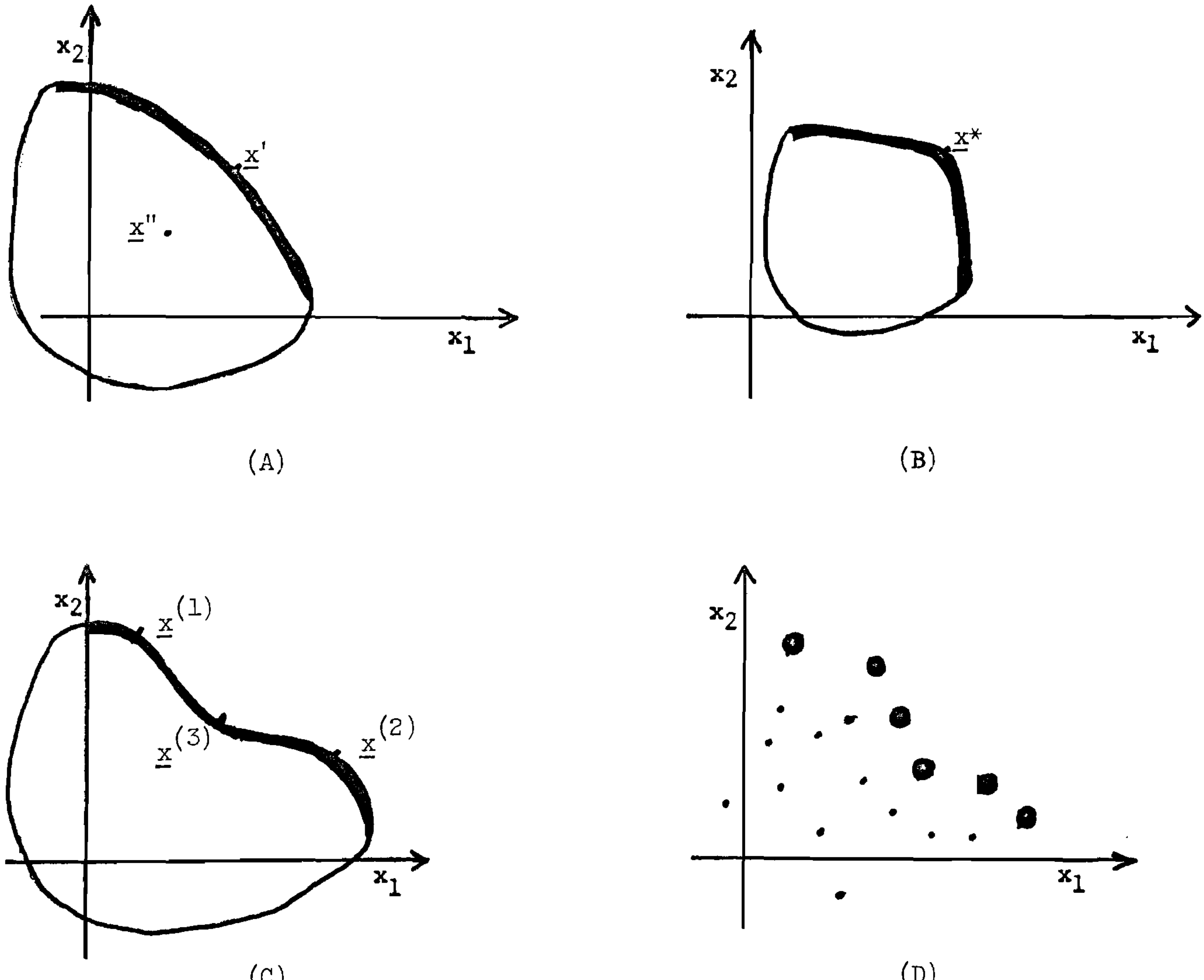


Figure 3.3

The Efficient Frontier for Various Sets of Consequences with Two Attributes

In some cases where the efficient set can be drawn it might be pretty obvious which $\underline{x}$ should be chosen. For example, in Fig. 3.3B the point $\underline{x}^*$ naturally suggests itself because one has to sacrifice so much of one attribute to gain so little of another attribute when moving slightly from $\underline{x}^*$. Admittedly, we are implicitly using cardinal concepts in making this last remark, but the natural units for the X_1 and X_2 evaluators might make such cardinal tradeoffs manifestly clear. We are not saying this is necessarily so; just that it might, on occasion, be so.

For values of $n > 3$ we cannot picture R and its efficient frontier. The next two sections describe two ways the decision maker can "move around" on an efficient frontier in order to locate a point that seems reasonably good. Later sections will describe procedures a decision maker can use to formally structure his preferences for points in the evaluation space. But meanwhile let's look at what can be done without completely specifying such a preference structure.

3.2.3 Exploring the Efficient Frontier: Use of Artificial Constraints*

The decision maker is faced with the following problem.

---

*Some references covering topics briefly discussed in this subsection are Dyer [1972]; Geoffrion, Dyer, and Feinberg [1972]; Kornbluth [1973]; Roy [1971]; and Schroeder [1974].

He must select an act $\underline{a}\varepsilon A$ so that he will be "satisfied" with the resulting n-dimensional payoff: $X_1(\underline{a})$, $X_2(\underline{a}),\ldots,$ $X_n(\underline{a})$. One procedure he might employ is to think of some "aspiration levels" $x_1^\circ, x_2^\circ, \ldots, x_n^\circ$ for the n attributes and pose the well-defined, mathematical problem: Is there an $\underline{a}\varepsilon A$ such that

$$X_i(\underline{a}) \geq x_i^\circ \text{ , for } i = 1,\ldots,n\,? \qquad (3.3)$$

Is it possible to satisfy these joint aspirations? If no, then the decision maker can change his joint aspirations to some point $x_1', x_2', \ldots, x_n'$. If yes--i.e., if an act $\underline{a}^+$ exists that satisfies (3.3)--then although we know that

$$X_i(\underline{a}^+) \geq x_i^\circ \text{ , for } i = 1,\ldots,n \text{ ,}$$

we still don't know that the point

$$(X_1(\underline{a}^+), X_2(\underline{a}^+), \ldots, X_n(\underline{a}^+))$$

is efficient. It may be dominated. We might continue our probing procedure by setting up another aspiration level $(x_1', \ldots, x_n')$ where

$$x_i' = X_i(a^\dagger) + \Delta_i \text{ , } i = 1,\ldots,n$$

and where the increment Δ_i is chosen in an ad hoc, intuitive manner that is a combination of wishful thinking and realism. Thus in an iterative manner the decision maker can investigate the frontier or "almost-frontier" of R. By informally keeping his preference in mind, he can choose a succession of aspiration levels which can move him around the region R until he reaches the limits of his patience, or until he figures that the expected gain of continuing the probing procedure is not worth the effort in time and cost of analysis.

Perhaps a more satisfactory variation of this procedure consists of setting aspirational levels for all attributes, save one. For example, suppose the decision maker selects aspiration levels $x_2^\circ, x_3^\circ, \ldots, x_n^\circ$ and seeks an $\underline{a}\varepsilon A$ which satisfies the imposed constraints

$$X_i(\underline{a}) \geq x_i^\circ \ , \text{ for } i = 2,3,\ldots, n \qquad (3.4)$$

and maximizes $X_1(\underline{a})$.

This maximization problem is in the form of the "standard optimization problem" of Section 3.1.1. If there is no feasible solution (i.e., no $\underline{a}\varepsilon A$ which satisfies (3.4)), then obviously the set of aspirations $x_2^\circ, \ldots, x_n^\circ$ has to be changed. But even if a feasible solution exists, the decision maker may be surprised at the maximum value of $X_1(\underline{a})$. If it is either too small or too large (as compared to what he "expected") he might want to change the original aspiration levels $x_2^\circ, \ldots, x_n^\circ$ and iterate the procedure.

Let the maximum of $X_1(\underline{a})$ subject to constraints $\underline{a}\varepsilon A$ and (3.4) be denoted by $M_1(x_2^\circ, \ldots, x_n^\circ)$. The notation emphasizes the point that the maximum depends on the aspiration levels $x_2^\circ, \ldots, x_n^\circ$. It is often the case that as a by-product of the solution procedure of the standard optimization problem, we get the local rate of change of M_1 as each of the constraints is released (all others remaining fixed).

In mathematical terms we obtain the partial derivatives

$$\frac{\partial}{\partial x_j^\circ} M_1(x_2^\circ,\ldots,x_n^\circ)$$

for $j = 2,\ldots,n$. Now the decision maker has a lot of information at his disposal. He chooses $x_2^\circ,\ldots,x_n^\circ$ and then as a result of the analysis he obtains

$$M_1(x_2^\circ,\ldots,x_n^\circ) \quad \text{and} \quad \frac{\partial}{\partial x_j^\circ} M_1(x_2^\circ,\ldots,x_n^\circ) \quad \text{for } j = 2,\ldots,n \quad .$$

He now has to decide either to remain satisfied with what he has or to probe further. If he decides to continue his search for a "satisfactory" solution he might wish to single out some index, say j, and investigate the behavior of

$$M_1(x_2^\circ,\ldots,x_{i-1}^\circ,x_j,x_{j+1}^\circ,\ldots,x_n^\circ)$$

as a function of x_j. That is, he might choose to keep intact all the previous constraints, other than x_j°, and to systematically observe what happens to M_1 as x_j moves over some given range. He does this even though he already knows the value of M_1 at x_j° and the derivative at this point, because this additional information may be useful, and the cost of the additional analysis may be quite small. Fig. 3.4 shows one possible result of such an analysis.

The above investigative, probing procedure is <u>ad hoc.</u> It is not precisely programmed. It requires a series of creative judgments from the decision maker. He has to decide on aspiration levels, on special investigations of the

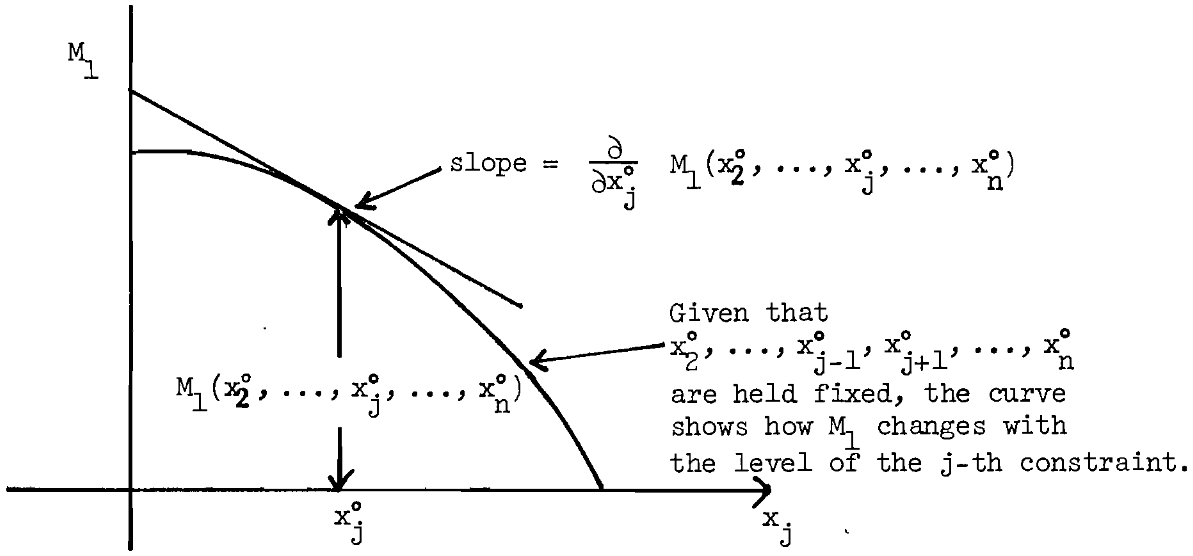


Figure 3.4

Exploring the Efficient Frontier Using Artificial Constraints sensitivity of payoffs (like M_1) to his arbitrarily imposed constraints, on setting new aspiration levels, and so on; and finally, he must decide when to be "satisfied" and stop. This probing procedure involves a continuing interaction between analyzing what is achievable and what is desirable. It proceeds incrementally, where the choice of each step is decided upon by the decision maker who must constantly weigh informally in his mind what he would like to get and what he thinks he might be able to get. Interactive computer programs have been written to help make this iterative probing operational. In the next subsection we shall discuss one more way of exploring the efficient frontier in n-space.

3.2.4 Exploring the Efficient Frontier: Use of Variable, Linear Weighted Averages*

In this section we shall pose an auxiliary mathematical problem, the solution of which will result in the identification of some point on the efficient frontier. By modifying the auxiliary problem, the decision maker can move along the efficient frontier until he is satisfied with the result.

For any $\underline{a} \epsilon A$ we assume, as before, there is the payoff $X_1(\underline{a}),\ldots,X_n(\underline{a})$. Let

$$\underline{\lambda} = (\lambda_1,\lambda_2,\ldots,\lambda_n) \tag{3.5a}$$

be any n-tuple for which

$$\lambda_i > 0, \quad \text{all } i \tag{3.5b}$$

---

*For procedures which generate the entire efficient frontier using linear programming, see Zeleny [1974].

and

$$\sum_{i=1}^{n} \lambda_i = 1 \; . \qquad (3.5c)$$

Define the auxiliary problem as follows: Choose $\underline{a} \varepsilon A$ to maximize

$$\sum_{i=1}^{n} \lambda_i X_i(\underline{a}) \; . \qquad (3.6)$$

We can also state the problem in an equivalent fashion as follows:Choose $\underline{x} \varepsilon R$ to maximize

$$\sum_{i=1}^{n} \lambda_i x_i \; . \qquad (3.7)$$

This auxiliary problem is in the form of a standard optimization problem. Let $\underline{x}^{\circ} = (x_1^{\circ},\ldots,x_n^{\circ})$ be a solution to this auxiliary problem. We now assert that $\underline{x}^{\circ}$ must lie on the efficient frontier. For, suppose it did not; then there would be an $\underline{x}'$ belonging to R which would dominate $\underline{x}^{\circ}$. But this cannot be, since in that case

$$\sum_{i=1}^{n} \lambda_i x_i' > \sum_{i=1}^{n} \lambda_i x_i^{\circ}$$

and therefore $\underline{x}^{\circ}$ would not be a maximizer of $\Sigma \lambda_i x_i$.

Hence corresponding to an n-tuple $\underline{\lambda}$ satisfying (3.5), the maximizer of $\Sigma \lambda_i x_i$ (for $\underline{x}$ in R), results in a point $\underline{x}^{\circ}$ which lies on the efficient frontier.

The geometry of this analysis is shown in Fig. 3.5 for n = 2, when $\underline{\lambda} = (.8,.2)$. The point $\underline{x}^{\circ}$ is a maximizer of

$$.8x_1 + .2x_2 \; .$$

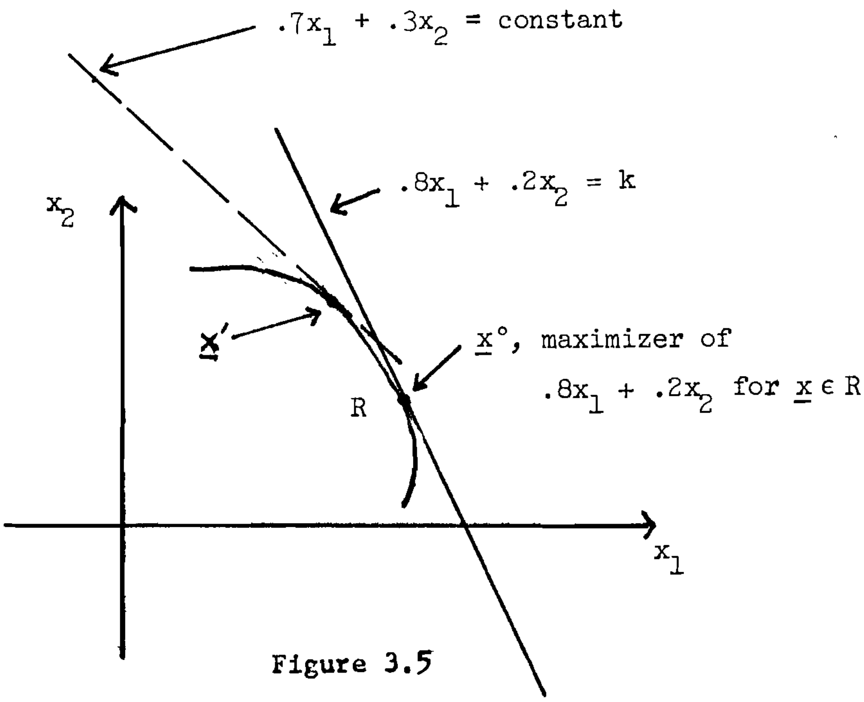


Figure 3.5

Exploring the Efficient Frontier Using Linear Weighting

The line through $\underline{x}^o$ of the form

$$.8x_1 + .2x_2 = k$$

(for a suitably chosen k) must be tangent to R at $\underline{x}^o$, since this line obviously contains $\underline{x}^o$ and no point of R can be to the right of this line (otherwise, $\underline{x}^o$ would not be a maximizer of $.8x_1 + .2x_2$).

Now the decision maker can query his psyche and ask himself whether he wants to settle for $\underline{x}^o = (x_1^o, x_2^o)$ or to explore the efficient frontier further. He knows that at $\underline{x}^o$ he can move along the frontier of R trading off Δ units of X_1 for approximately 4Δ units of X_2. That tradeoff is only precisely true in a limiting sense but for practical purposes we can think of 1 to 4 as the (local) marginal rate of substitution of X_1 for X_2 at the frontier point $\underline{x}^o$. Suppose the decision maker, upon reflection, feels that the value of x_2^o is too low in comparison with x_1^o (i.e., he would be willing to give up some of X_1 to get more of X_2). He can then resolve the auxiliary problem by looking for a maximizer of, say

$$.7x_1 + .3x_2 \quad ,$$

for $\underline{x} \varepsilon R$. If $\underline{x}' = (x_1', x_2')$ is such a maximizer, then $\underline{x}'$ will also be on the efficient frontier of R and $\underline{x}'$ will lie northwest of $\underline{x}^o$ as seen in Fig. 3.5. At $\underline{x}'$ the (local) marginal rate of substitution will be Δ units of X_1 for $7\Delta/3$ units of X_2. And so the process goes.

Of course, if n = 2, the efficient frontier can be pictured. The real power of the technique can best be appreciated for higher values of n where the geometry can only be imagined but not drawn. For example, if the choice of $\underline{\lambda} = (\lambda_1, \ldots, \lambda_n)$ gives rise to the associated maximizer $\underline{x}^\circ = (x_1^\circ, \ldots, x_n^\circ)$, and if x_i° appears to be unsuitably low, then the auxiliary maximization problem can be recycled with an increased value of λ_i. This will result in an increase--to be precise, it will not result in a decrease--in the optimal level x_i in the new maximization problem. The decision maker by looking at the points he has already obtained on the efficient frontier must decide when to be satisfied. By manipulating the λ_i's he can always move to different points on the efficient frontier. Once again he is asked informally to balance what he would <u>like</u> to get with what he <u>thinks he can achieve.</u> If the efficient frontier is convex, with no local dips or valleys, the procedure which manipulates the λ_i's can generate any point on the frontier. In the non-convex case, special techniques can be employed to map out these dips, but since this procedure is not in the main stream of our concern, we will not explore these variations.

It is possible to proceed further. One might want to formalize some variation of the above iterative procedure and prove convergence to an optimum. Of course, if one proceeds along these lines, one would have to imagine that lurking in the background there is a complete ordering of the points in n-space which is called upon at each step of the iteration in order to guide the choice of each incremental adjustment. Since this line of approach does not generalize readily to the probabilistic case, which is after all our main orientation, we shall not pursue the numerous analytical points emanating from the discussion in this subsection.

3.3. STRUCTURING OF PREFERENCES AND VALUE FUNCTIONS

We now turn our attention to a new tack--one which formalizes the decision maker's preferences for points in the consequence space. As is commonly done in economics, we initially forget about the set of achievable points in n-space (i.e., the set R in the Section 3.2) and discuss the decision maker's preferences for consequences in n-space, whether they actually belong to R or not. Only after formalizing these preferences, do we then investigate the problem of finding a point in R that will yield his greatest preference.

3.3.1 Lexicographical Ordering

As our first illustration we shall examine an approach which we believe is more widely adopted in practice than it deserves to be. However, it has the merit of simplicity and it can be easily administered. Our objection is that it is naively simple. It is called lexicographical ordering.

A lexicographical ordering is like the ordering found in a lexicon or dictionary: a' $\succ$ a" if and only if:

a. $X_1(a') > X_1(a'')$

or

b. $X_i(a') = X_i(a'')$ for $i = 1,\ldots,k$, and $X_{k+1}(a') > X_{k+1}(a'')$, for some $k = 1,\ldots,n-1$.

In other words we assume that the evaluators $X_1,\ldots,X_n$ are ordered according to importance. Act a' is preferred to a" if it merely has a higher score on X_1--regardless of how well or poorly it does on other evaluators. Only if there is a tie on X_1, does evaluator X_2 come into consideration.Only if there is a tie on X_1 and X_2, does evaluator X_3 come into consideration. And so on. Naturally, we can generalize this formulation by permuting the prominence of the evaluators. We can, for example, make X_3 most important, followed by X_1, followed by ...

Notice that if $\underline{x}'$ and $\underline{x}''$ are distinct points in an evaluation space,then they cannot be indifferent with a lexicographical ordering.

A lexicographical ordering is easy to understand and, in some (very rare!) cases, it might reflect the "true" beliefs of the decision-making unit. However,it is our belief that--leaving aside "administrative ease"--it is rarely appropriate. But, of course, "administrative ease," is an important meta-evaluator in its own right, and cannot be ignored. Hence, we do observe cases where lexicographical orderings are employed.

<u>A variant of lexicographical ordering with aspiration levels</u>,

Suppose we order the evaluators in importance and for convenience let us use the natural ordering 1, 2,.... For each evaluator X_i set an aspiration level x_i° and posit the following rules: a' $\succ$ a'' whenever

a. $X_1(a') > X_1(a'')$ and $X_1(a'') < x_1$

(i.e., X_1 overrides all else as long as X_1 aspirations are not met), or

b. $X_1\ (a') \geq x_1^{\circ}$

$X_1\ (a'') \geq x_1^{\circ}$

$X_2\ (a') > X_2(a'')$ and $X_2(a'') < x_2^{\circ}$

(i.e., if X_1 aspirations are met, then X_2 overrides all else as long as X_2 aspirations are not met).

An so forth. If all aspiration levels are met, then one may be willing to give up some of X_1 for a suitably large increase in X_2, and so on. In this ordering system two distinct points $\underline{x}'$ and $\underline{x}''$ might be indifferent provided that $x_j' > x_j^{\circ}$ and $x_j'' > x_j^{\circ}$, for all j.

Again we feel that such an ordering procedure, if carefully scrutinized, will rarely pass a test of "reasonableness," but for administrative purposes such an ordering might indeed be imposed.

In the sequel we shall only deal with preference structures that are less dogmatic in the sense that: if $\underline{x}'$ is an interior point of R, then for a suitably small decrease in x_i' there will be a suitably large compensating increase in x_j'. In two-space, this means that every point $\underline{x}$ lies on some indifference curve.

3.3.2 Indifference Curves

Fig. 3.6. depicts an example of how a decision maker might structure his preferences for points in a two-dimensional evaluation space. This example assumes that the decision maker is indifferent between achieving $\underline{x}'$ or $\underline{x}''$ and this is portrayed by having both $\underline{x}'$ and $\underline{x}''$ on the same indifference curve. The point $\underline{x}'''$ is preferred to $\underline{x}'$ (by the decision maker) and hence $\underline{x}'''$ lies on a higher (or more preferred) indifference curve.

We imagine that through any point $\underline{x}$ in an n-dimensional consequence space there is an indifference surface connecting all points that are indifferent to $\underline{x}$. These indifference surfaces will be curves for n=2. We shall assume throughout, that in the opinion of the decision maker, any two points $\underline{x}^{(1)}$ and $\underline{x}^{(2)}$ are comparable in the sense that one, and only one, of the following holds:

a. $\underline{x}^{(1)}$ is indifferent* to $\underline{x}^{(2)}$ (written: $\underline{x}^{(1)} \sim \underline{x}^{(2)}$,

b. $\underline{x}^{(1)}$ is preferred to $\underline{x}^{(2)}$ (written: $\underline{x}^{(1)} \succ \underline{x}^{(2)}$),

c. $\underline{x}^{(1)}$ is less preferred than $\underline{x}^{(2)}$ (written: $\underline{x}^{(1)} \prec \underline{x}^{(2)}$).

We write $\underline{x}^{(1)} \succsim \underline{x}^{(2)}$ to mean "not $\underline{x}_1^{(1)} \prec \underline{x}^{(2)}$" and assume all the relations $\sim, \succ, \succsim$ to be transitive.

We shall say that a preference structure is defined on the consequence space if any two points are comparable and no intransitivities exist. We assume, also, that the de-

*Less elliptically, and more grammatically, we could say, "The decision maker is indifferent between $\underline{x}^{(1)}$ and $\underline{x}^{(2)}$."

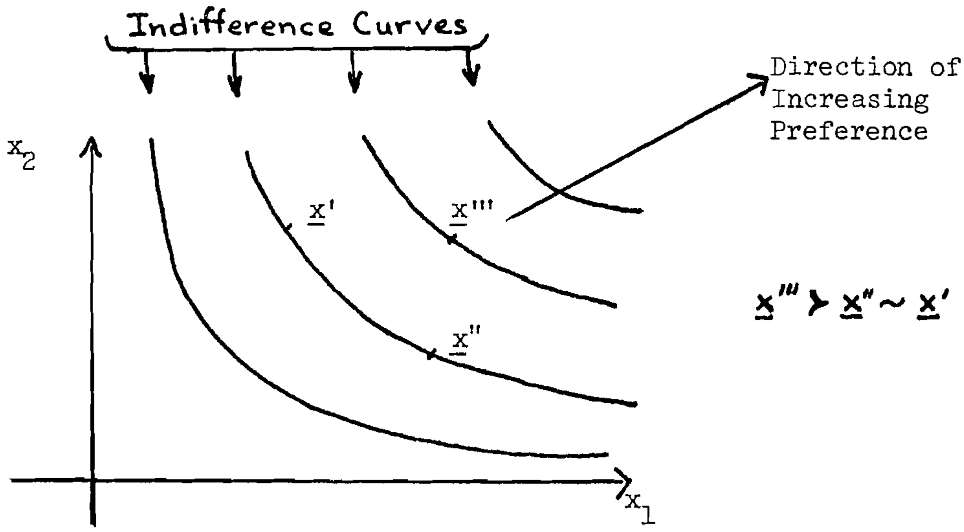


Figure 3.6

A Set of Indifference Curves

cision maker believes that in a specified decision context there is some particular preference structure that is appropriate for him.

Once the decision maker has specified his preference structure he can proceed to formalize his problem, namely: Find $a^{\circ} \varepsilon A$ such that

$$\underline{X}(a^{\circ}) \succsim \underline{X}(a) \quad , \text{ for all } a \varepsilon A$$

where

$$\underline{X}(a) \equiv (X_1(a), X_2(a), \ldots, X_n(a)) \quad .$$

Or, alternatively stated: Find $\underline{x}^{\circ} \varepsilon R$ such that

$$\underline{x}^{\circ} \succsim \underline{x} \quad , \text{ for all } \underline{x} \varepsilon R \; .$$

Fig. 3.7 depicts the geometry of this maximization problem.

3.3.3 Value Functions

A function v, which associates a real number $v(\underline{x})$ to each point $\underline{x}$ in an evaluation space, is said to be a <u>value function</u> representing the decision maker's preference structure provided that

$$\underline{x}' \sim \underline{x}'' \Longleftrightarrow v(\underline{x}') = v(\underline{x}'') \quad , \tag{3.8a}$$

and

$$\underline{x}' \succ \underline{x}'' \Longleftrightarrow v(\underline{x}') > v(\underline{x}'') \quad . \tag{3.8b}$$

Some typical examples of value functions for n = 2 are:

$$v(\underline{x}) = c_1 x_1 + c_2 x_2 \quad , \quad \text{where } c_1 > 0, \; c_2 > 0 \quad ,$$

$$v(\underline{x}) = x_1^{\alpha} x_2^{\beta} \quad , \qquad \text{where } \alpha > 0, \; \beta > 0 \quad ,$$

$$v(\underline{x}) = c_1 x_1 + c_2 x_2 + c_3 (x_1 - b_1)^{\alpha} (x_2 - b_2)^{\beta} \; .$$

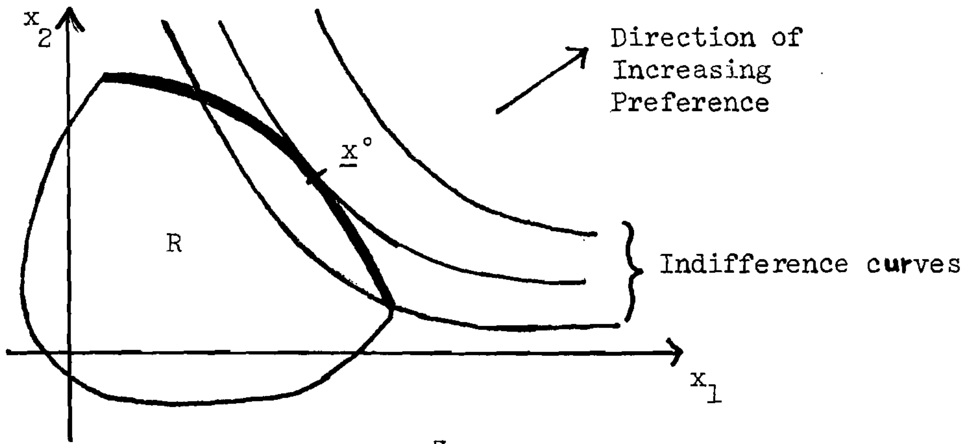


Figure 3.7

The Consequence $\underline{x}°$ is the Best Consequence in R

If v is a value function reflecting the decision maker's preferences, then his problem can be put into the format of the standard optimization problem:
Find $\underline{a} \varepsilon A$ to maximize $v[\underline{X}(\underline{a})]$.

We shall see later, that there is a subtle interplay between formulating a preference structure and finding a corresponding value function. Indeed, we may employ value functions to <u>help</u> a decision maker articulate his preferences.

3.3.4 Indifference Curves and Value Functions

Given a value function v, any two points $\underline{x}'$ and $\underline{x}''$ such that $v(\underline{x}') = v(\underline{x}'')$ must be indifferent to each other and must lie on the same indifference surface. Hence we see that given v it is possible, in principle, to find the indifference surfaces. More generally we see that a knowledge of v <u>uniquely</u> specifies an entire preference structure. The converse, however, is not true: a preference structure does <u>not uniquely</u> specify a value function. Suppose v_1 is a value function consistent with a given preference structure.Then if $T(\cdot)$ is any strictly monotonically increasing real-valued function (of a real variable), as depicted in Fig. 3.8 and if we now define $v_2(\underline{x}) \equiv T[v_1(\underline{x})]$, then it is immaterial whether we choose $\underline{a} \varepsilon A$ to maximize v_1 or v_2.

<u>Definition</u>. Given T as defined, we shall say that the value functions v_1 and $v_2 \equiv T(v_1)$ are <u>strategically equivalent</u>

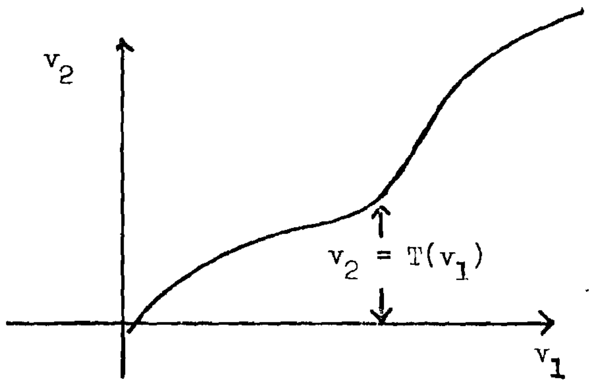


Figure 3.8

A Monotonically Increasing Real-Valued Function T Relating Two Strategically Equivalent Value Functions v_1 and v_2

for decision making purposes and write this as $v_1 \sim v_2$.

If, for example, all x_i are positive and

$$v_1(\underline{x}) = \sum_i k_i x_i \quad , \quad k_i > 0, \text{ all } i \quad ,$$

then

$$v_2(\underline{x}) = \sqrt{\sum_i k_i x_i}$$

and

$$v_3(\underline{x}) = \log\left(\sum_i k_i x_i\right)$$

would be strategically equivalent to v_1. All three functions are representations of the same preference structure. Indeed, for operational purposes, given v we will want to choose T such that the value function T(v) is easy to manipulate mathematically.

3.4 PREFERENCE STRUCTURES AND VALUE FUNCTIONS FOR TWO ATTRIBUTES

For notational convenience we shall label the two attributes by X and Y instead of X_1 and X_2. We repeat that X and Y shall each be assumed to be positively oriented: the more of any component the better for any fixed level of the other component.

3.4.1 The Marginal Rate of Substitution

Suppose you are given a concrete problem where X and Y are specified desirable attributes and suppose you are asked: If Y is increased by Δ units, how much does X have to decrease in order for you to remain indifferent?

Clearly in many instances your answer depends on the levels x of X and y of Y. If at the point (x_1,y_1) you are willing to give up $\lambda\Delta$ units of X for Δ units of Y, then we will say that the marginal rate of substitution of X for Y at (x_1,y_1) is λ. In other words, λ is roughly the amount of X you are just willing to "pay" for a unit of Y, given that you presently have x_1 of X and y_1 of Y. Figure 3.9 depicts this case. Strictly speaking we should take the limit as Δ approaches 0. Throughout we assume that we are in a well-behaved world where all functions have smooth second derivatives.

The marginal rate of substitution at (x_1,y_1)--as we are using it--is the negative reciprocal of the slope of the indifference curve at (x_1,y_1). Thus, if we have indifference curves, then we can calculate local substitution rates.* In this section we will develop some methods for doing the reverse: that is, we will think about how marginal rates of substitution can help us construct indifference

---

*Mathematical Digression: If the indifference curve through (x_1,y_1) is given by

$$v(x,y) = c \ ,$$

then the marginal rate of substitution λ at (x_1,y_1) can be obtained from the following formula:

$$\lambda = -\frac{dx}{dy}\bigg|_{x_1,y_1} = \frac{v'_y(x_1,y_1)}{v'_x(x_1,y_1)}$$

where v'_x and v'_y are the partial derivatives of v with respect to the first and second arguments respectively.

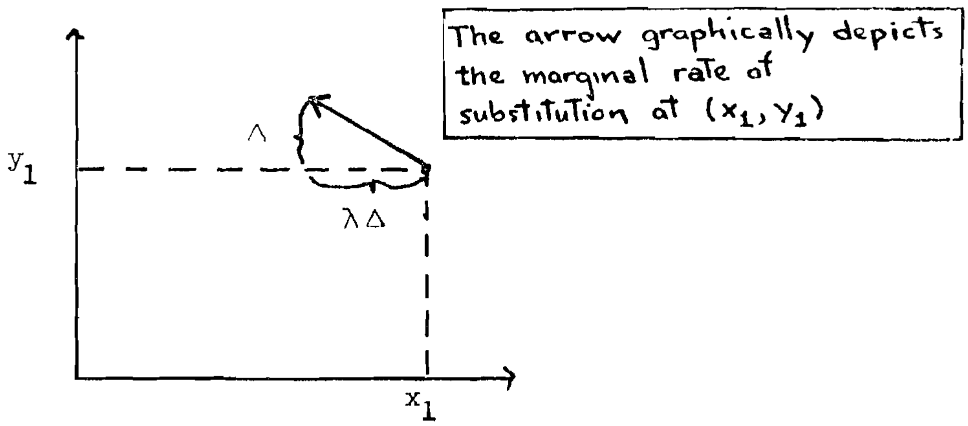


Figure 3.9

The Marginal Rate of Substitution of X for Y at (x_1, y_1) is λ

curves*.

3.4.2 The General Case

We shall now investigate how marginal rates of substitution might depend on the levels of X and Y, that is, on (x_1, y_1). A straightforward procedure is first to hold x_1 fixed and look at the substitution rates as a function of y_1, and second to hold y_1 fixed and look at the substitution rates as a function of x_1.

A typical case would be the following. Suppose the substitution rate at (x_1, y_1), the point a in Figure 3.10 is λ_a. If we hold x_1 fixed, we might find that the substitution rates increase with a decrease in Y and decrease with an increase in Y. This is illustrated at points b and c in Figure 3.10. The changes in the substitution rates mean that the more of Y we have, the less of X we would be willing to give up to gain a given additional amount of Y. In Figure 3.10 we can see that for the same increase in Y, the sacrifice of X is less at c than at b.

Similarly, if we hold y_1 fixed, we might find that the substitution rates decrease with decreasing X and increase with increasing X. This is illustrated at points

---

*MacCrimmon and Toda [1969] introduce a procedure for determining indifference curves and present experimental results. An interactive computer program for utilizing the procedure and related experience with its use are found in MacCrimmon and Siu [1974]. See also Toda [1974].

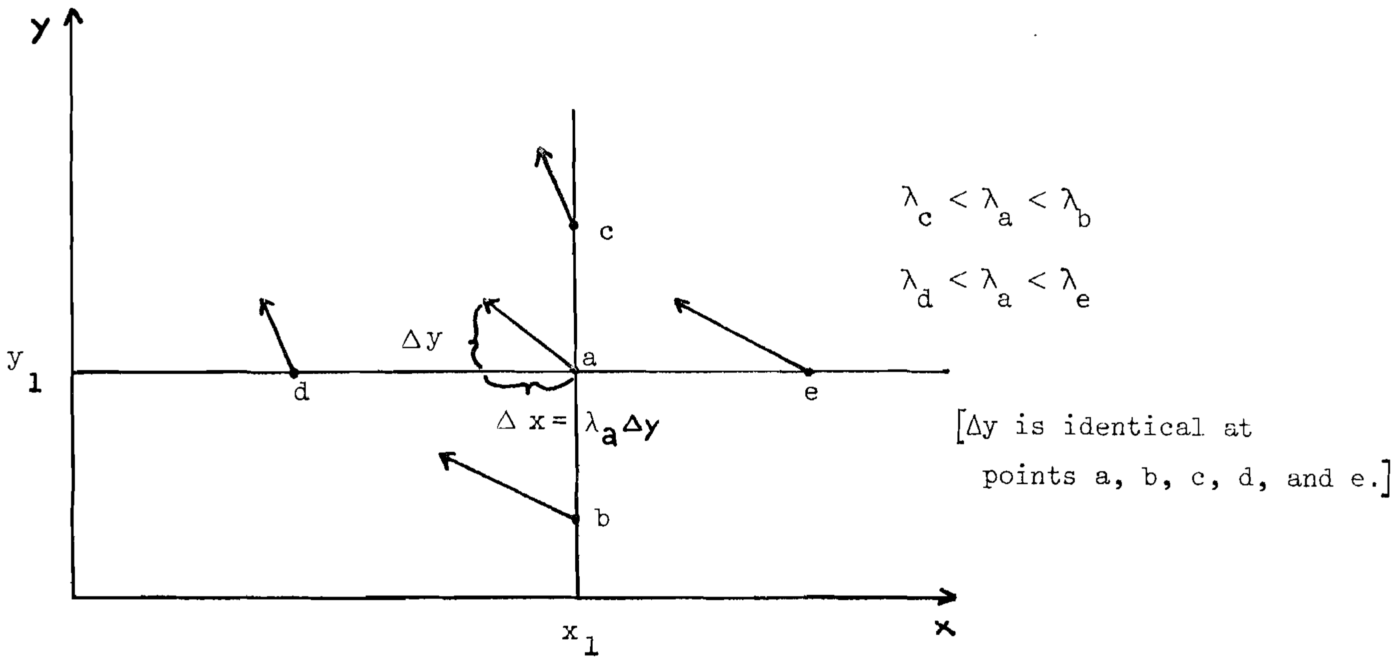


Figure 3.10

The Marginal Rate of Substitution as a Function of X and Y d and e in Figure 3.10. The interpretation is that additional units of X become less important relative to Y the higher the x value, and that we are therefore willing to substitute more X per additional unit of Y. This behavior is consistent with indifference curves of the shape given in Figure 3.6.

In many applications it is convenient to let X stand for monetary consequences. Now in this case, if $(x',y') \sim (x'',y'')$, then we can say that the decision maker is <u>just willing to pay</u> an amount $x'' - x'$ for a change of Y from y' to y'', when the monetary change taken place from the base of x'. If h is some positive amount it definitely does <u>not</u> follow that, in general,

$$[(x',y') \sim (x'',y'')] \quad \text{implies} \quad [(x' + h,y') \sim (x'' + h,y'')] \;.$$

That is, the amount the decision maker is just-willing-to-pay for a change from y' to y'' will depend on the monetary base he is starting from. It generally is not possible to "price-out" a change from y' to y'' without specifying the absolute level of X. The next two subsections consider those special cases where changes in Y can be "priced-out" independently of the X starting position. A more general discussion of "willingness-to-pay" arguments is found in Section 3.8.

3.4.3 <u>Constant Substitution: Case of Linear Indifference Curves</u>

An extreme special case of substitution rates occurs when the substitution rate at (x_1, y_1) does not depend on the values x_1 and y_1. That is, the marginal substitution rate is also the global substitution rate, applicable at any point and to substitutions in any amounts. In this case, the indifference curves are of the form

$$x + \lambda y = \text{constant} , \qquad (3.9)$$

and a suitable value function for this preference structure is

$$v(x,y) = x + \lambda y . \qquad (3.10)$$

Since in this case the local substitution rate is the global substitution rate, when assessing λ, the analyst does not have to ask localized questions involving small changes in x and y. The decision maker can base his assessment of λ on sizable, psychologically meaningful changes in x and y.

Sometimes a decision maker may be of the opinion that for his problem the substitution rates should be constant, but he may have difficulty assigning a value to λ. In practice it may not be necessary to determine λ exactly. For example, in a problem involving the choice of one of several actions, the decision maker might calculate λ-intervals, such that action a_1 is best if

$$\lambda \leq \lambda_1 ,$$

action a_2 is best if

$$\lambda_1 < \lambda \leq \lambda_2 ,$$

and so on. Figure 3.11 illustrates such intervals.

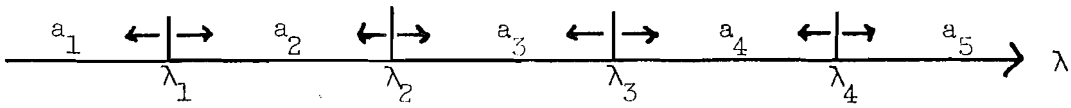


Figure 3.11

The Optimal Act as a Function of the Substitution Rate λ

In some problem, it may be clear that, although the exact value of λ is unknown, λ falls in the interval (λ_2,λ_3) and thus a_3 is best. If λ is close to λ_2 it may not be clear whether λ is greater than or less than λ_2, and thus whether a_2 or a_3 should be chosen. But in this case, a_2 and a_3 are almost at a standoff, so it may not be necessary to worry too much about which one is chosen and certainly a_1, a_4, and a_5 can be eliminated from consideration.

3.4.4 Constant Substitution Rates with a Transformed Variable

Suppose that the marginal rate of substitution λ at (x_1,y_1) depends on y_1 but not on x_1. That is, suppose that the amount the decision maker is willing to pay in X units for additional Y units depends on the level of Y but not on the level of X. (Even if this supposition does not hold exactly, it may hold approximately for x-values in a given range of concern and a convenient "lie" may not be inappropriate.) Four typical substitution rates for this case are illustrated in Figure 3.12.

An example of the kind of composite value function that produces this pattern of local substitution rates is

$$v(x,y) = x + v_Y(y) \ , \qquad (3.11)$$

where we use the symbol $v_Y(.)$ to indicate a function of single variable y.

A major question is the following. If a decision maker feels that substitution rates depend on y but not on x, how

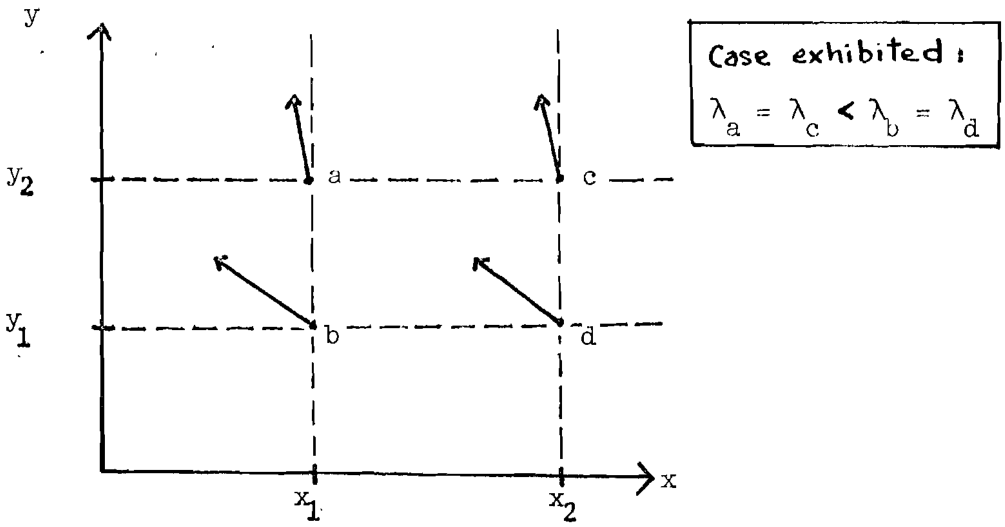


Figure 3.12

Substitution Rates Depending on y but not on x can this qualitative requirement help in the assessment of an appropriate v function? We now show in this case that v may be expressed as in (3.11).

If you are at (x_1,y_1), how much should you be willing to pay in X-units to increase Y from y_1 to y_2? To answer this question, let the marginal rate of substitution (x,y) be denoted by $\lambda(y)$, which shows the dependence of λ on y but not on x. As a first order approximation, for a small Δ increment in Y, you should be just willing to spend $\lambda(y)\cdot\Delta$ in X-units. Hence to go from y_1 to y_2 you should be just willing to pay in X units the amount

$$\int_{y_1}^{y_2} \lambda(y)\, dy \quad .$$

Let y_o be the minimum value of Y that is of concern in our problem. Define the function

$$v_Y(y') = \int_{y_o}^{y'} \lambda(y)\, dy \; . \tag{3.12}$$

The function v_Y can be thought of as the global substitution function between Y and X. In terms of the v_Y function, the decision maker is indifferent between

$$(x_1,y_1) \quad \text{and} \quad (x_1 - \left[v_Y(y_2) - v_Y(y_1)\right], y_2) \quad .$$

This is to say that an increase from y_1 to y_2 is worth $v_Y(y_2) - v_Y(y_1)$ in X units.

We have just informally argued an important result.

Theorem 3.1. The marginal rate of substitution between X and Y depends on y and not on x if and only if there is a value function v of the form

$$v(x,y) = x + v_Y(y) , \qquad (3.13)$$

where v_Y is a value function over attribute Y. Pruzan and Jackson (1963) offer a slightly different presentation of this same result.

<u>Assessment of v_Y</u>: The measurement problem associated with (3.13) boils down to an appropriate assessment of v_Y. It is usually difficult for subjects to give meaningful quantitative responses for small changes in attribute levels. Thus, in most circumstances, the analyst should <u>not</u> assess v_Y by first assessing $\lambda(y)$ and then using (3.12). Rather he should get at v_Y another way, and if he then wants to find $\lambda(y)$, he can invert (3.12) to calculate

$$\lambda(y) = \frac{d}{dy} v_Y(y) . \qquad (3.14)$$

One way to obtain v_Y is as follows: Arbitrarily set $v_Y(y_0) = 0$. With this choice of origin we can now interpret $v_Y(y)$ as the amount (in X units) the decision maker is just-willing-to-pay to go from y_0 to y. Thus the analyst can, in principle, obtain direct assessments of v_Y at selected points $y_1, y_2 \ldots$ and "fair in" a curve. The analyst might be well-advised, however, first to attempt to learn more about the qualitative structure of v_Y before getting involved in quantitative details. For example, it will often be the case that the decision maker would be willing to pay less and less for a positive, fixed change of Δ units in Y as the value of y increases. In other words he might feel that

$$v_Y(y+\Delta) - v_Y(y) < v_Y(y) - v_Y(y-\Delta) \quad \text{all } y, \Delta > 0 \; ; \qquad (3.15)$$

it is worth less to go from y to $y + \Delta$ than from $y - \Delta$ to y, regardless of the value of y or Δ (positive). A qualitative determination of the appropriateness of (3.15) implies that v_Y is strictly concave--i.e., it exhibits, in the vernacular of classical economics--a decreasing marginal evaluation. (Notice that we shun the expression decreasing marginal "utility" because we choose to use the term "utility" in a more precise fashion. See Section 4.4.) If the analyst learns that an appropriate shape for v_Y is concave, as is shown in Figure 3.13, then he can draw v_Y reasonably accurately if he ascertains numerical values for just a few points.

In order not to leave the impression that v_Y is necessarily concave, let us consider another common type of qualitative structure for v_Y. Imagine that the decision maker feels that there is some small interval about a level y_1, say, where things go "critical." Going from $y_1 - \Delta$ to $y_1 + \Delta$ might be much more important than going from $y_1 + \Delta$ to $y_1 + 3\Delta$ or going from $y_1 - 3\Delta$ to $y_1 - \Delta$. By qualitative probing the analyst might ascertain that this decision maker's v_Y curve is shaped somewhat like that depicted in Figure 3.14.

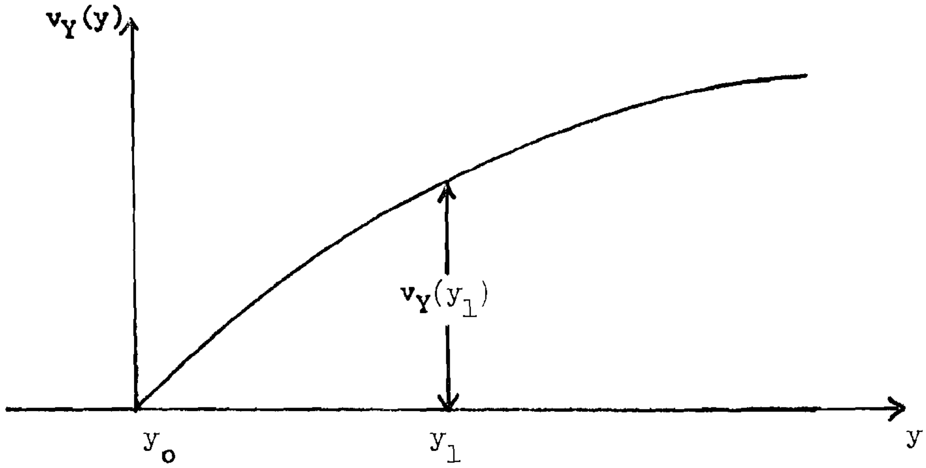


Figure 3.13

A Concave Value Function for Attribute Y

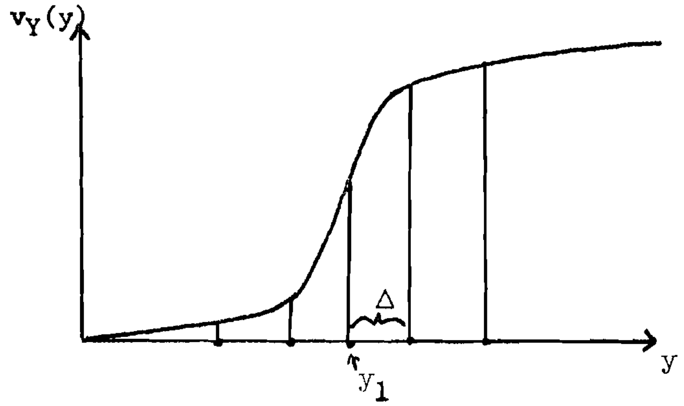


Figure 3.14

A Value Function That is Not Concave

Change of scale for linearization. If the marginal rate of substitution depends on y but not on x, then the indifference curves will be horizontal translates of each other. One indifference curve can generate the other just by sliding it horizontally as shown in Figure 3.15A. The indifference curves can be "straightened out" by change of the Y-variable to a Z-variable by means of the function v_Y. Thus, if we define

$$z = v_Y(y) \ , \tag{3.16}$$

then the point (x,y) in Fig. 3.15A becomes (x,z) in Fig. 3.15B where z and y are related by (3.16). The indifference curve C in Fig. 3.15A gets transformed into the straight line L with slope -1 in Figure 3.15B.

In the transformed coordinates x and z, the indifference curves are parallel straight lines. There is not a constant substitution rate between X and Y but there is a constant substitution rate (of 1) between X and Z where $z = v_Y(y)$. In the (x,z) evaluation space an appropriate value function is

$$v(x,z) = x + z \ . \tag{3.17}$$

3.4.5 The Corresponding Tradeoffs Condition: An Additive Value Function

In general the marginal rate of substitution at (x_1,y_1) depends on the level of x_1 and on the level of y_1. It may be, however, that we can transform the X-scale into

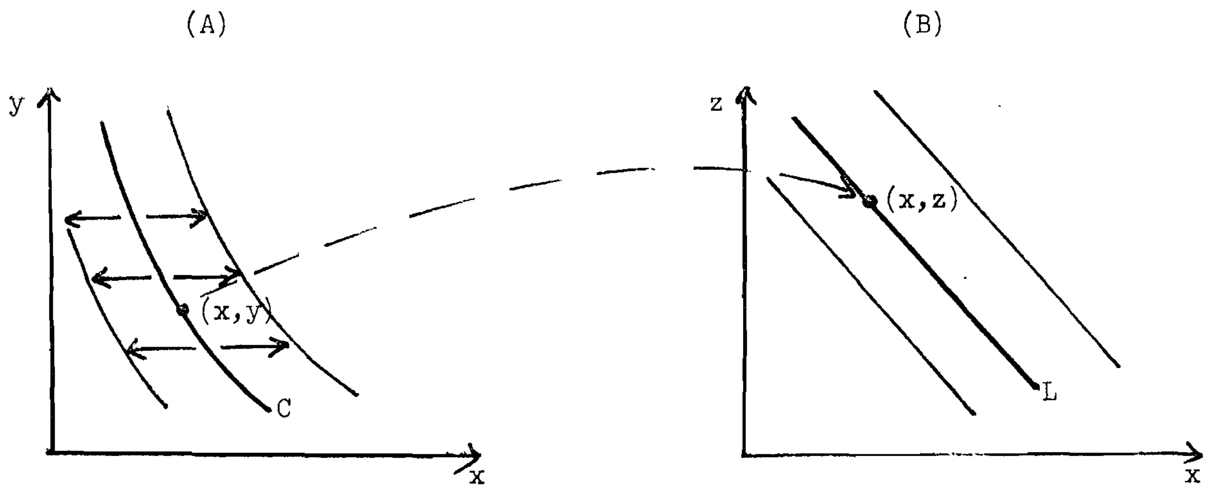


Figure 3.15

A Change of Scale to Linearize Indifference Curves a W-scale and the Y-scale into a Z-scale such that the substitution rate at (w_1, z_1) would not depend on the level of w_1 or z_1. Then we would have the constant substitution rate case discussed in subsection 3.4.3.

An Additive Value Function. Consider four points A:(x_1, y_1), B:(x_1, y_2), C:(x_2, y_1), and D:(x_2, y_2) as shown in Fig. 3.16. Suppose the following holds:

1. At (x_1, y_1) an increase of b in Y is worth a payment of a in X;
2. At (x_1, y_2) an increase of c in Y is worth a payment of a in X;
3. At (x_2, y_1) an increase of b in Y is worth a payment of d in X.

The question is: at (x_2, y_2) an increase of c in Y is worth what payment in X? If the answer is that it is worth a payment of d in X--that is, in Figure 3.16 the question mark (?) is answered "d", and if this holds regardless of the values of x_1, x_2, y_1, y_2, a, b, c, and d, then we will say that the Corresponding Tradeoffs Condition is satisfied. This test provides us with necessary and sufficient conditions for an important result. But first let us define the concept of additivity which will simplify the statement of the next result.

Definition: A preference structure is additive if there exists some value function reflecting that preference structure that can be expressed by

$$v(x,y) = v_X(x) + v_Y(y) .$$

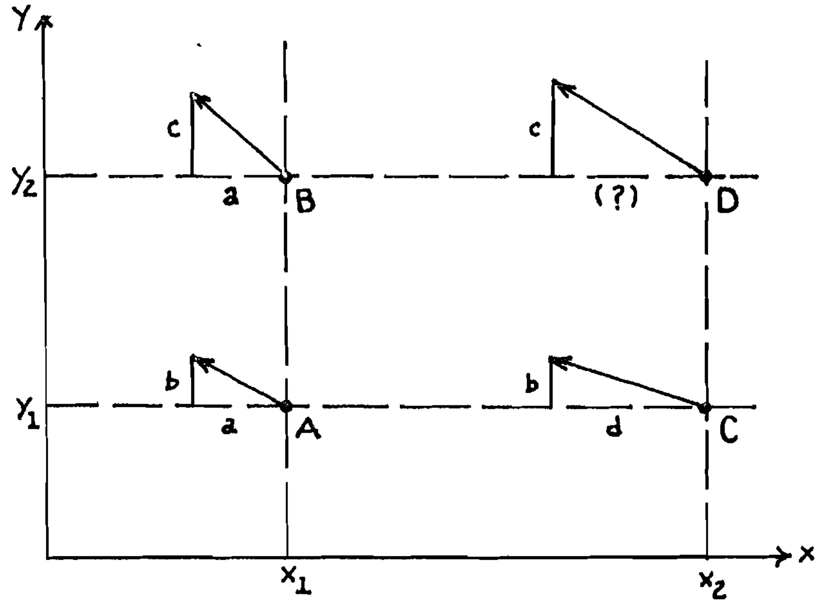


Figure 3.16

Substitution Rates Consistent with An Additive Value Function when ? Equals Amount

If a given preference structure, for example, has a value function

$$v_1(x,y) = (x - \alpha_1)^{\alpha_2} (y - \beta_1)^{\beta_2}$$

then that preference function would be additive since then

$$\log v_1(x,y) = \alpha_2 \log (x - \alpha_1) + \beta_2 \log (y - \beta_1)$$

and an additive **v** can be defined as $\log v_1$.

Theorem 3.2 A preference structure is additive and therefore has an associated value function of the form

$$v(x,y) = v_X(x) + v_Y(y) \quad , \qquad (3.18)$$

where v_X and v_Y are value functions if and only if the Corresponding Tradeoffs Condition is satisfied.

Clearly, given the additive value function (3.18), the Corresponding Tradeoffs Condition is met. However, the converse, proven by Luce and Tukey [1964], is much more difficult to show. In the next subsection, the conjoint scaling procedure used to illustrate the assessment of the additive value function also demonstrates informally the validity of Theorem 3.2. A formal proof is not given here.

3.4.6 Conjoint Scaling: The Lock-Step Procedure

Suppose that the Corresponding Tradeoffs Condition is met implying the existence of v_X and v_Y. How might we go about finding them? One procedure we might adopt is the following.

Let x_o and y_o be the lowest values of X and Y under consideration.

1. Define

   $$v(x_o,y_o) = v_X(x_o) = v_Y(y_o) = 0 . \qquad (3.19)$$

   This sets up the origin of measurement.

2. Choose $x_1 > x_o$ and arbitrarily set $v_X(x_1) = 1$.
   This sets up the unit of measurement.

3. Ask the decision maker to give a value of Y, say y_1, such that

   $$(x_1,y_o) \sim (x_o,y_1) \quad ,$$

   where $\sim$ stands for "is indifferent to". Define $v_Y(y_1) = 1$.

4. Ask the decision maker to give a value of X, say x_2, and a value of Y, say y_2, such that

   $$(x_2,y_o) \sim (x_1,y_1) \sim (x_o,y_2) \quad .$$

   Define

   $$v_X(x_2) = v_Y(y_2) = 2 .$$

5. A necessary condition for this scaling procedure to work is that

   $$(x_1,y_2) \sim (x_2,y_1) \quad .$$

   But as is easily seen from Fig. 3.17 this condition holds if the Corresponding Tradeoffs Condition works. Compare Fig. 3.17 with Fig. 3.16 and identify points labelled A,B,C, and D in each. In Fig. 3.17, the

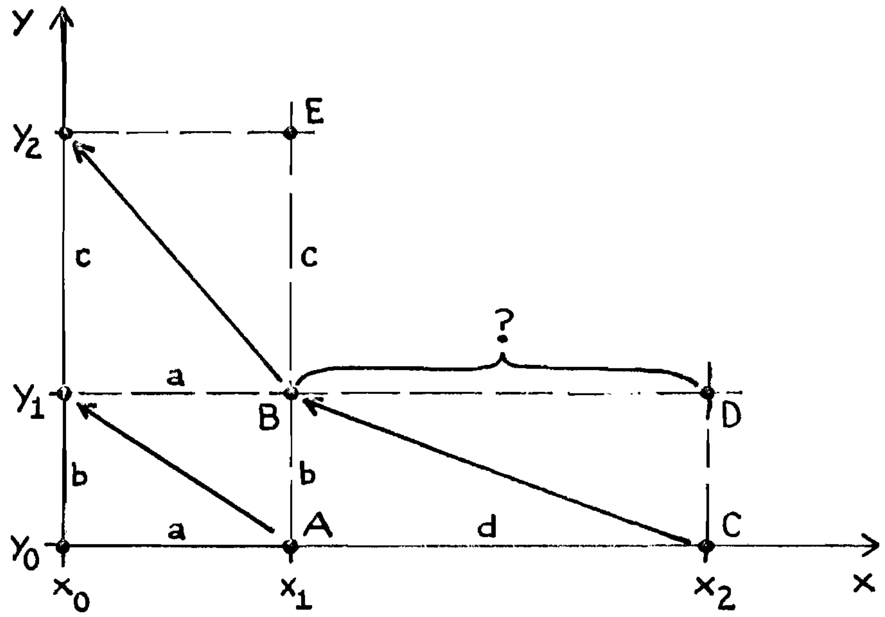


Figure 3.17. Conjoint Scaling With An Additive Preference Structure

Corresponding Tradeoffs Condition implies that the distance in X-units from B to C must be d and hence points C and E are indifferent.

6. Assuming step 5 is passed, ask the decision maker to choose (x_3, y_3) such that

   $$(x_3, y_o) \sim (x_2, y_1) \sim (x_1, y_2) \sim (x_o, y_3) \quad .$$

   Define

   $$v_X(x_3) = v_Y(y_3) = 3 \ .$$

7. As in step 5 above, a necessary condition for this scaling procedure to work is that

   $$(x_3, y_1) \sim (x_2, y_2) \sim (x_1, y_3) \quad .$$

   You might want to check that the above is implied by the Corresponding Tradeoffs Condition.

8. Continue in the same manner as above.

9. Plot these few points, as in Figure 3.18, fair in smooth v_X and v_Y curves and agree tentatively to let

   $$v(x,y) = v_X(x) + v_Y(y) \ .$$

10. As a precautionary measure check a few pairs of points for "reasonableness." To this end let us define x_k and y_k such that

    $$v_X(x_k) = v_Y(y_k) = k \ .$$

    Now we can check, for example, if

    $$(x_1, y_o) \sim (x_{.5}, y_{.5}) \quad .$$

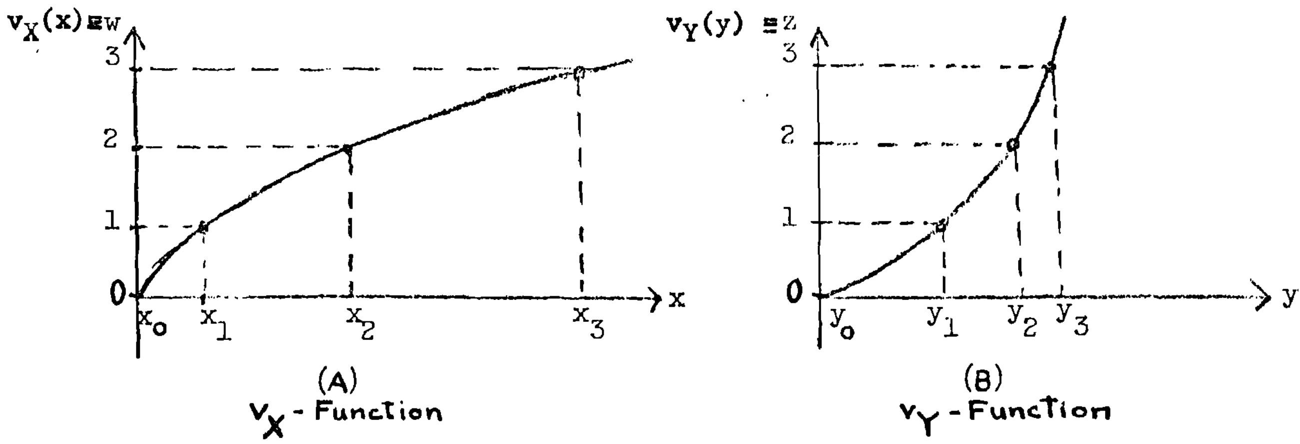


Figure 3.18. Consistently Scaled Value Functions for Attributes X and Y

If not, you might alter the points $(x_{.5}, 0.5)$ and $(y_{.5}, 0.5)$ on the v_X and v_Y curves.

Notice how the v_X and v_Y functions are intrinsically interwined. We cannot interpret completely one without the other.

The above method of generating v_X and v_Y constitutes a constructive heuristic (almost) proof showing that the validity of the Corresponding Tradeoffs Condition implies the existence of an additive preference structure. The construction was only demonstrated on a grid of points and one would need to subdivide the intervals (say by a "halving technique") and sprinkle in some continuity somewhere to complete the proof. Note also the implicit use of a "solvability condition" which is not formally stated: We selected, for example, x_0, y_0, and x_1 and then glibly assumed the existence of y_1 that solved the indifference equation

$$(x_0, y_1) \sim (x_1, y_0) \quad .$$

Similarly we obtained x_2 and y_2 as solutions to indifference equations.

3.4.7 An Alternative Conjoint Scaling Procedure: The Mid-Value Splitting Technique

Two preliminary definitions will facilitate the presentation of an alternate procedure for assessing v_X and v_Y. Assume the Corresponding Tradeoffs Condition is valid.

<u>Definition:</u> The pair (x_a,x_b) is said to be <u>differentially value-equivalent</u> to the pair (x_c,x_d)--where $x_a < x_b$ and $x_c < x_d$--if whenever one is just willing to go from x_b to x_a for a given increase of Y, then one would be just willing to go from x_d to x_c for the same increase in Y. Or stated in another manner, if at any point y' of Y one is willing to "pay" the same amount of Y for the increase of X from x_a to x_b as for the increase from x_c to x_d, then (x_a,x_b) is differentially value-equivalent to (x_c,x_d).

<u>Definition:</u> For any interval $[x_a,x_b]$ of X its <u>mid-value</u> point x_c is such that the pairs (x_a,x_c) and (x_c,x_b) are differentially value-equivalent.

Observe two things about this definition. First: in order to define a mid-point x_c of $[x_a,x_b]$ we exploited the existence of a second attribute Y. Second: if the decision maker, starting at y' is willing to give up in Y the same amount to go from x_a to x_c as from x_c to x_b, then the same condition ($c=c'$ in Fig. 3.19) must prevail starting at any other level y" <u>provided</u> the Corresponding Tradeoffs Condition holds. The argument can be seen readily from Fig. 3.19. We label points A,B,C,D to help the reader make the necessary correspondences with Fig. 3.16.

Let the range of X be $x_0 \leq x \leq x_1$, of Y be $y_0 \leq y \leq y_1$, and assume that the Corresponding Tradeoffs Condition is passed*. We now seek a value function v that can be expressed

*In this subsection the subscripts on the symbols x and y are used differently than they were used in the previous subsection. We also now assume that v is bounded.

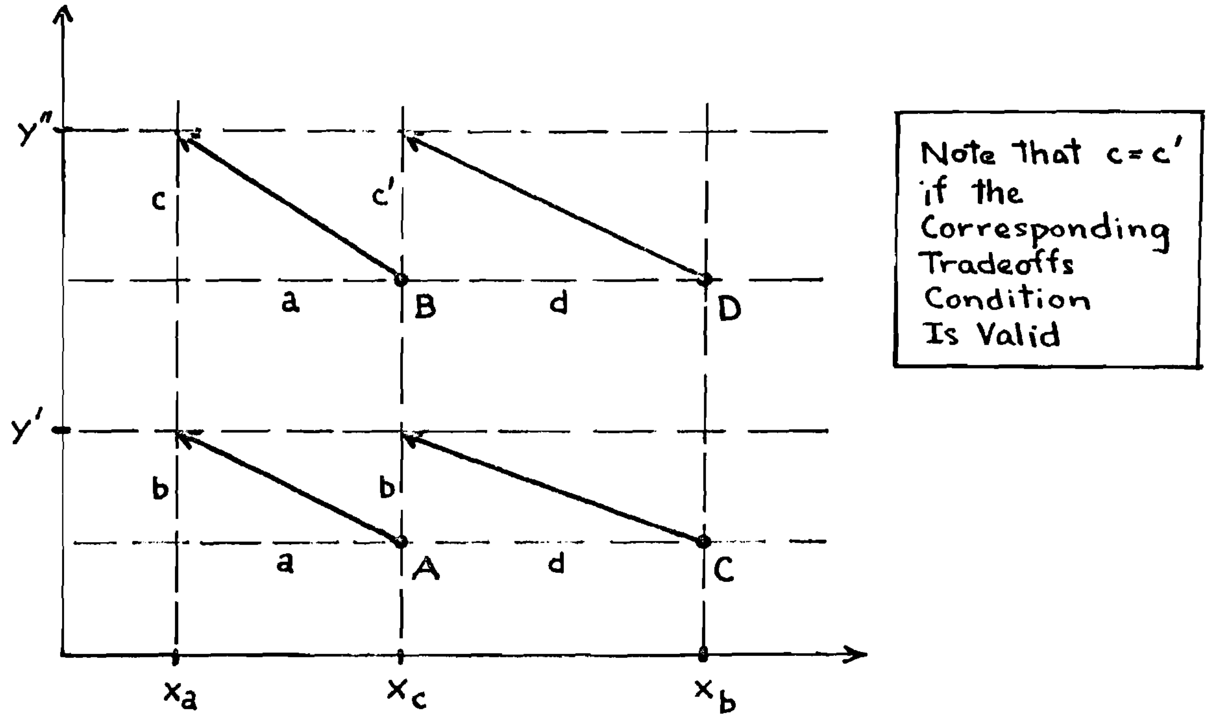


Figure 3.19. The Existence of the Mid-Value Point x_c Requires the Corresponding Tradeoffs Condition

in the form

$$v(x,y) = \lambda_1 v_X^* (x) + \lambda_2 v_Y^* (y) , \qquad (3.20)$$

where

$$\text{a. } v_X^* (x_o) = 0 \text{ and } v_X^* (x_1) = 1 , \qquad (3.21a)$$

$$\text{b. } v_Y^* (y_o) = 0 \text{ and } v_Y^* (y_1) = 1 , \qquad (3.21b)$$

$$\text{c. } \lambda_1 > 0, \lambda_2 > 0, \text{ and } \lambda_1 + \lambda_2 = 1 . \qquad (3.21c)$$

The assessment procedure is as follows:

<u>Procedure:</u>

a. Obtain v_X^* as follows: (1) Find the mid-value point of $[x_o, x_1]$; call it $x_{.5}$ and let $v_X^*(x_{.5}) = .5$. (2) Find the mid-value point, $x_{.75}$, of $[x_{.5}, x_1]$ and let $v_X^*(x_{.75}) = .75$. (3) Find the mid-value point $x_{.25}$ of $[x_o, x_{.5}]$ and let $v_X^* (x_{.25}) = .25$. (4) As a consistency check, ascertain that $x_{.5}$ is the mid-value point of $[x_{.25}, x_{.75}]$; if not, juggle the entries to get consistency. (5) Fair in the v_X^* curve passing through points (x_k, k) for $k = 0, 1, .5, .75, .25$ and perhaps additional points obtained by a mid-value splitting technique.

b. Repeat the same process for v_Y^*.

c. <u>Finding the scale factors</u> λ_1 <u>and</u> λ_2: Choose any two (x,y) pairs that are indifferent, say (x',y') and (x'',y''). We then have

$$v(x',y') = v(x'',y'')$$

or

$$\lambda_1 v_X^*(x') + \lambda_2 v_Y^*(y') = \lambda_1 v_X^*(x'') + \lambda_2 v_Y^*(y'') .$$

Since $v_X^*(x')$, $v_Y^*(y')$, $v_X^*(x'')$ and $v_Y^*(y'')$ are now known numbers and since $\lambda_1 + \lambda_2 = 1$ we can solve for λ_1 and λ_2.

3.4.8 A Hypothetical Illustrated Assessment

In order to demonstrate the interaction process between an analyst and decision maker, we present below an imagined dialogue between an interrogator and a very cooperative repondent.

In the natural units of attributes X and Y, assume that X(a) ranges over the interval 7 to 92 and Y(a) ranges over the interval -9 to 8. So, for convenience let us choose $x_o = 0$, $x_1 = 100$, $y_o = -10$, $y_1 = 10$, which are consistent with the scaling conventions of (3.21).

Question	Hypothesized Answer
1. Suppose Y is at 0 and X at 20. If Y were decreased by 1 unit how much more X would you need to just offset it? Don't be exact, give a rough answer.	I would want to move to $x = 25$.
2. Keep Y at 0 and let X be at 60. How much would 1 unit of Y buy of X at this point? Again, all I want is a rough answer.	Say $x = 70$.

Question	Hypothesized Answer
3. All right. At $y = 0$ it would cost roughly 1 unit of Y to push you from $x = 20$ to 25 and from 60 to 70. Is that right?	Yes, roughly.
4. O.K. Now think hard about this one. At another value of Y, say at $y = 5$ would you pay the same amount to go from $x = 20$ to 25 as from 60 to 70?	What's hard about that? I already said I would pay the same for the change 20 to 25 as for the change 60 to 70. But the absolute amount of Y I would pay would depend on the level of Y I'm at. I might pay 1 unit of Y at $y = 0$ and 3 units of Y at $y = 5$. Is that O.K.?

5. Sure. That's reasonable.

(At this point in the conversation, the interrogator might presume that the Corresponding Tradeoffs Condition is satisfied even though, strictly speaking, he must be sure that the same type of response would be forthcoming for more general values of X and Y. Also at this point of the dialogue the interrogator might query the respondent about concavity or convexity of the functions v_X and v_Y. This is omitted for the sake of of brevity. The interrogator next proceeds to describe the mid-value point of any interval.)

Question	Hypothesized Answer
6. Suppose you're at $y = 0$. Would you pay more of Y to change X from 0 to 50 or 50 to 100?	I would pay more to go from 0 to 50.
7. More to go from 0 to 10 or 10 to 100?	More to go from 10 to 100.
8. Give me a value, x' say, such that you would give up the same in Y to go from 0 to x' as from x' to 100.	About $x' = 20$.
9. In our vernacular then, 20 is the mid-value point between 0 and 100. We label 20 by $x_{.5}$. What is your mid-value point between 20 and 100?	Let's say 45. I'd pay the same to go from 20 to 45 as 45 to 100.
10. In that case $x_{.75} = 45$. What is your mid-value point between 0 and 20?	Oh, about 7.
11. Fine. This means that $x_{.25} = 7$. Does 20 seem like a good mid-value between 7 and 45?	Sure.
12. Now let's turn to the Y value. What is the mid-value point between -10 and 10.	Say, -2.
13. The mid-value between -2 and 10.	Say, 3.

Question	Hypothesized Answer
14. The mid-value between -10 and -2.	-7.

(The analyst now plots these few points as shown in Figure 3.20 and fairs in v_X and v_Y curves.)

15. I have to trouble you for a couple of more questions. Which (x,y) pair would you prefer (0,10) or (100,-10)? In other words if you were at (0,-10) would you rather push Y up to its limit of 10 or X up to its limit of 100?	The X varaible is more critical. I would rather have (100,-10) than (0,10).

(This answer implies that $\lambda_1 > \lambda_2$.)

16. O.K. then. Give me a value x such that you are indifferent between (x,-10) and (0,10). In other words, I'm asking you to consider the following. Imagine that you're at (0,-10). How much would you have to push X up to be equivalent to Y going from -10 to 10.	I don't know. I would say about 60. But I feel awfully woozy about that.

(The analyst draws the Figure 3.21)

If we assume that (60,-10) is indifferent to (0,10) then we have

$$v(60,-10) = v(0,10) \quad ,$$

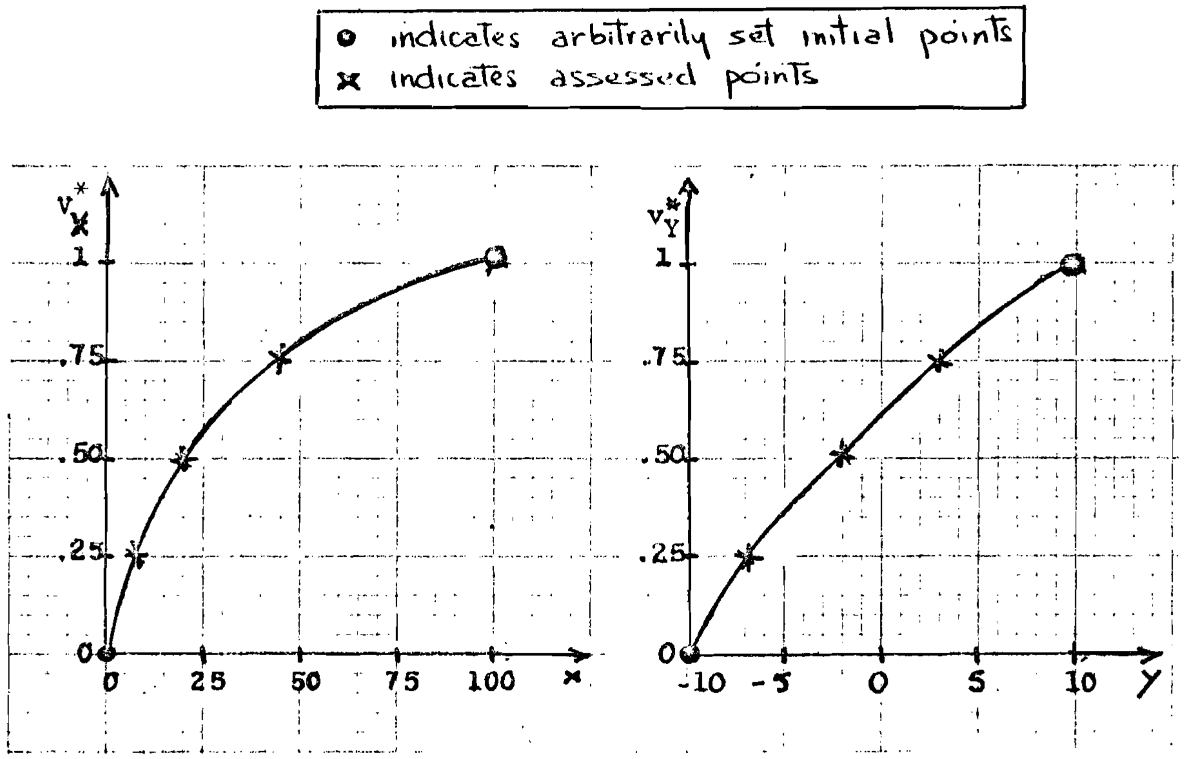


Figure 3.20. The Assessed Value Functions for Attributes X and Y

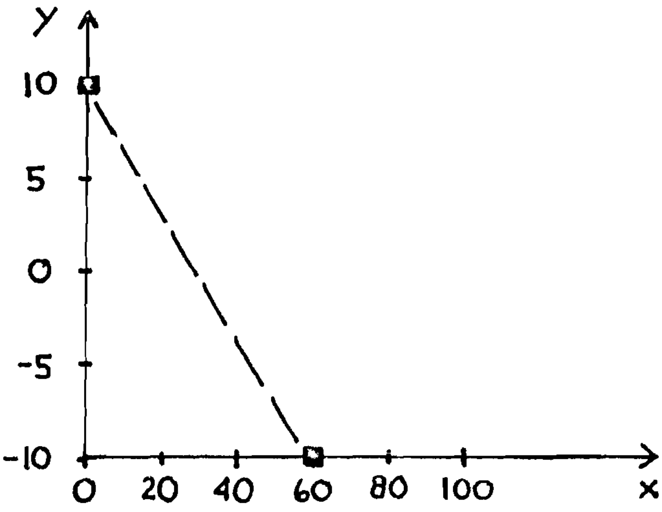


Figure 3.21. Specifying The Value Tradeoffs Between X and Y

or

$$\lambda_1 v_X^*(60) + \lambda_2 v_Y^*(-10) = \lambda_1 v_X^*(0) + \lambda_2 v_Y^*(10) .$$

Since

$$v_X^*(60) = .85,\ v_Y^*(-10) = v_X^*(0) = 0, \text{and } v_Y^*(10) = 1,$$

this implies

$$85\lambda_1 = \lambda_2 ;$$

and since $\lambda_2 = 1 - \lambda_1$, we have

$$\lambda_1 = 1/1.85 = .54 \quad \text{and} \quad \lambda_2 = .46 .$$

Or perhaps we should say:" λ_1 is a woozy .54."

We could think of this procedure as a first approximation to a suitable value function v. One should now look at a few pairs that have the same v-values and ask the decision maker if he would consider these pairs to be roughly indifferent. In other words we still might want to do a "fine-tuning" of the v_X^* and v_Y^* curves and of the λ_1, λ_2 values. Furthermore, if the λ_1 value (remember $\lambda_2 = 1 - \lambda_1$) were deemed the "weakest link of the chain", then it might be appropriate to do sensitivity or breakeven analyses with respect to the λ_1 values. It is important to reflect that it would not be possible to run such sensitivity studies on λ_1 without the preliminary structuring of the problem. This is often the case: <u>in order to run sensitivity studies for certain critical variables, one often has to structure the less sensitive part of the problem in a precise manner.</u>

3.4.9. <u>Some Words of Advice</u>

If the decision maker has hard-formed judgments, it may often be the case in practice that a value function cannot be found of the form

$$v(x,y) = v_X(x) + v_Y(y) \ .$$

Nonetheless, such a value function may hold approximately. In other cases, it may be important for ease of analysis of explanation to concoct a value function of this form. The decision maker may begin the conjoint-scaling procedure and see along the way if the checking conditions are plausible.

3.5 THE CASE OF THREE ATTRIBUTES

We can straightforwardly generalize the results we obtained in Section 3.4 to the case of three evaluating criteria. Instead of the two evaluators X and Y we will consider the three evaluators X,Y, and Z. The evaluators map any act a in the action space into a point $[X(a), Y(a), Z(a)]$ in the three-dimensional consequence space.

3.5.1 Conditional Preferences

We will begin by considering a conditional preference structure in the (x,y) space given an assumed value of Z, say z'.
Definition: Consequence (x',y') is conditionally preferred to (x",y") given z' if and only if (x',y',z') is preferred to (x",y",z').
Conditional indifference is defined analogously and thus we can talk about conditional indifference curves in the (x,y) space given z'.

In general, the conditional preference structure for attributes X and Y given the value of the Z attribute is z' will depend on the value z'. For example, the marginal rate of substitution at some point (x_1,y_1) might depend on z'. In some cases, however, the conditional preference structure in the (x,y) space given z' may not depend on z'. We are thus led to the following definition:

Defintion: The pair of attributes X and Y is preferentially independent of Z if the conditional preferences in the (x,y) space given z' do not depend on z'.

Notice that if the pair {X,Y} is preferentially independent of Z, then the substitution rate between X and Y at the point (x_1,y_1) given z' does not depend on z', for all x_1,y_1, and z'. Thus, the set of indifference curves in X, Y space does not depend on z'. Furthermore, because of the preferential independence condition, these curves have the same preference ordering.

Suppose that the pair {X,Y} is preferentially independent of Z. In this case we can say that if

$$(x_1,y_1,z') \succsim (x_2,y_2,z') \quad ,$$

where the symbol $\succsim$ is read: "is preferred or indifferent to", then

$$(x_1,y_1,z) \succsim (x_2,y_2,z) \quad , \text{ for all } z \; .$$

The following two examples indicate some cases of possible preferential independence.

Suppose the three attributes of a proposed construction project are

Q = quality,

T = time-to-completion (negatively oriented) ,

C = cost (negatively oriented) .

In some circumstances the value tradeoffs between quality and time-to-completion may not depend on the cost of the project. In this case $\{Q,T\}$ would be preferentially independent of C. Also, we might find that given a quality level q', the preference structure in the (time, cost) subspace does not depend on the particular level of q'; in other words $\{T,C\}$ may be preferentially independent of Q. Similarly, $\{a,c\}$ may be preferentially independent of T. Whether or not any one of these preferential independent assertions would, in fact, be valid depends on the particular setting of the problem.

A second example concerns a proposed program with attributes

B_1 = benefit of type 1 ,

B_2 = benefit of type 2 ,

C = cost (negatively oriented) .

If the two types of benefits must be kept in balance, then $\{B_1,C\}$ would not be preferentially independent of B_2 and $\{B_2,C\}$ would not be preferentially independent of B_1. However, it might be plausible to expect that $\{B_1,B_2\}$ would be preferentially independent of C.

3.5.2 Reduction of Dimensionality

How can we exploit, in our measurement techniques, the fact that a particular decision maker may feel that $\{X,Y\}$ is preferentially independent of Z? In the next section we shall develop special techniques for the case where each pair of attributes is preferentially independent of the remaining attribute. But now let us assume that all we can justify is that $\{X,Y\}$ is preferentially independent of Z. Here is one way we might proceed.

Consider the conditional preference structure for X and Y, given some value z'. Observe that the particular value z' is really immaterial because of our hypothesis of preferential independence. We shall only consider the special case where each of the conditional indifference curves in the (x,y) space intersects some line $y = y'$ for a suitably chosen y'. We shall refer to y' as a base value for Y. (If no such y' exists, then the procedure we are about to describe will have to be modified a bit.) Now the indifference curve through a typical point (x,y) will intersect the line $y = y'$ at some value (x',y') as shown in Fig. 3.22. Observe that x' depends on the choice of y' and on the point (x,y). In order to emphasize this observation we write

$$x' = T(x,y;y') \quad . \tag{3.22}$$

Also notice that in terms of three space, we have

$$(x,y,z) \sim (x',y',z) \quad , \text{ for all } z \ . \tag{3.23}$$

Hence the preferential comparison of any two triplets

$$(x_1,y_1,z_1) \quad \text{versus} \quad (x_2,y_2,z_2)$$

can be transformed into the preferential comparison of

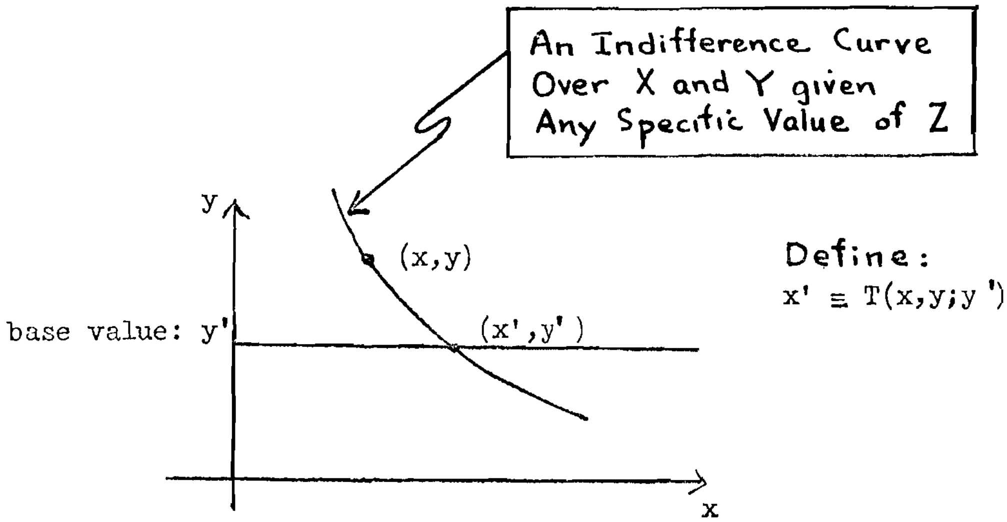


Figure 3.22. Exploiting Preferential Independence To Reduce The Dimensionality of a Problem

$$(x_1', y', z_1) \text{ versus } (x_2', y', z_2)$$

where

$$x_1' = T(x_1, y_1; y') \text{ and } x_2' = T(x_2, y_2; y') .$$

Thus our overall measurement task now reduces formally to a consideration of our conditional preference structure for $\{X,Z\}$ given the level of Y is y'. Instead of comparing

$$(x_1, y_1, z_1) \text{ and } (x_2, y_2, z_2)$$

in three-space, we now must make the conditional comparison of

$$(x_1', z_1) \text{ and } (x_2', z_2)$$

given y'. We have essentially used our hypothesis to reduce one three-dimensional comparison to two-dimensional comparisons.

<u>Some Words about the Transformation T.</u> Let the set of acts be labelled $A = \{a_1, \ldots, a_i, \ldots, a_n\}$. Once again assume that $\{X,Y\}$ is preferentially independent of Z. If n is small, then for each a_i it may not be outlandish to ask the decision maker directly for a value $X'(a_i)$ such that he is conditionally indifferent between

$$[X(a_i), Y(a_i)] \text{ and } [X'(a_i), y'] .$$

Answers to these n questions may be a lot easier to obtain then to get the full conditional preference structure in the (x,y)-plane.

If n is very large, this procedure is not operational. If, however, in the (x,y)-plane we can justify a value function v of the form

$$v(x,y) = v_X(x) + v_Y(y)$$

(see subsection 3.4.5), then $x' = T(x,y;y')$ will be such that

$$v_X(x') - v_X(x) = v_Y(y) - v_Y(y') \quad ,$$

and this may be a feasible operation to implement.

If n is large and no simple v function can be assumed, then we're in trouble; but still life is not hopeless. One might, for example, choose a reasonable number of points $(x_1,y_1),\ldots, (x_j,y_j),\ldots,(x_m,y_m)$, for say m = 10 or so, and by direct questioning get for each j a value x'_j when

$$(x'_j,y') \sim (x_j,y_j)$$

or equivalently where

$$x'_j = T(x_j,y_j;y') \quad .$$

By carefully investigating the dependence of x'_j on x_j and y_j(remember y' is fixed for all j), one might concoct a reasonable, simple compromise function T that fits the data reasonably well and can be used to extrapolate an x' value for any other (x,y) pair. We shall not even begin to enumerate the myriad of techniques that can be employed for this type of data-fitting procedure.

Of course, if {X,Y} is preferentially independent of Z, then instead of bringing each y to a base position y' and defining x' by (3.22) and (3.23), we could bring x to a base position x', say, and define y' to be such that

$$(x,y,z) \sim (x',y',z) \quad , \text{for all } z \quad .$$

This reduction would then be followed by a conditional preference analysis of Y and Z given x = x'. One must be imaginative in choosing the most convenient reduction procedure. There are still other possibilities. For example, suppose in a given context it is natural to expect y to be approximately a multiple h of x. In this case for any (x,y) pair we might choose a value x' such that

$$(x,y,z) \sim (x',hx',z) \quad , \quad \text{for all } z \quad .$$

This reduction would then be followed by a conditional preference analysis of X and Z given the understanding that y is not free but is always an h-multiple of x.

3.5.3 Mutual Preferential Independence and the Existence of an Additive Value Function*

If preferences for (x,y,z) triplets are consistent with a v-function of the additive form

$$v(x,y,z) = v_X(x) + v_Y(y) + v_Z(z) \quad ,$$

then clearly

a. $\{X,Y\}$ is preferentially independent of Z ,

b. $\{X,Z\}$ is preferentially independent of Y ,

c. $\{Y,Z\}$ is preferentially independent of X .

What is much more important, however, and quite surprising, is that the converse is true.

Theorem 3.3. A value function v may be expressed in an additive form

---

*It is assumed throughout this section that all three evaluators are essentially relevant--i.e., that the preference structure cannot be full described in terms of only two of the three evaluators.

$$v(x,y,z) = v_X(x) + v_Y(y) + v_Z(z) \quad , \qquad (3.24)$$

where v_X, v_Y, and v_Z are single attribute value functions, if and only if {X,Y} is preferentially independent of Z, {X,Z} is preferentially independent of Y, and {Y,Z} is preferentially independent* of X.

This result was first proven by Debreu [1960]. A sightly more general proof is found in Krantz et al. [1971]. Since formal proofs do appear in the literature, our discussion will avoid formalities and attempt merely to illustrate the plausibility of the result. Before proceeding, we should define an important term.

Definition. If each pair of attributes is preferentially independent of its complement, the attributes are pairwise preferentially independent.

Hence, in shorthand vernecular, Theorem 3.3 says that additivity coimplies pairwise preferential independence.

Something is truly remarkable about Theorem 3.3. Remember that in order to get an additive representation for two evaluators X and Y we had to impose the stringent Corresponding Tradeoffs Condition. Nothing of that sort is required here. If all we know is that {X,Y} is preferentially independent

---

*The condition that each pair of attributes must be preferentially independent of the remaining attribute will be weakened in the next subsection. Roughly, any two of the three preferential independence assumptions will be shown to imply the third.

of Z, then we cannot say that conditional preferences for X and Y will satisfy the Corresponding Tradeoffs Condition. But once we assume pairwise preferential independence, then the conditional preference strucutre for any pair of evaluators, given any level of the remaining evaluator, clears the Corresponding Tradeoffs hurdle. Without giving a formal proof of these assertions, let's see how these assertions can be made plausible.

Recall how we constructed the v_X and v_Y functions using the conjoint scaling technique for two evaluators. (See subsection 3.4.6) We first arbitrarily chose values x_0, y_0, and x_1. Then in succession we used the decision maker's preferences to successively generate y_1, x_2, and y_2. Up to that point no requirement was made of the Corresponding Tradeoffs Condition. The first place that this condition had to be invoked was to justify the indifference of (x_1,y_2) and (x_2,y_1). Now how does bringing in Z and imposing pairwise preferential independence avoid this condition? Well let's back up a bit and start the measurement process from the beginning for three evaluators.

1. First choose x_0, y_0, and z_0 and let

   $$v(x_0,y_0,z_0) = v_X(x_0) = v_Y(y_0) = v_Z(z_0) = 0 .$$

2. Next arbitrarily choose x_1 and define y_1 and z_1 such that

   $$(x_1,y_0,z_0) \sim (x_0,y_1,z_0) \sim (x_0,y_0,z_1) .$$

   Let

   $$v_X(x_1) = v_Y(y_1) = v_Z(z_1) = 1 .$$

3. Notice now, how mutual preferential independence works to allow us to conclude that $(x_1,y_1,z_o) \sim (x_1,y_o,z_1) \sim (x_o,y_1,z_1)$.

   For example, from step 2 we know that (x_1,y_o) and (x_o,y_1) are conditionally indifferent given z_o. Hence they must be conditionally indifferent given z_1, or $(x_1,y_o,z_1) \sim (x_o,y_1,z_1)$.

   Similarly from step 2 we know that $(x_1,z_o) \sim (x_o,z_1)$ given y_o, and hence from the preferential independence of {X,Z} from Y, it is true also given y_1. But this implies $(x_1,y_1,z_o) \sim (x_o,y_1,z_1)$.

4. Next define x_2,y_2, and z_2 such that $(x_2,y_o,z_o) \sim (x_o,y_2,z_o) \sim (x_o,y_o,z_2) \sim (x_1,y_1,z_o)$.

Now we are ready to discuss the crucial point which we referred to earlier: How do we know without a Corresponding Tradeoffs Condition that

$$(x_2,y_1,z_o) \sim (x_1,y_2,z_o) \quad ?$$

The trick is to show that

$$(x_2,y_1,z_o) \sim (x_1,y_1,z_1)$$

and

$$(x_1,y_2,z_o) \sim (x_1,y_1,z_1) \quad ,$$

and by transitivity of indifference we're home.We know that

$$(x_2,y_o,z_o) \sim (x_1,y_o,z_1) \quad ,$$

and since {X,Z} is preferentially independent of Y, we can freely change y_0 to y_1 for the above indifference relation. This shows

$$(x_2,y_1,z_0) \sim (x_1,y_1,z_1) \quad .$$

One completes the demonstration by showing in an analogous manner that

$$(x_1,y_2,z_0) \sim (x_1,y_1,z_1) \quad .$$

While the above argument is far from being a proof it should make the theorem seem much less mysterious — even transparent. But, of course, there is a big gap between heuristic plausibility and a formal proof.

3.5.4 Weakening the Additivity Assumptions

Our interest in results such as Theorem 3.3 is mainly to take a set of fundamental assumptions--in this case the preferential independence assumptions--about a decision maker's preferences and from these, ascertain a specific convenient mathematical expression consistent with these preferences. In any problem, we first try to check for the appropriateness of the conditions and then assess subjectively the decision maker's value function. Thus, it is important to reduce, if possible, the number of conditions implying a particular functional form for one's preferences.

This subsection discusses the following operationally useful result.

Theorem 3.4 If

a. $\{X,Y\}$ is preferentially independent of Z ,

b. $\{Y,Z\}$ is preferentially independent of X ,

then

c. $\{X,Z\}$ is preferentially independent of Y .

A formal proof of Theorem 3.4 is found in Gorman [1968a]. Here, let us try to provide some intuitive insights into this result.

In Fig. 3.23 let the points A and B have a common y-coordinate and assume $A \sim B$. To show that $\{X,Z\}$ is preferentially independent of Y, we must show that if we modify the y-coordinate of A and B (keeping the y-coordinates equal) then the modified points remain indifferent. First choose a point C which has an x-coordinate in common with A, a z-coordinate in common with B and such that $C \sim A \sim B$. Now since $A \sim C$ and $\{Y,Z\}$ is preferentially independent of X, it follows that $D \sim E$. Also since $B \sim C$ and $\{X,Y\}$ is preferentially independent of Z, it follows that $D \sim F$. Hence, by transitivity, we have $E \sim F$. Now we started with $A \sim B$ and have shown that if we change the common y-coordinate by an amount Δ the resulting points F and E are indifferent. This does not prove our result since the distance Δ is chosen in a special way and is not arbitrary. But now we can repeat the process on E and F. And so on. In order to gain another degree of flexibility we also could have started the process with a point such as G where $G \sim A \sim B$. Thus we see that if we simultaneously slide the points A and B to any one of several specified y-levels, the resulting points

...To page 171a

will remain indifferent. We can repeat the argument using other points on the indifference curve through A and B and spread them out in such a way that one obtains additional points on the indifference curve through points E and F. Now one might reasonably suspect that with a sprinkling of continuity and differentiability thrown in, the result we want should follow. It does.

3.6 THE CASE OF MORE THAN THREE ATTRIBUTES

Let $X_1, \ldots, X_i, \ldots, X_n$ be n-evaluators that map any act $\underline{a}$ into a point $[X_1(a), \ldots, X_i(a), \ldots, X_n(a)] = \underline{X}(a)$ in an n-dimensional consequence space. We shall continue to assume that for any two points $\underline{x}'$ and $\underline{x}''$ in the consequence space that either $\underline{x}' \succsim \underline{x}''$ or $\underline{x}'' \succsim \underline{x}'$--if both hold then we say that $\underline{x}' \sim \underline{x}''$ and if [not $\underline{x}' \succsim \underline{x}''$] holds, we shall say that $\underline{x}'' \succ \underline{x}'$--and that the preference relation $\succsim$ is transitive.

We shall have occasion in the sequel to examine a point $\underline{x}$ by concentrating on a designated subset of its attributes as an entity and on the complementary set of attributes as an entity. For example, if n = 5 then we might want to partition $\underline{x}$ into two subvectors (x_1, x_3, x_4) and (x_2, x_5). If we let

$$\underline{y} = (x_1, x_3, x_4)$$

and

$$\underline{z} = (x_2, x_5) \quad ,$$

then we can think of $\underline{x}$ as displayed as the pair $(\underline{y}, \underline{z})$ where $\underline{y}$ involves attributes 1,3, and 4 and $\underline{z}$ involves attributes 2 and 5. More generally we shall talk about

...To page 172

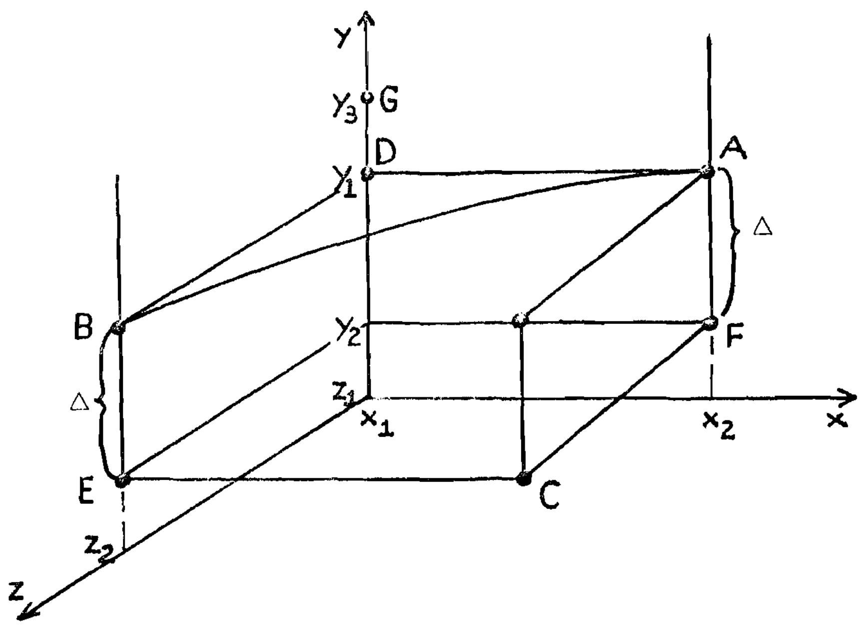


Figure 3.23. A Graphical Argument to Illustrate a Relationship Among Preferential Independence Conditions

$$\underline{x} = (\underline{y}, \underline{z})$$

where $\underline{y}$ represents those components of $\underline{x}$ on a previously specified subset of the indices $\{1,\ldots,n\}$ and $\underline{z}$ represents $\underline{x}$ on the complementary set of indices. Without any loss of generality we can always permute the indices so that we can think of $\underline{y}$ as representing $\underline{x}$ on the first s indices and $\underline{z}$ as representing $\underline{x}$ on the last n-s indices so that

$$\underline{y} = (x_1,\ldots,x_s) \quad \text{and} \quad \underline{z} = (x_{s+1},\ldots,x_n) \ .$$

In a natural manner we shall also extend this convention to talk about partitioning the attributes into two sets

$$\underline{Y} = \{X_1,\ldots,X_s\} \quad \text{and} \quad \underline{Z} = \{X_{s+1},\ldots,X_n\} \ .$$

Definition: We shall say that $\underline{y}'$ is conditionally preferred or indifferent to $\underline{y}''$ given $\underline{z}'$ if and only if

$$(\underline{y}',\underline{z}') \gtrsim (\underline{y}'',\underline{z}') \ .$$

Thus, we can talk about the conditional preference structure amongst attributes $\underline{Y}$ given that the complementary attributes are held fixed at $\underline{z}'$.

3.6.1 Preferential Independence

Definition: The set of attributes $\underline{Y}$ will be said to preferentially independent of the complementary set $\underline{Z}$ if and only if the conditional preference structure in the $\underline{y}$ space given $\underline{z}'$ does not depend on $\underline{z}'$. More symbolically, $\underline{Y}$ is preferentially independent of $\underline{Z}$ if and only if for some $\underline{z}'$,

$$[(\underline{y}',\underline{z}') \succsim (\underline{y}'',\underline{z}')] \Longrightarrow [(\underline{y}',\underline{z}) \succsim (\underline{y}'',\underline{z})] \text{ ,all } \underline{z},\underline{y}',\underline{y}'' \ .$$

As an example, there may be several benefit attributes and several cost attributes, and it may happen (this will not necessarily be the case!) that the conditional preferences amongst various packages of benefit levels may not depend on the particular costs involved. If the benefit vector $\underline{y}'$ is deemed better than the benefit vector $\underline{y}''$ at cost $\underline{z}'$, the same may hold at any other cost, $\underline{z}$. In this case we would say elliptically that "benefits are preferentially independent of costs."

If the decision maker feels that the set of attributes $\underline{Y}$ is preferentially independent of the set of complementary attributes $\underline{Z}$, then he can concentrate his efforts on structuring his preferences amongst $\underline{y}$'s holding $\underline{z}'$ fixed, knowing full well that this effort does not have to be repeated for different levels of $\underline{z}$. In this case it is meaningful for the decision maker to structure a value function v_Y defined on $\underline{y}$'s without having to specify a particular $\underline{z}'$. In particular v_Y, to be a valid value function, must be such that

$$(\underline{y}',\underline{z}) \succsim (\underline{y}'',\underline{z}) \Longleftrightarrow v_Y(\underline{y}') \geq v_Y(\underline{y}'') \quad . \qquad (3.25)$$

If $\underline{Y}$ is preferentially independent of $\underline{Z}$ we shall write $\underline{y}' \succsim \underline{y}''$ to mean $(\underline{y}',\underline{z}') \succsim (\underline{y}'',\underline{z}')$ for all $\underline{z}'$. Similarly, the notation $\underline{y}' \sim \underline{y}''$ means $(\underline{y}',\underline{z}') \sim (\underline{y}'',\underline{z}')$.

If $\underline{Y}$ is preferentially independent of $\underline{Z}$ it does <u>not</u> necessarily follow that $\underline{Z}$ is preferentially independent of $\underline{Y}$. However, the following holds.

<u>Theorem 3.5</u> If $\underline{Y}$ is preferentially independent of $\underline{Z}$, then

$$[(\underline{y}',\underline{z}') \succsim (\underline{y}',\underline{z}'')] \Rightarrow [(\underline{y},\underline{z}') \succsim (\underline{y},\underline{z}'')]$$

for all $\underline{y} \sim \underline{y}'$.

The result follows from the following string of relations which follow from the hypotheses and the meaning of preferential independence:

$$(\underline{y},\underline{z}') \sim (\underline{y}',\underline{z}') \succsim (\underline{y}',\underline{z}'') \sim (\underline{y},\underline{z}'') \quad .$$

The above theorem says that if $\underline{Y}$ is preferentially independent of $\underline{Z}$, then the conditional preference structure in the $\underline{Z}$-space given $\underline{y}$ depends on $\underline{y}$ only through its indifference surface. If v is an appropriate value function of argument $(\underline{y},\underline{z})$ then the above theorem also says that: if $\underline{Y}$ is preferentially independent of $\underline{Z}$, then $v(\underline{y},\underline{z})$ depends on $\underline{y}$ via its value function $v_Y(\underline{y})$.

If $\underline{Y}$ is preferentially independent of $\underline{Z}$ and if also $\underline{Z}$ is preferentially independent of $\underline{Y}$, then the preference structures in the $\underline{y}$ and $\underline{z}$ spaces can be considered separately. In particular, in this case, if v, v_Y, and v_Z are appropriate value functions of arguments $(\underline{y},\underline{z})$, $\underline{y}$, and $\underline{z}$ respectively, then we have

$$v(\underline{y},\underline{z}) = f\,[v_Y(\underline{y}),v_Z(\underline{z})] \quad .$$

Operationally, this means that the decision maker can structure his preferences for $\underline{y}$'s, without worrying about $\underline{z}$'s, and for $\underline{z}$'s, without worrying about $\underline{y}$'s. Then he must worry about tradeoffs between $v_Y(\underline{y})$ and $v_Z(\underline{z})$, which is a problem we analyzed earlier in Section 3.4 where we considered the case of two evaluators. We are thus led to the following question: If $v_Y(\underline{Y}) = v_Y^{\circ}$ and $v_Z(\underline{Z}) = v_Z^{\circ}$, how much are you (the decision maker) willing to give up in v_Y-units to increase v_Z from v_Z° to v_Z'? The trouble with this question is that the value functions v_Y and v_Z are not necessarily intuitively meaningful--they are only meaningful up to monotone transformations. Well, what can be done? One suggestion is the following: Suppose that

$$\underline{Y} \equiv \{X_1, X_2, \ldots, X_s\}$$

and

$$\underline{Z} \equiv \{X_{s+1}, X_{s+2}, \ldots, X_n\} \quad .$$

Choose typical values $x_2^{\circ}, \ldots, x_s^{\circ}, x_{s+2}^{\circ}, \ldots, x_n^{\circ}$ and consider the conditional preference structure in the (x_1, x_{s+1})-space given $x_2^{\circ}, \ldots, x_s^{\circ}, x_{s+2}^{\circ}, \ldots, x_n^{\circ}$. This is a "thinkable" task.

If, for example, in this subspace

$$(x_1', x_{s+1}') \sim (x_1'', x_{s+1}'')$$

given $x_2^{\circ}, \ldots, x_s^{\circ}, x_{s+2}^{\circ}, \ldots, x_n^{\circ}$, then this would mean that in the V_Y, V_Z space we would have

$$(v_Y(x_1', x_2^{\circ}, \ldots, x_s^{\circ}), v_Z(x_{s+1}', x_{s+2}^{\circ}, \ldots, x_n^{\circ})) \sim$$
$$(v_Y(x_1'', x_2^{\circ}, \ldots, x_s^{\circ}), v_Z(x_{s+1}'', x_{s+2}^{\circ}, \ldots, x_n^{\circ})) \quad .$$

Roughly, we can help structure indifference curves in the v_Y, v_Z space by examining tradeoffs between a pair of components, one from the $\underline{Y}$ set and one from the $\underline{Z}$ set, holding all other components fixed.

3.6.2 Mutual Preferential Independence and the Existence of an Additive Value Function

Definition: The attributes $X_1, \ldots, X_n$ will be said to be mutually preferentially independent if every subset $\underline{Y}$ of these attributes is preferentially independent of its complementary set of evaluators.

Recall from the previous section concerning the three attribute case that mutual preferential independence implied the existence of an additive value function*. The result is also valid for cases with more than three attributes.The general result is

Theorem 3.6 Give attributes $X_1,\ldots,X_n, n \geq 3$, an additive value function

$$v(x_1,x_2,\ldots,x_n) = \sum_{i=1}^{n} v_i(x_i) \qquad (3.26)$$

where v_i is a value function over X_i, exists if and only if the attributes are mutually preferentially independent.

Formal proofs of this theorem are found in Debreu [1960], Fishburn [1970], and Krantz et al. [1971]. Pruzan and Jackson [1963] also state this result. Since we have already informally argued through the three attribute case, we will avoid repeating the essential arguments here.

Furthermore, the argument for $n > 3$ can be made to depend on the argument for $n = 3$ by partitioning $X_1,\ldots,X_n$ into three vector variables and using the additivity results for the three dimensional case.

In the next section, we assess in some detail a four

---

*In the next subsection, it is shown that, for three or more attributes, pairwise preferential independence is equivalent to mutual preferential independence.

attribute value function in a hypothetical setting. This will again bring out some of the flavor of the relationship between preferential independence conditions and additive value functions.

3.6.3 Weakening the Additivity Assumptions

Theorem 3.6 is very useful in the sense that the additive value function is about as simple as one can find. However, as it is now written, the number of preferential independence conditions which we would need to verify get astronomically large as n gets even modestly large--say 10. Clearly, for a general n, there are $n(n-1)/2$ pairs of attributes which must be preferentially independent of their respective complements, and this says nothing about the triples of attributes, etc. Fortunately, results in Leontief [1947a,1947b] and in Gorman [1968a,1968b] save us much potential work. Let us first state this result and then discuss its use.

Theorem 3.7 Let Y and Z be subsets of the attribute set $S \equiv \{X_1, X_2, \ldots, X_n\}$ such that Y and Z overlap, but neither is contained in the other, and such that the union $Y \cup Z$ is not identical to S. If Y and Z are each preferentially independent of their respective complements, then the following sets of attributes:

(i) $Y \cup Z$,

(ii) $Y \cap Z$,

(iii) $Y - Z$ and $Z - Y$,

(iv) $(Y - Z) \cup (Z - Y)$,

are each preferentially independent of their respective complements. The reader can consult Gorman's [1968a] paper for a formal proof of this result.*

To gain some insight into the meaning of Theorem 3.7, let us assume that $S = \{X_1, X_2, X_3, X_4\}$, $Y = \{X_1, X_2\}$, and $Z = \{X_2, X_3\}$. The theorem says that if $\{X_1, X_2\}$ and $\{X_2, X_3\}$ are preferentially independent of $\{X_3, X_4\}$ and $\{X_1, X_4\}$ respectively, then

(i) the union $Y \cup Z$, namely $\{X_1, X_2, X_3\}$, is preferentially independent of X_4,

(ii) the intersection $Y \cap Z$, which is X_2, is preferentially independent of its complement $\{X_1, X_3, X_4\}$,

(iii) X_1 as $Y-Z$ and X_3 as $Z-Y$ are preferentially independent of their respective complements, and

(iv) $\{X_1, X_3\}$ is preferentially independent of $\{X_2, X_4\}$.

The two most important parts of Theorem 3.7--at least from an application's viewpoint--are (i) and (iv). These two results permit us to reduce the number of requisite pre-

---

*Given that each of $Y_1, Y_2, \ldots, Y_m$ is a subset of $S \equiv \{X_1, X_2, \ldots, X_n\}$ and is preferentially independent of its complement, one can repeatedly use Theorem 3.7 to obtain all the implied preferential independence conditions and hence, to simplify the resulting value function as much as possible. A general result in this spirit is proven in Section 6.9 using the 'utility independence' analog to Theorem 3.7.

ferential independence conditions necessary to invoke the additive value function of Theorem 3.6 to n-1, where n is the number of attributes.

The informal proof of Theorem 3.4 in subsection 3.5.4 lends some insight into why part (iv) of Theorem 3.7 is true. However, let us try to offer the concept of why part (i) is valid.

The essence of the proof can be shown from considering the special case where we let

$$\underline{x} = (x_1, x_2, x_3, \underline{x}_4)$$

and consider the case where

$$Y = \{X_1, X_2\}, \quad Z = \{X_2, X_3\}.$$

If both Y and Z are preferentially independent of their complementary sets, we shall now show that

$$Y \cup Z = \{X_1, X_2, X_3\}$$

is also preferentially independent of its complementary set. We must show

$$[(x_1', x_2', x_3', \underline{x}_4^*) \succsim (x_1'', x_2'', x_3'', \underline{x}_4^*)] \Longrightarrow [(x_1', x_2', x_3', \underline{x}_4) \succsim (x_1'', x_2'', x_3'', \underline{x}_4)], \quad \text{all } \underline{x}_4. \qquad (3.27)$$

That is if $(x_1', x_2', x_3') \succsim (x_1'', x_2'', x_3'')$ given $\underline{x}_4^*$, it is also true given any $\underline{x}_4$. Let x_1''' be such that

$$(x_1''', x_2'') \sim (x_1', x_2'), \qquad (3.28)$$

and note that this assertion makes sense since $\{X_1, X_2\}$ is preferentially independent of its complementary set*.

*Here we assume that x_2' and x_2'' were chosen such that a x_1''' satisfying (3.28) exists. The solvability and continuity assumed throughout this chapter (see Section 3.1) imply this existence.

From the hypothesis of (3.27) and (3.28) we have

$$(x_1''', x_2'', x_3', \underline{x}_4^*) \succsim (x_1'', x_2'', x_3'', \underline{x}_4^*) \quad . \tag{3.29}$$

But since $\{X_1, X_3\}$ is preferentially independent of $\{X_2, \underline{X}_4\}$, (3.29) implies, for any $\underline{x}_4$, that

$$(x_1''', x_2'', x_3', \underline{x}_4) \succsim (x_1'', x_2'', x_3'', \underline{x}_4) \quad . \tag{3.30}$$

By (3.28) together with the hypothesis that $\{X_1, X_2\}$ is preferentially independent of its complement, we find

$$(x_1', x_2', x_3', \underline{x}_4) \sim (x_1''', x_2'', x_3', \underline{x}_4) \quad . \tag{3.31}$$

From (3.31), (3.30), and transitivity we get the right hand side of (3.27). This proves our assertion. As a consequence of this result, we have the important

Corollary. If every pair of attributes is preferentially independent of its complementary set, then the attributes are mutually preferentially independent.

The argument generalizes and can be formalized by mathematical induction. If it's true for any subset of k attributes ($k \geq 2$) it can be shown to be true for k+1 evaluators. The details are omitted.

3.6.4 Selecting Preferentially Independent Sets of Attributes

Note that as a result of Theorem 3.7, there are numerous possible combinations of preferentially independent sets of attributes which imply mutual preferential independence among the members of $\{X_1, X_2, \ldots, X_n\}$. A simple combination is that

$\{X_i, X_{i+1}\}$ be preferentially independent of its complement for $i = 1,2,\ldots,n-1$.

In order to see how this works, let $n = 5$ and assume that each of the sets

$$\{X_1, X_2\}, \{X_2, X_3\}, \{X_3, X_4\}, \{X_4, X_5\}$$

has the preferential independence (P.I.) property--that is, each is preferentially independent of its complement. We then conclude from Theorem 3.7 part (iv) that

$$\{X_1, X_3\}, \{X_2, X_4\} \text{ and } \{X_3, X_5\}$$

also have the P.I. property. Repeating, we next get that

$$\{X_1, X_4\} \text{ and } \{X_2, X_5\}$$

have the P.I. property. Finally, we see that $\{X_1, X_5\}$ also has the P.I. property. Thus we see that each pair has the P.I. property and we know from the previous corollary that therefore every triplet must have the P.I. property. And so on.

Another set of n-1 assumptions which implies mutual preferential independence among $\{X_1, X_2, \ldots, X_n\}$ is that the pairs $\{X_1, X_i\}$, $i = 2,3,\ldots,n$, are each preferentially independent of its complement. The reasoning is similar to that above.

As a more involved example, suppose there are five attributes $\{X_1, X_2, \ldots, X_5\}$ and that the following sets are preferentially independent of their complements:

(a) $\{X_1, X_2\}$,

(b) $\{X_2, X_3\}$,

(c) $\{X_1,X_2,X_3,X_4\}$, and

(d) $\{X_2,X_3,X_4,X_5\}$.

It is a simple matter to show that (a) to (d) imply mutual preferential independence. Together (a) and (b) imply $\{X_1,X_2,X_3\}$ is preferentially independent of $\{X_4,X_5\}$, which when combined with (d) implies by part (iii) of Theorem 3.7 that $\{X_4,X_5\}$ is preferentially independent of its complement. By the same reasoning, (a) and (d) imply $\{X_3,X_4,X_5\}$ is preferentially independent of $\{X_1,X_2\}$, which together with (c) implies that $\{X_3,X_4\}$ is preferentially independent of $\{X_1,X_2,X_5\}$. Hence we have that $\{X_i,X_{i+1}\}$, $i = 1,2,3,4$, are preferentially independent of their respective complements from which mutual preferential independence among the X_i directly follows.

Clearly, in practice, it would not be reasonable to check directly for all possible preferential independence conditions. A little judgment on which are most likely to yield useful results could facilitate the assessment process considerably. Ting [1971] suggests a few guidelines which may help in this. An important one is to look for natural attribute groups. For instance, in an example dealing with siting of a nuclear power plant, the first level of disaggregation in the objectives hierarchy may specify the overall objective in terms of consideration for monetary costs, environmental impact, human health, and political factors. Each of these may be further specified and involves multiple attributes. However, it may be natural at this first level to have the decision maker ascertain that his preferences for attributes in various combinations of these groups do not depend on the other groups levels. Perhaps at this point, one could conclude that an additive value function existed defined over these four major attribute groups giving us something like

$$v(m,e,h,p) = v_M(m) + v_E(e) + v_H(h) + v_P(p) \quad ,$$

where M,E,H, and P represent monetary, environmental, health, and political considerations respectively. One could then try to utilize the preferential independence concept on the attributes within each grouping and hopefully further specify the decision maker's value structure.

In Section 3.8, we discuss the technique of pricing-out nonmonetary variables. For certain problems, this approach, which involves separately considering each nonmonetary attribute paired with a monetary attribute, may be reasonably natural for identifying preferential independence conditions. More details on the actual verification procedures for preferential independence are given in the assessment Section 6.6 for multiattribute utility functions.

3.6.5 Value Functions With Partial Additivity

Before concluding this section, we should indicate that even when mutual preferential independence does not hold, the existence of any preferential independence properties that do hold may help considerably in structuring the value function.

Theorem 3.8 Given $\{X_1, X_2, X_3, X_4\}$, if $\{X_1, X_2\}$ and $\{X_2, X_3\}$ are preferentially independent of their respective complements, a value function v exists of the form

$$v(x_1, x_2, x_3, x_4) = f(y, x_4) \tag{3.32}$$

where $y = v_1(x_1) + v_2(x_2) + v_3(x_3)$ and f is increasing in its first variable.

A proof of this result is in Gorman [1968a].

Note that $v_1(x_1) + v_2(x_2) + v_3(x_3)$ can be thought of as a conditional additive value function over attributes X_1, X_2, and X_3 given that X_4 is fixed at some convenient level. This level does not matter since by the conditions of Theorem 3.8, it follows from Theorem 3.7 that $\{X_1, X_2, X_3\}$ is preferentially independent of X_4.

Since the X_i in Theorem 3.8 can designate vector attributes, the theorem represents a general attribute case. It is important to realize that this result can be used several times--perhaps corresponding to different levels in the objectives hierarchy--in structuring the same value function.

3.6.6 Using the Additive Value Function

As illustrated in earlier two attribute assessments of the additive value function, rather than using the form

$$v(x_1, x_2, \ldots, x_n) = \sum_{i=1}^{n} v_i(x_i) \tag{3.33}$$

directly, when v is bounded, it may be more convenient to scale v and each of the single attribute value functions from zero to one. Thus, we will have the additive value function of the form

$$v(x_1, x_2, \ldots, x_n) = \sum_{i=1}^{n} \lambda_i v_i(x_i) \quad , \tag{3.34}$$

where v and v_i, i=1,2,...,n, in (3.34) are scaled from zero to one and

$$\sum_{i=1}^{n} \lambda_i = 1 \ , \quad \lambda_i > 0 \quad . \tag{3.35}$$

Equations (3.33) and (3.34) are both additive value functions and given consistent scaling, they are equivalent. The assessment of (3.34) is illustrated in the next section.

3.7 ASSESSMENT OF AN ADDITIVE VALUE FUNCTION: AN ABSTRACT HYPOTHETICAL EXAMPLE*

In this section we shall illustrate by means of an example how a decision maker might assess an additive value function over four attributes.

Suppose that you, the decision maker, have to choose amongst 75 (say) alternative acts and that each act can be evaluated in terms of four attributes. Table 3.1 summarizes these evaluations. For example, act A_1 has a score of 7.5 on attribute X_1, a score of 344 on attribute X_2, a score

---

*In Section 7.2, the work of James Roche in utilizing the procedures illustrated here for evaluating alternative instructional programs in a public school system is discussed.

of 0.47 on X_3, and 12.15 on X_4. For act A_i, the recorded scores are x_{1i}, x_{2i}, x_{3i}, and x_{4i} on attributes X_1 to X_4 respectively. Let us assume that attributes X_1, X_2, and X_3 are positively oriented in the sense that you would prefer higher scores on each of these attributes, but assume that attribute X_4 is negatively oriented in the sense that you would prefer lower scores*.

Your problem is: Given performance evaluation of these 75 acts on these four attributes, which act should you single out as being best for you? That is, how can you systematically probe your feelings about these attributes so that you could force yourself to articulate your underlying preference structure?

For the time being observe, however, that A_{75} cannot be a serious contender for "best" since A_1 is better than A_{75} on each of the four attributes--remember that for the 4th attribute 12.15 is better than 12.92. In technical jargon A_{75} is dominated by A_1.

---

*This assertion implicitly assumes that each attribute, taken individually is preferentially independent of its complement.

TABLE 3.1

Performance Measures of Alternative Acts

on Four Attributes

Note that act A_1 dominates act A_{75}

Act	Attributes X_1 (↗)	X_2 (↗)	X_3 (↗)	X_4 (↘)
A_1	7.5	344	.47	12.15
A_2	3.7	268	.79	12.20
⋮	⋮	⋮	⋮	⋮
A_i	x_{1i}	x_{2i}	x_{3i}	x_{4i}
⋮	⋮	⋮	⋮	⋮
A_{75}	6.7	250	.24	12.92
Lowest (rounded)	2.0	200	.15	12.00
Highest (rounded)	9.0	400	.90	13.50

Performance Profile of A_1: (7.5, 344, .47, 12.15)

On the bottom of Table 3.1 note that the 75 entries under attribute X_1 lie within the interval from 2.0 to 9.0. The entries under attribute X_2 lie in the interval 200 to 400. Similarly the ranges for attributes X_3 and X_4 are recorded.

...To page 187a

Observe, once again that for attribute X_4 no act is better than 12.00 or worse than 13.50.

The four numbers x_{1i}, x_{2i}, x_{3i}, and x_{4i} associated with act A_i can be thought of as the profile of A_i and the profiles of acts A_1 and A_2 are shown in Fig. 3.24.

3.7.1 Legitimacy of the Additive Value Function

Now let us suppose that you, the decision maker, feel that any pair of attributes is preferentially independent of the others. Thus for example, suppose that the tradeoffs for attributes X_2 and X_3 say, keeping the levels of attributes X_1 and X_4 fixed, do not depend on the particular values of these fixed levels. And so on for each pair of attributes. Now as we indicated in Section 3.6, your preferences, if they are to be fully articulated in a manner consistent with the above preferential independence assumptions, must be in a form that can be characterized as consistent with a value function v of the form

$$v(x_1,x_2,x_3,x_4) = \sum_{j=1}^{4} \lambda_j \; v_j(x_j)$$

where

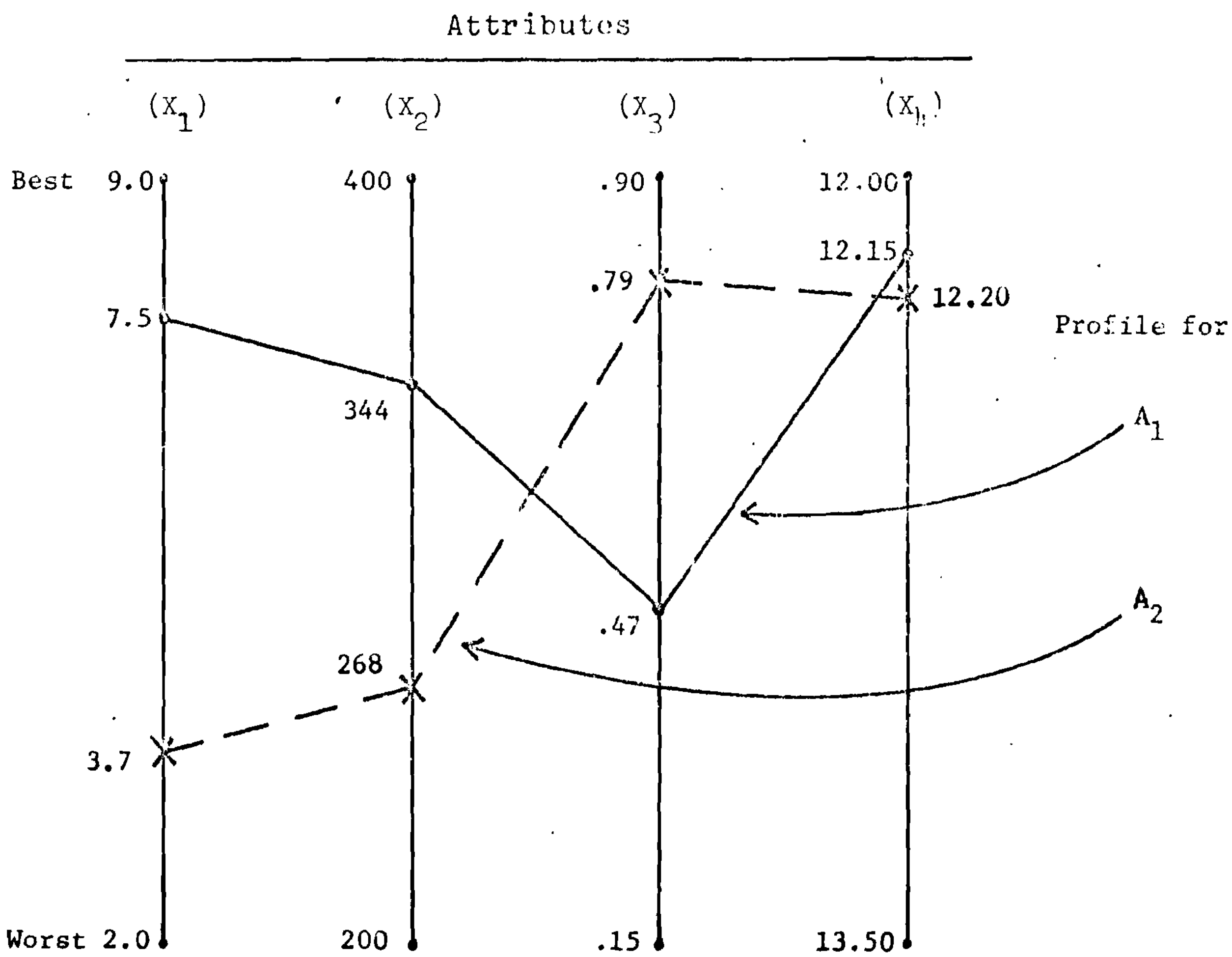
a. $v_j(\text{worst } x_j) = 0$, $v_j(\text{best } x_j) = 1$, $j = 1,\ldots,4$;

b. $0 < \lambda_j < 1$, $j = 1,\ldots,4$;

c. $\sum_{j=1}^{4} \lambda_j = 1$.

We can think of the function v_j defined over the attribute score x_j as the j-th component value function and the

... To page 188



Performance Profiles of Acts A_1 and A_2 for a Hypothetical Example

Figure 3.24

λ_j as the weight associated with attribute X_j. For our illustrative example, we note from Table 3.1 that the worst x_1 score is 2.0 and the best is 9.0. It will turn out later that the determination of the λ_j-weights are intimately related to the ranges of the scales.

The problem you now face is to determine appropriate v_j functions and λ_j weights. By so doing you will have articulated your underlying preference structure for x-profiles.

3.7.2 Assessment of Component Value Functions.

One procedure for determining the v_1, v_2, v_3, and v_4 functions is described and illustrated in subsection 3.4.7. Let us merely illustrate in sketchy form how one might assess v_1.

First we normalize v_1 by letting $v_1(2.0) = 0$ and $v_1(9.0) = 1$. We then seek the subjective mid-point, let us call it $m_{.5}$ say, of the interval 2.0 to 9.0. That is we want to find the value $m_{.5}$ for which $v_1(m_{.5}) = .5$. We ask for that knife-edge point where the intervals $(2.0, m_{.5})$ and $(m_{.5}, 9.0)$ are differentially value equivalent. The value $m_{.5}$ is such that if

$$(2.0, b, c, d) \sim (m_{.5}, b', c', d')$$

then

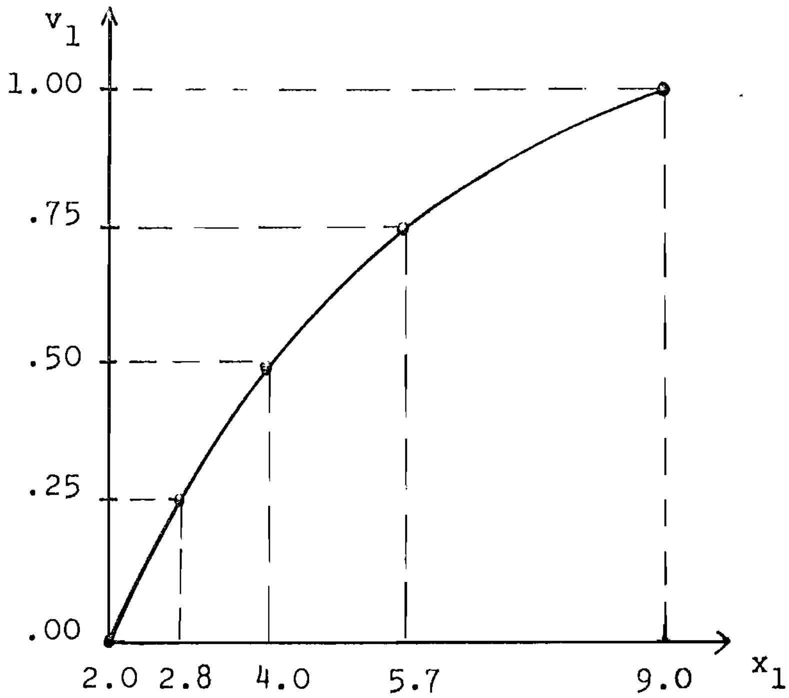
$$(m_{.5}, b, c, d) \sim (9.0, b', c', d') \quad .$$

If one gives up a certain amount of attributes X_2, X_3 and X_4-- i.e. by going from (b,c,d) to (b',c',d')--to go from 2.0 to $m_{.5}$, then one should be willing to give up exactly the same amount to go from $m_{.5}$ to 9.0.

Well let's say the mid-value of 2.0 to 9.0 is 4.0. We then go through the same procedure for determining the mid-point of the interval 2.0 to 4.0. Let's say it is 2.8 so that $v_1(2.8) = .25$. Similarly let the mid-point of the range 4.0 to 9.0 be 5.7 so that $v_1(5.7) = .75$. These points can now be plotted as shown in Fig. 3.25 and a curve v_1 can be faired through these five points. Or alternatively more mid-points could be determined before fairing in the curve. It depends on how much accuracy is needed. We reiterate the point we have made several times earlier: it may be desirable to run consistency checks (e.g. finding the mid-point of 2.8 to 5.7) and to police the inconsistencies so that a coherent set of compatible responses is obtained. In addition, one may wish initially, before specific numbers are chosen, to check in a qualitative way whether v_1 is concave, convex, or is perhaps more complicated in shape.

3.7.3 Assessment of Scaling Constants.

Some special notation should help our discussion of the λ_j's. For the j-th attribute let w_j represent the worst value and b_j the best value. Then for positively oriented scales we would have $w_j \leq x_j \leq b_j$. Let I be the complete set of attribute indices; in our example $I = \{1,2,3,4\}$. Let T be a subset of I and $\bar{T}$ be the complementary set to T, or $\bar{T} = I - T$. Let $\underline{x}^T$ be that profile where all the component x_j's are equal to b_j for $j \in T$ and equal to w_j for $j \in \bar{T}$. Thus, for example if $T = \{2, 3\}$, then



Assessment of the Component Value Function v_1

Figure 3.25

$$\underline{x}^T = \underline{x}^{\{2,3\}} = (w_1, b_2, b_3, w_4) \quad .$$

Since $v_j(w_j) = 0$ and $v_j(b_j) = 1$, we know that

$$v(\underline{x}^T) = \sum_{j \varepsilon T} \lambda_j$$

so when $T = \{2,3\}$, then $v(\underline{x}^T) = \lambda_2 + \lambda_3$.

Also define

$$\lambda(T) \equiv \sum_{j \varepsilon T} \lambda_j \ .$$

Notice that when T consists of the single element set $\{j\}$ we have

$$v(\underline{x}^{\{j\}}) = \lambda_j = \lambda(\{j\}) \quad .$$

Our task is to suggest techniques for the determination of the λ_j's. One suggestion is to start off by ranking the profiles $\underline{x}^{\{1\}}, \ldots, \underline{x}^{\{4\}}$. Suppose, for example, that you feel that

$$\underline{x}^{\{2\}} \succ \underline{x}^{\{1\}} \succ \underline{x}^{\{4\}} \succ \underline{x}^{\{3\}} \quad .$$

This would imply that for you

$$\lambda_2 > \lambda_1 > \lambda_4 > \lambda_3 \quad .$$

Next, you could try to get more refined inequalities by comparing say

$$\underline{x}^{\{2\}} \text{ vs } \underline{x}^{\{1,3,4\}} \quad .$$

If in this paired comparison $\underline{x}^{\{2\}}$ were preferred then we could infer that $\lambda_2 > .5$.

Observe that when you are asked to compare $\underline{x}^T$ to $\underline{x}^S$

you are essentially asked the following question: "Suppose the $\underline{x}$-profile were at the worst case, (w_1,w_2,w_3,w_4), and you had the option of improving some of the w_j's from the worst to the best position. Would you rather improve the levels of the attributes in the subset T or subset S?"

This method of analysis usually only provides inequalities for the λ_j's. In some special cases precise numerical values can be deduced if there are indifferences. For example, if $\underline{x}^{\{T\}}$ and $\underline{x}^{\{\bar{T}\}}$ are indifferent, then $\lambda(T) = .5$. But this is not the usual case.

Let us continue with the special case where

$$\lambda_2 > \lambda_1 > \lambda_4 > \lambda_3 \quad .$$

Now compare the two profiles,

$$(w_1,x_2,w_3,w_4) \quad \text{vs} \quad \underline{x}^{\{1\}} \quad ,$$

and manipulate the level of x_2 until indifference is reached. Suppose this occurs at $x_2 = 350$; that is, suppose

$$(2.0,350,.15,13.50) \sim (9.0,200,.15,13.50) .$$

Then we have

$$v(2.0,350,.15,13.50) = v(9.0,200,.15,13.50)$$

or

$$\lambda_2 v_2(350) = \lambda_1 \quad ,$$

and since it is assumed the component v_2 function has already been assessed, we can find $v_2(350)$. Suppose it is

$$v_2(350) = .6 \quad ,$$

so that

In a similar fashion we can determine the proportional relationships between λ_4 and λ_2, and between λ_3 and λ_2. Assume in particular that

$$(2.0, 240, .15, 13.50) \sim (2.0, 200, .15, 12.00)$$

and

$$v_2(240) = .4 \quad ,$$

so that

$$.4\lambda_2 = \lambda_4 \quad ; \tag{3.37}$$

also assume that

$$(2.0, 210, .15, 13.50) \sim (2.0, 200, .90, 13.50)$$

and

$$v_2(210) = .1 \quad ,$$

so that

$$.1\lambda_2 = \lambda_3 \quad . \tag{3.38}$$

From equations (3.36), (3.37), (3.38) and

$$\lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 = 1 \quad ,$$

we conclude that

$$\lambda_1 = .286 \quad , \quad \lambda_2 = .476 \quad , \quad \lambda_3 = .048 \quad , \quad \lambda_4 = .190 \quad .$$

We repeat, as we have so often in the past, that it may be desirable to ask additional questions thereby getting an over determinate system of equations, fully expecting that the set of responses would in practice be inconsistent. These inconsistencies can be used by the analyst to "force" the decision maker to rethink through his preferences. Hopefully, reasons for the original inconsistencies can be found, and from this, a consistent set of preferences established.

3.7.4 Additional Comments on the λ-Function.

The λ-function defined on subsets of I satisfies the usual rules of a probability measure:

a. $\lambda(T) \geq 0$, for $T \subset I$;

b. $\lambda(I) = 1$;

c. if S and T are disjoint,

$$\lambda(S \cup T) = \lambda(s) + \lambda(T) \quad .$$

Thus finding the λ-function is not unrelated to the problem of finding suitable probability assignments over a finite sample space. Very often in assigning the weighting measure λ, just like in assigning a probability measure, it is not natural to initially assign weights at the atomic level--i.e., to assign numbers for $\lambda_1, \lambda_2, \ldots$. Rather it may be more appropriate to make initial assignments to subsets (e.g. to assign values to $\lambda(T)$ for special subsets) and to make conditional assignments. Let us illustrate. Suppose, for example, we consider a ten-attribute case with the hierarchical structure shown in Figure 3.26. In this case let

$$I = \{1,2,\ldots,10\}$$

$$A = \{1\}, \quad B = \{2,3,4\} , \quad C = \{5,6\} , \quad D = \{7,8,9,10\},$$

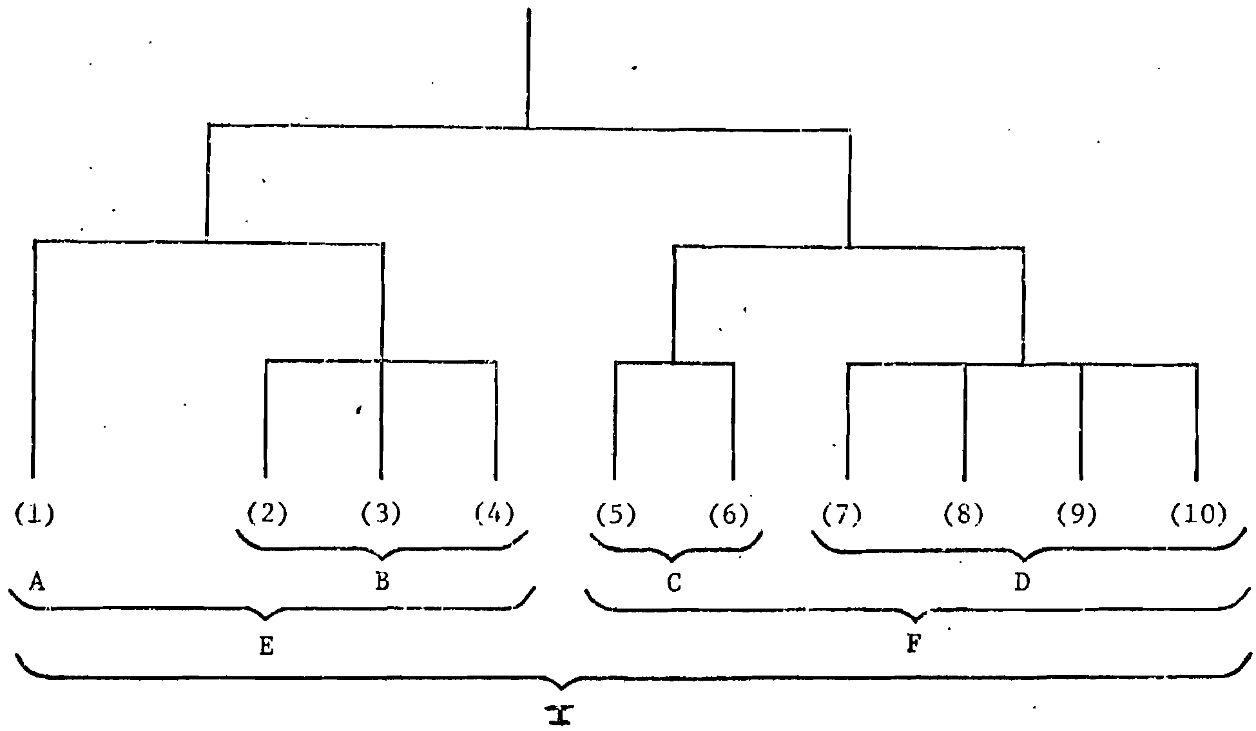


Figure 3.26

The Hierarchical Structure of Objectives Used in Evaluating Scaling Constants

$$E = A \cup B \quad , \qquad F = C \cup D \quad .$$

In such a hierarchical example, it might be natural to compare

$$\lambda(E) \quad \text{vs} \quad \lambda(F)$$
$$\lambda(A) \quad \text{vs} \quad \lambda(B)$$
$$\lambda(C) \quad \text{vs} \quad \lambda(D) \quad .$$

Taking our cue from probability theory, it might also be appropriate to define conditional weighting functions, such as $\lambda(B|E)$ which could be defined as

$$\lambda(B|E) = \frac{\lambda(B)}{\lambda(E)} \qquad \text{for } B \subset E \quad ,$$

where $\lambda(B|E)$ gives the "weighting importance" of attribute set B within the subset E, or the conditional weighting of B within E.

In hierarchical attribute sets with many attributes it is critically important to isolate components of the problem and to make conditional assessments. In Figure 3.27, we have concocted some hypothetical conditional assignments. For example, we have let

$$\lambda(E) = .6 \qquad \text{and} \qquad \lambda(F) = .4 \quad ,$$
$$\lambda(A|E) = .5 \qquad \text{and} \qquad \lambda(B|E) = .5 \quad ,$$
$$\lambda(\{2\}|B) = .5 \; , \; \lambda(\{3\}|B) = .3 \quad , \text{ and } \quad \lambda(\{4\}|B) = .2 \; ,$$
$$\lambda(C|F) = .8 \qquad \text{and} \qquad \lambda(D|F) = .2 \quad ,$$

and so on.

To find λ_3, say, we have

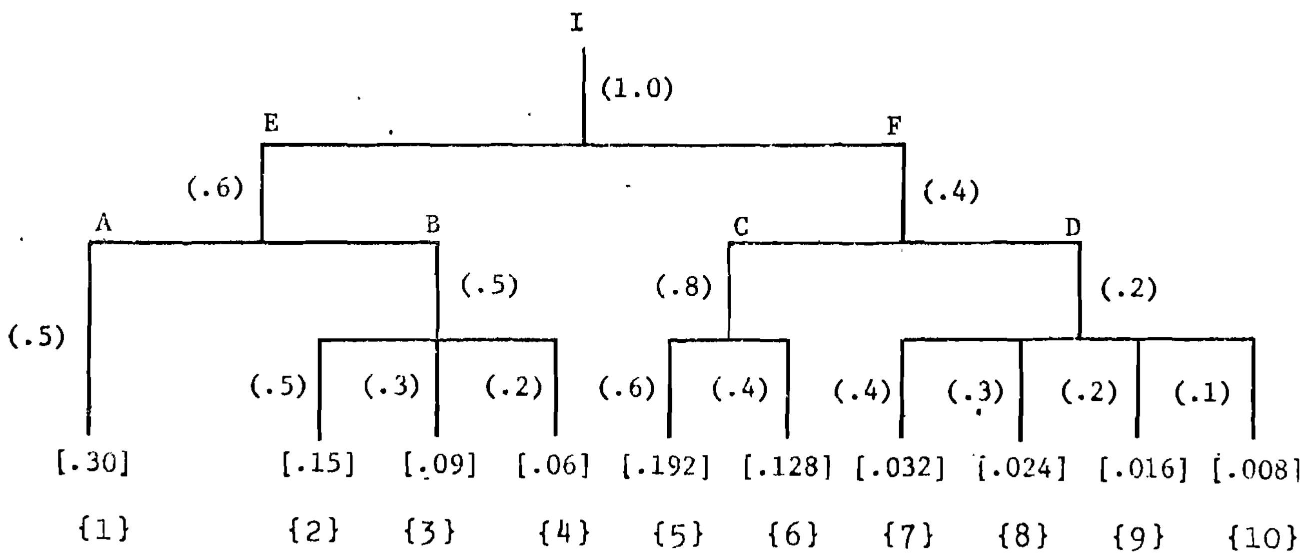


Figure 3.27

Hypothetical Scaling Constants in a Hierarchical Structure

$$\lambda_3 = \lambda(\{3\}|B) \quad \lambda(B|E) \; \lambda(E)$$
$$= .3 \times .5 \times .6 = .09 \quad .$$

In a similar manner we get all the individual λ_j's, which are displayed in the second row from the bottom in Fig. 3.27.

In a problem such as this it might be clear, for example, how to assign conditional weights within subsets E and F but one might be hard pressed to apportion weights between E and F. But an ability to structure part of the problem might make it possible to run meaningful sensitivity analyses on those critical assessments which are the hardest to make. The above remark about sensitivity analysis--and remarks similar to it which we have already made and will make numerous times in the sequel--are especially important if more than one decision maker is involved in the decision process.

3.8. WILLINGNESS-TO-PAY

Consider an attribute structure with a monetary attribute M, measured in monetary units m, and other attributes are $X_1, X_2, \ldots, X_n$. Paired comparisons are then of the form

$$(m^1, x_1^1, \ldots, x_n^1) \quad \text{versus} \quad (m^2, x_1^2, \ldots, x_n^2) \quad ,$$

or more compactly

$$(m^1, \underline{x}^1) \qquad \text{versus} \quad (m^2, \underline{x}^2) \quad .$$

3.8.1 Pricing Out.

In many contexts--but we insist not all!-- it is natural to proceed by "costing out" or "pricing out" the $\underline{x}$-components. For example, we might single out some particular $\underline{x}$-profile, let us call it $\underline{x}^*$ and ask such a question: "Starting from the profile $(m^o,\underline{x}^o)$ how much would you just be willing-to-pay to alter $\underline{x}^o$ to the base case $\underline{x}^*$?" We are asking the decision maker in essence to find the value m in the indifference equation

$$(m,\underline{x}^*) \sim (m^o,\underline{x}^o) \quad .$$

The willingness-to-pay would then be $m - m^o$.

If one had to evaluate a limited number of alternatives $(m^i,\underline{x}^i)$ for i = 1 to N, and if one determined for each i a value m_i^* such that

$$(m_i^*,\underline{x}^*) \sim (m^i,\underline{x}^i) \quad , \quad i = 1,\ldots,N,$$

then one could rank the N alternatives in terms of the numbers m_1^* to m_N^*.

This procedure becomes even more attractive under some special structure. For example, in the indifference equation

$$(m,\underline{x}^*) \sim (m^o,\underline{x}^o) \quad ,$$

the willingness-to-pay for changing $\underline{x}^o$ to $\underline{x}^*$ might (in a special case) not depend on the level m^o. This simplifies things. However, if this is not the case, and if the number N of alternatives is large, then the dependence of $m - m^o$ on m^o becomes a particularly bothersome complication.

If the dimensionality of $\underline{x}$ is large, it is helpful to price-out the transformation of $\underline{x}^o$ to $\underline{x}^*$ in stages.

For example, we might want first to consider the component x_j and modify it to the base x_j^*. We are then led to the indifference equation

$$(m,x_1^o,\ldots,x_{j-1}^o,x_j^*,x_{j+1}^o,\ldots,x_n^o) \sim (m^o,\underline{x}^o) \quad .$$

In general without special assumptions the willingness-to-pay $m - m^o$ will depend not only on x_j^o and x_j^* but also on m^o, and on $x_1^o,\ldots,x_{j-1}^o,x_{j+1}^o,\ldots,x_n^o$. If, however, the monetary attribute M and attribute X_j, taken as a pair, are preferentially independent of the complementary set of attributes, then we can 'price out' the change from x_j^o to x_j^* without worrying about the levels of the other attributes. We still, of course, have to worry about the initial monetary levels m^o.

If the pair $\{M,X_j\}$ is preferentially independent of the complementary set for each j, then we can price out the attributes in sequence. For example, suppose

$$(m^o + \Delta^1,x_1^*,x_2^o,x_3^o,\ldots) \sim (m^o,x_1^o,x_2^o,x_3^o,\ldots) \quad ,$$

so that Δ^1 is what we "pay" for the transformation of x_1^o to x_1^*; in general Δ^1 will depend on m^o (but not on $x_2^o,x_3^o,\ldots$). Next suppose that

$$(m^o + \Delta^1 + \Delta^2,x_1^*,x_2^*,x_3^o,\ldots) \sim (m^o + \Delta^1,x_1^*,x_2^o,x_3^o,\ldots)$$

so that Δ^2 is the price we "pay" for transforming x_2^o to x_2^* and this will depend, with the assumptions we've made, on $m^o + \Delta^1$, on x_2^o and on x_2^* but not on the other x's. And so on. When we price out the transformation of x_j^o to x_j^*, it will unfortunately depend on $m^o + \Delta^1 + \ldots$ $\ldots + \Delta^{j-1}$, unless of course we explicitly assume otherwise.

If, in general, the pair $\{M,X_j\}$ is preferentially independent of the complementary set of attributes for all j <u>and if</u> the quantity Δ^j in the indifference equation

$$(m^o + \Delta^j, x_1^o, \ldots, x_{j-1}^o, x_j^*, x_{j+1}^o, \ldots, x_n^o) \sim (m^o, \underline{x}^o)$$

does not depend on m^o, for each j, then life becomes especially attractive. Then we can price out the transformation of x_j^o to x_j^* without first determining in a sequential order the values of Δ^1, $\Delta^2, \ldots, \Delta^{j-1}$.

In some circumstances it may not be possible to assume that $\{M, X_j\}$, $j = 1, 2, \ldots, n$, is pairwise preferentially independent of the complementary set of attributes. One might, however, in some contexts be able to partition the $\underline{X}$ attributes into two subsets, Y and Z, so that in a suggestive notational fashion we can express

$$(m, \underline{x}) \quad \text{as} \quad (m, \underline{y}, \underline{z}) \quad .$$

If the attribute set {M,Y} is preferentially independent of Z, one can price out a transformation say of $\underline{y}^o$ to $\underline{y}^*$ and in doing so we would not have to worry about the $\underline{z}^o$-levels.

The willingness-to-pay procedure has its virtues.It is easy to explain and that in itself should not be underestimated. Unfortunately, it is often applied in a manner that can only be justified under certain assumptions when indeed these assumptions cannot be fully justified. We're referring here to the assumptions:

i. the money attribute taken together with any other single attribute is preferentially independent of the others, and

ii. the marginal rate of substitution between money and any other attribute does not functionally depend on the monetary level.

It should also be pointed out that <u>even if</u> the above assumptions make sense in a given context it does not necessarily follow that the willingness-to-pay procedure should be followed. In many cases it may be too painful and too unnatural to try to price out a transformation of $\underline{x}^o$ to $\underline{x}^*$ or even of x_j^o to x_j^*. In some circumstances it may be more natural to directly attempt to specify the preference structure as discussed in Sections 3.3 to 3.7.

For some interesting examples where willingness-to-pay arguments are used in a multiattribute context, see several of the publications of the Decision Analysis Group of Stanford Research Institute: Matheson and Roths [1967], Stanford Research Institute [1968], Boyd et al. [1971], and Howard, Matheson, and North [1972].

3.8.2 Dominance and Extended Dominance

There are loads of tricks one can use for processing preferences short of establishing a full value function and it is hopeless here to try to be very systematic about describing many of the tricks of the trade. But one point that has been exploited by us in practice bears some emphasis. It is not easy to make the kinds of tradeoffs that we have been glibly describing. If one could avoid making some of these vexing tradeoffs, then this should be exploited. One obvious device is to exploit the concept of dominance introduced in Section 3.2. If we compare

$$\underline{x}' = (x_1',\ldots, x_n') \text{ and } \underline{x}'' = (x_1'',\ldots,x_n'')$$

and if x_j' is preferred to x_j'' for all j (or preferred or indifferent for all j and strictly preferred for some j) then $\underline{x}''$ can be eliminated as a contender if $\underline{x}'$ is available. Getting rid of dominances may solve the problem. Fine, if this is the case!

Now suppose that we try the above reduction by dominance and the decision problem is not resolved...the usual situation. Furthermore suppose that we can partition $\underline{x}$ into $(\underline{y},\underline{z})$ and let us suppose that we can "price out" $\underline{y}$'s in terms of the $\underline{z}$'s by transforming each $\underline{y}$ to some base--$\underline{y}^*$, say. That is, for the i-th alternative $(\underline{y}_i,\underline{z}_i)$, we solve the indifference equation

$$(\underline{y}_i,\underline{z}_i) \sim (\underline{y}^*,\underline{z}_i^*)$$

for $\underline{z}_i^*$. Let us assume this is repeated for i = 1,...,N. Now once again one can investigate dominance relations amongst the restrictive profiles, $\underline{z}_1^*,\ldots,\underline{z}_N^*$. Of course, this latter type of extended dominance does incorporate the subjective reduction of $(\underline{y}_i,\underline{z}_i)$ to $(\underline{y}^*, \underline{z}_i^*)$ for i = 1,...,N.

If the processes of dominance and extended dominance help to isolate a best act, then this would be a welcome bonus. More generally, however, the elimination of alternatives have other beneficial effects: it is usually helpful to have fewer real alternatives to consider, since with a reduction in the number of alternatives one must consider there is likely to be a diminution in the size of the interval that is necessary for each of the scalar attribute scales. And this restriction of the intervals for each attribute, in turn, makes it more palatable to adopt various assumptions like preferential independence (and other variations to be introduced later). To illustrate this last point suppose we consider the case of three attributes and are contemplating whether or not it is legitimate to assume that attributes 1 and 2 are preferentially independent of attribute 3. This might be a reasonable assumption to investigate (a palatable lie) provided that the range of values of attribute 3 is sufficiently narrow. We might not be able to make this convenient assumption if the third attribute varies widely. And here is where some preliminary work on dominance and especially extended dominance may have a significant impact.

3.9 BRIEF SUMMARY AND GUIDE TO SOME RELEVANT LITERATURE

The basic objective of this chapter was to present techniques for assessing multiattribute value functions. Once the decision maker articulates a value function, which implies a preference ordering over all multidimensional evaluations, the subsequent analysis must then examine the set of technologically achievable evaluations and choose a best evaluation in this set. The two processes, determination of achievability and articulation of a preference structure can be kept separate and fused at the very end of the analysis. Indeed in this book we concentrate almost exclusively on the latter of these two processes. However, at the beginning of this chapter we did describe a very informal mechanism for intertwining these two processes: one first finds a point on the efficient frontier of achievable evaluations and then one moves around this frontier in a manner that improves one's preferences at each step. This is done in a rather ad hoc manner that does not require a full specification of one's preference structure. While this procedure may sometimes be effective in some special, highly structured problems (e.g., in linear programming problems with more than a single linear objective function), in most of the applied problems dealt with in this book this informal, interactive, search procedure is not very useful--especially when probabilistic concerns are introduced. We therefore concentrate our attention on the aspect of the decision problem dealing with the articulation of preferences. We do it also in a manner that will enable us later on to bring in probabilistic considerations.

Sections 3.4 to 3.6 provide a number of representation theorems which break down the assessments of the value function into component parts. The key concept in all these reduction techniques is that of <u>preferential independence</u>. Because there is considerable power in the implications of overlapping sets of attributes being preferentially independent of their complements, the two-attribute case cannot be dealt with nearly so nicely as cases with three or more attributes. Most of the important representation theorems provided conditions for expressing the value function v in the additive form

$$v(x_1, x_2, \ldots, x_n) = \sum_{i=1}^{n} v_i(x_i) \quad ,$$

where the v_i are consistently scaled single-attribute value functions. A complete example illustrating the assessment of such a function is given in Section 3.7.

A common practice of many analysts is to 'price out'--that is, bring down to some standard level--all the non-monetary attributes into some (single) monetary attribute. A comparison of alternatives is then made only in terms of the 'adjusted' levels of the monetary attribute. The requisite assumptions necessary for such an approach to be valid are strong. These are discussed in Section 3.8.

Most of this chapter is expository in nature, since as indicated throughout, the fundamental results are due to others. Our approach has been to state an important result and then to informally argue through the reasoning to obtain it. Formal proofs of the theorems have been referenced to the original articles and the technical literature.

We would like now to present a cursory review of the literature. The purpose is merely to suggest some sources where an interested reader may search for more depth than provided here. We will, however, try to mention some classic works that are directly relevant.

Leontief [1947a,1947b] investigated properties of functions of several variables which provided for separability, breaking the original function into a function defined as functions over distinct subsets of the original variables. His results were local in nature rather than global. Debreu [1960] provided the first axiomatization implying the existence of an additive value function for three or more attributes. His elegant proof was topological in nature. An alternative algebraic proof of additivity was given by Luce and Tukey [1964] in their paper introducing 'conjoint measurement' for the two-attribute case. Several extensions to conjoint measurement were made by individuals such as Krantz [1964), Luce [1966], and Tversky [1967]. For a complete summary of this field we highly recommend the book Foundations of Measurement by Krantz, Luce, Suppes and Tversky [1971]. In a general measurement context, this book also presents representation theorems for a number of more general value functions than those considered in this chapter. This includes the large class of value functions which can be represented by polynomial structures. A recent addition to this class of literature is Fishburn [1974a].

An important contribution toward separating the assessment of a value function into a number of component parts is Gorman [1968a]. His results allow us to reduce greatly the number of conditions necessary to imply a value function is additive, thus, making the techniques more operational. Ting [1971] discusses many techniques for decomposing the assessment of preferences and suggests some guidelines for verifying the assumptions necessary to use the results.

CHAPTER 4

UNIDIMENSIONAL UTILITY THEORY

This chapter concerns unidimensional utility theory: the assignment of utilities to consequences that are described in terms of one scalar attribute. The general problem addressed can be stated simply. A decision maker must choose among several alternatives $A_1, A_2, \ldots, A_n$, each of which will eventually result in a consequence describable in terms of attribute X. However, the decision maker does not know exactly what consequence will result from each of the various alternatives, but he can assign probabilities to the various possibilities which might result from any course of action. What should he do?

4.1. THE MOTIVATION FOR UTILITY THEORY*

The power of the concept of utility and the grounds for our interest in it is this: If an appropriate utility is

---

*Sections 4.1 through 4.8 present an expository account of much of the standard literature of single attribute utility theory. It draws heavily on the research work in the last fifteen years of Robert Schlaifer, Kenneth Arrow, John Pratt, and Richard Meyer. Readers who are thoroughly familiar with the concepts and results in Pratt [1964] may wish to skim briefly these sections.

assigned to each possible consequence and the expected utility of each alternative is calculated, then the best course of action is the alternative with the highest expected utility. Different sets of axioms which imply the existence of utilities with the property that expected utility is an appropriate guide for consistent decision making are presented in von Neumann and Morgenstern [1947], Savage [1954], Luce and Raiffa [1957], Pratt et al. [1965], and Fishburn [1970]. The next subsection informally reviews the basic ideas of the theory.

In terms of our double dichotomy of Chapter 1 depicted in Fig. 4.1, the problem addressed in this chapter is a special case of the general problem of Chapter 3 in the sense that we are concerned with only one unidimensional attribute but a generalization in the sense that uncertainty is now involved. One might ask why, when we spend most of Chapters 1 and 2 arguing that most important "real world" problems require more than one attribute to adequately summarize consequences, do we allocate a chapter solely to the unidimensional case? Our reason is three-fold. First a thorough understanding of unidimensional utility theory and the associated techniques in implementing the theory is essential for work on the multiattribute problem involving uncertainty, second there are some important problems where one scalar attribute may be adequate, and third we shall show that many multidimensional utility problems can be reduced to unidimensional ones by using

	Unidimensional	Multidimensional
Certainty		Chapter 3
Uncertainty	Chapter 4	Chapters 5 and 6

Figure 4.1 Double Dichotomy of Decision Problems

some of the techniques of the previous chapter. These are expanded on later in this section.

4.1.1 Basic Fundamentals of the Utility Theory

We are assuming that most of our readers are familiar with the basic fundamentals of utility theory, but as a review to some and a short introduction to others we offer the following.

Suppose we start out with just n consequences labelled $x_1, x_2, \ldots, x_n$. It is immaterial at this point what the underlying scales of these x's are. Each x could be a scalar, a vector, or a paragraph of prose describing this consequence. It is important, however, that the decision maker can rank the consequences in order of his preference, and we shall assume the labelling is such that x_1 is less preferred to x_2, which is less preferred to x_3, and so on. In symbols, we assume

$$x_1 \prec x_2 \prec x_3 \prec \cdots \prec x_n \quad . \tag{4.1}$$

Now suppose the decision maker is asked to express his preferences for probability distributions over these consequences. For example, the decision maker is asked to state his preference between act a' and a'' where

i) Act a' will result in consequence x_i with probability p_i', for i = 1,2,...,n. Of course, $p_i' \geq 0$, all i, and $\Sigma_i p_i' = 1$.

ii) Act a'' will result in consequence x_i with probability p_i'', for i = 1,2,...,n. Again $p_i'' \geq 0$, all i, and $\Sigma_i p_i'' = 1$.

Notice that there are an infinity of potential probability distributions over this finite set of consequences. Now suppose the decision maker asserts that, for each i, he is indifferent between the following two options:

<u>Certainty Option:</u> Receive x_i

<u>Risky Option</u>: Receive x_n (the best consequence) with probability π_i and x_1 (the worst consequence) with the complementary probability $1 - \pi_i$.

Let us denote the risky option by $\langle x_n, \pi_i, x_1 \rangle$. Furthermore, the decision maker is consistent in the sense that he assigns $\pi_n = 1$ and $\pi_1 = 0$ and the π's are such that

$$\pi_1 < \pi_2 < \ldots < \pi_n \quad . \tag{4.2}$$

Comparing (4.2) with (4.1) we can see that the π's can be thought of as a numerical scaling of the x's.

The fundamental result of utility theory is that the <u>expected value</u> of the π's can also be used to numerically scale probability distributions over the x's. To illustrate the reasoning, let us reconsider the choice between act a' (which results in x_i with probability p_i') and act a" (which results in x_i with probability p_i''). If we associate to each x_i its scaled π_i value then the expected π-scores for acts a' and a"--let us label these by $\bar{\pi}'$ and $\bar{\pi}''$-- are

$$\bar{\pi}' = \Sigma_i p_i' \pi_i$$

and

$$\bar{\pi}'' = \Sigma_i p_i'' \pi_i \quad .$$

There are compelling reasons for the decision maker to rank order act a' and a" in terms of the magnitudes of $\bar{\pi}'$ and $\bar{\pi}''$. The argument briefly follows: Take act a'. It results with probability p_i' in consequence x_i. But x_i in turn is considered by the decision maker as indifferent to a π_i chance at x_n and complementary chance at x_1. So in effect act a_i' is equivalent to giving the decision maker a $\bar{\pi}'$ chance at x_n and a complementary chance at x_1. Similarly a" yields a chance of $\bar{\pi}''$ at x_n and a complementary chance at x_1. This completes the argument which rests heavily on the <u>substitution</u> of the risky option $< x_n, \pi_i, x_1 >$ for each x_i. The pros and cons of this substitution idea,which lies at the core of utility theory, are discussed in Raiffa [1968].

Now if we transform the π's into u's by means of a positive linear transformation

$$u_i = a + b\pi_i \ , \quad b > o \ , \quad i = 1,\ldots,n \ ,$$

then we have

$$u_1 < u_2 < \ldots < u_n$$

and it is easy to see that for probabilistic choice (such as between a' and a") the expected u values rank order a' and a" the same way as the expected π values. For example,

$$\bar{u}' = \Sigma_i \; p_i' \; u_i = \Sigma_i \; p_i'(a + b \; \pi_i) = a + b \; \bar{\pi}' \ .$$

If, however, we were to transform the π's into a new scale--call it w--by a monotone transformation other than a positive linear transformation, then the w's would reflect preferences for the simple consequence $x_1, x_2, \ldots, x_n$ but would not necessarily reflect preferences for probabilistic alternatives such as a' and a".

If one is sold on the merits of the above argument as we, the authors, are then the critical issue becomes: How can one assess, in a responsible manner, appropriate π-values? This is really the essence of our problem. If the x's are themselves scalars there are, as we shall see in this chapter, ways of thinking about the assessment problem, which exploit this underlying structure. If the x's are multidimensional vectors, we will in subsequent chapters describe techniques for structuring the assessment problem.

4.1.2 Alternative Approaches to the Risky Choice Problem

Does the decision maker need the full power of utility theory to make choices amongst risky alternatives? Can he get by, in practice, with less formal machinery, or can he circumvent the use of subjective judgements altogether and use more objective measures like means and variances?

Of course, in special cases one can get by with less paraphernalia than is needed for the maximization of expected utility. Suppose the possible impacts of two alternatives A and B can be described by the probability density functions f_A and f_B in Fig. 4.2A or alternatively by the cumulative probability distributions in Fig. 4.2B, where we have denoted the attribute of importance as X. Let F_A and F_B denote the cumulative distribution functions of A and B respectively. Notice from Fig. 4.2B, the proba-

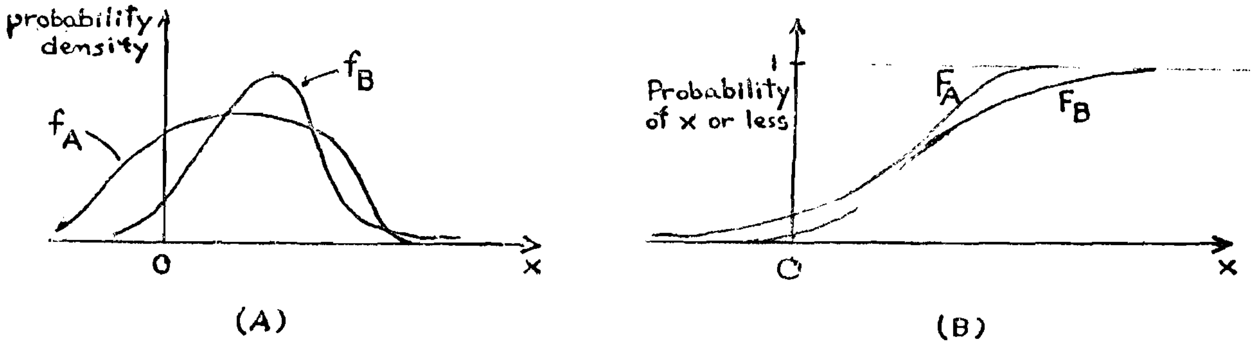


Figure 4.2. An Illustration of Probabilistic Dominance

bility that any outcome is x or less is greater for alternative A than for alternative B. Thus, if we just knew, for instance, that more of X is preferred to less of X, it would be appropriate to conclude that B should be preferred to A. In such a case we say alternative A is <u>probabilistically dominated</u> by alternative B. When such situations occur, one can use less information than contained in the complete utility function over X to make responsible, consistent decisions. This conclusion would not be readily apparant, however from Figure 4.2A. Of course, one is always not so lucky to be able to invoke probabilistic dominance.

There are cases where two cumulative distribution functions F_A and F_B for alternatives A and B intersect (so that no probabilistic dominance is present) but where a bit of subjective informal common sense might help one to make a choice without much ado. Often one merely has to look at F_A and F_B and without any formal procedures whatsoever come to a comfortable decision. But this again depends usually on extreme differences. Life is often more complicated and the choice is not readily apparant. One would <u>like to</u> probe one's basic feelings more systematically--and here, of course, the full power of utility theory comes to the fore. But let's look first at some so-called objective procedures.

One simple proposal is to use the <u>expected value</u> of the uncertain outcome as a guide. Here one requires only a knowledge of the probability distributions to calculate the expected value for each alternative. For certain problems, this may be appropriate. However, many decision makers would probably not be different between the following acts:

act A ≡ earn \$ 100,000 for sure,

act B ≡ earn \$ 200,000 or \$ 0, each with probability 0.5,

act C ≡ earn \$ 1,000,000 with probability 0.1 or \$0 with probability 0.9,

act D ≡ earn \$ 200,000 with probability 0.9 or lose \$ 800,000 with probability 0.1.

Notice that for each of the acts, the expected amount earned is exactly \$ 100,000, and so the expected value of the consequence would not be an appropriate criterion for a decision maker with a preference among these acts.

A possible criticism of this illustration might point out that "Naturally act A is preferred to the others since there is no uncertainty associated with the outcome. However, if a measure of uncertainty, such as the variance of the possible outcome, was used in addition to the expected outcome, we should be able to correctly order preferences for alternatives." This claim seems plausible but it is not always correct. Simple calculation will show that both acts C and D above have the same expected outcomes and variances and hence, any evaluation scheme based on just the mean and variance of the outcome would necessarily imply indifference between acts C and D. Various investigations have indicated that many people do have a preference between C and D, and thus, no mean-variance criterion can correctly represent their preferences.

Even if some mean-variance criterion seems appropriate for evaluating alternatives in a specific problem, one has to establish an appropriate preference order over the two attributes "expected outcome" and "outcome variance." This task, which may require assessing a value function over these two attributes, could be more involved than originally assessing a utility function over the single outcome attribute.

There are a myriad of other ad hoc schemes that can be found in the literature, but to our mind, no proposal other than maximization of expected utility withstands the scrutiny of careful examination. Let us cite one further proposal there. Let the uncertain outcome resulting from a given alternative be denoted by $\tilde{x}$. This proposal suggests that the distribution of $\tilde{x}$ be summarized by two indices:

a. $\alpha = P\left[\tilde{x} \leq x_o\right]$, the probability that $\tilde{x}$ is less than some critical aspiration level x_o;

and

b. $\beta = E\left[\tilde{x} \mid \tilde{x} \geq x_o\right]$, the conditional expectation of $\tilde{x}$, given that $\tilde{x}$ attains the aspiration level x_o.

The analyst can then compute the pair (α, β) for each alternative and set up a simple two-dimensional value function. For example, one might want to maximize β subject to the condition that $\alpha \leq .05$. Ad hoc procedures of this kind can be easily destroyed by citing extreme examples but then the retort usually is: "Oh, in such extreme examples we would modify our (α, β) proposal by imposing another constraint such as ..." There have been endless debates of this kind in the literature and suffice it to say, here, that we authors become more and more committed to the principle of maximization of expected utility, the younger we get and the more arguments we hear. Of course, this in itself should not be a compelling argument to you but we are reporting what we evidently feel is a relevant empirical fact.

4.1.3 Relevance of Unidimensional Utility Theory to Multiattribute Problems

Our motivation stated above for introducing unidimensional utility theory concerns mainly the usefulness for the concept of utility itself and relies on the fact that this usefulness can be easily illustrated with the unidimensional case. There is another very important reason. Namely, in many of the techniques we shall describe for assessing multiattribute utility functions, an essential component part is the assessment of unidimensional utility functions over single attributes. That is, our procedures often provide a basis for reducing the problem of assessing a multiattribute utility function into one of assessing some consistently scaled unidimensional utility functions. A thorough knowledge of unidimensional utility theory is needed for this latter task.

For instance, although the consequences of a problem may only be adequately described in terms of n attributes, it may be possible using the techniques discussed in Chapter 3 to reduce the dimensionality of the attribute space from n to (n-1). If $n = 2$, we then have a unidimensional problem. If $n > 2$, successive reduction of the dimensionality may lead us to the unidimensional case.

In Chapter 3, the techniques discussed suggested procedures for obtaining a value function $v(\underline{x})$ for all possible outcomes $\underline{x}$. Since value is unidimensional and $v(\underline{x}') = v(\underline{x}'')$ if and only if $\underline{x}'$ and $\underline{x}''$ are equally prefereable, it is appropriate to assess a utility function $u[v(\underline{x})]$ over the unidimensional attribute "value" and thus associate a utility with each possible consequence $\underline{x}$. The exact manner in which this is done is discussed in Chapter 5.

An alternative approach which does not require a value function in multiattribute situations is to verify assumptions implying a specific form of the utility function. The simplest example of this in two dimensions is the additive utility function $u(y,z) = u_Y(y) + u_Z(z)$, where u_Y and u_Z are consistently scaled unidimensional utility functions. The point is that both $u_Y(y)$ and $u_Z(z)$ can be assessed using the techniques discussed in this chapter.

The assumptions needed to justify an additive form such as

$$u(x_1,\ldots,x_n) = \sum_i k_i u_i(x_i)$$

or various multiplicative forms, such as

$$u(x_1,\ldots,x_n) = \prod_i [\alpha_i + \beta_i u_i(x_i)] ,$$

require various utility independence assumptions to be introduced in the sequel. However, even in cases where such independence does not hold we shall often have to introduce conditional univariate utility functions, such as: the conditional utility of x_i given that a summary index Y is at level y^o, say.

In summary, we can state that univariate utility functions will be an essential ingredient in all the multivariate theory to be developed in ensuing chapters.

4.1.4 Examples of Unidimensional Decision Problems

Let us cite some examples where one attribute might adequately summarize consequences for decisional purposes. A company's objective is to maximize profits. In this case, the attribute chosen to describe consequences might be incremental cash flow, or monetary asset position, or net monetary profit, etc. One attribute may be better than another in the sense that the decision maker can more easily express his preferences over different amounts of that attribute. The choice of which attribute to use is obviously subjective and left largely to the discretion of the analyst with consultation of the decision maker.*

---

* In business contexts, it is often preferable to use assets rather than incremental flows because it helps avoid some idiosyncratic behavior (e.g. the zero illusion) in the assessment procedures, and also it is easier to examine dynamic problems. See Schlaifer [1969], pages 163 - 165.

The question of how to choose an attribute, whether or not it is sufficient to describe consequences, etc., was discussed in detail in Chapter 2.

Many of the concepts covered in this chapter will use money as the unidimenional attribute. The main reasons for this are (1) many of the past interests and results in utility theory deal with this special case, and (2) most readers have already thought about or could think about their preferences for various amounts of money. Hence, a better intuitive feeling for the concepts of preference and risk introduced in the chapter will likely be developed using money as the primary attribute than would be the case if a less familiar attribute were used in illustrations. However, the concepts to be introduced are relevant to other unidimensional problems of importance. Let us indicate a few examples.

The emergency services, such as ambulance, police, fire, etc., respond to requests for help by dispatching an emergency vehicle (ambulance, etc.) to the scene "as soon as possible." An obvious choice of a measure of effectiveness in this case is <u>response time</u>, the elapsed time from receipt of the call requesting help to arrival on the scene of an emergency vehicle. Larson [1972] and Savas [1969] have chosen this attribute in some of their work on police systems and ambulance services, respectively.

In many queuing situations, whether it involves automobiles at toll booths or customers at a checkout counter, the objective is good service and this might be measured in terms of the attribute "delay time." Another queuing problem concerns the congestion occurring at the major airports. With this situation, a prime objective of the people responsible for operating these airports is the efficiency of runway operations. Blumstein [1959], Odoni [1972], and others have built analytical models of landing and departure operations, and they measure effectiveness of the various operating policies in terms of "the number of runway operations per hour," and this becomes the single attribute of concern.

In a medical context, basic univariate attributes might be: the cure-rate of some medical treatment; the number of severe side effects that result from use of a drug; and so on.

As a final example, we consider the following unpleasant situation: A country is seized by an epidemic, and the medical director of the country must choose an alternative for curtailing the death caused by this epidemic. An attribute which would describe the consequences of his actions might be number of deaths caused by the epidemic. In a variation of this problem the underlying attribute might be the probability that "severe consequences" will occur.

4.1.5 Outline of the Chapter

In the next section, a [illegible] method of assigning utilities to consequences is presented. This method is not operationally adequate when there are many consequences, since the direct method requires a subjective input from the decision maker for each assignment of utility and there may be too many such inputs to handle practically. In these situations, it may be desirable and necessary to construct a utility function u which assigns a utility u(x) to any possible consequence x over a continuous range of possibilities. Sections 4.3 through 4.7 develop a framework for examining monotonically increasing utility functions defined on a real-valued variable--that is, for cases where more of the variable is preferred to less. This framework is extended to decreasing and nonmonotonic utility functions in Section 4.8. The next two sections respectively, suggest a procedure for assessing unidimensional utility functions and report examples of such assessments. Section 4.11 and 4.12 extend the ideas of the chapter to conditional unidimensional utility theory and provide a transition to the multiattribute case considered in Chapters 5 and 6.

4.2 DIRECT ASSESSMENT OF UTILITIES FOR CONSEQUENCES

Let us denote the possible consequences of a decision as $x_1, x_2, \ldots, x_N$. Then, because utility is relative and not absolute, to establish an origin and unit of measure, we can arbitrarily assign utilities to two of the consequences and then assess utilities for the other consequences relative to those two. This procedure is probably easier to illustrate if we define x^o and x^* as a least preferred and a most preferred consequence. The use of "a least preferred" rather than "the least preferred" indicates there could be more than one consequence with the same degree of preference.

Now, to set our scale, let us assign

$$u(x^*) = 1 \text{ and } u(x^o) = 0 ,$$

and assess for each other consequence x, a probability π such that x is indifferent to the lottery $\langle x^*, \pi, x^o \rangle$, yielding a π chance at x^* and a $(1-\pi)$ chance at x^o. Then, because the utility of x must equal the expected utility of the lottery, we assign

$$u(x) = \pi u(x^*) + (1-\pi) \;\; u(x^o) = \pi .$$

If utilities were assessed in this manner for all x's, there would be many possible consistency checks. For instance, let x', x'', and x''' designate an increasing preference sequence and let the alternative x'' for certain be indifferent* to the lottery $\langle x''', p, x' \rangle$; then for consistency, p must be such that

$$u(x'') = pu(x''') + (1-p)u(x'),$$

or

$$p = \frac{u(x'')-u(x')}{u(x''')-u(x')} .$$

---

*Less elliptically we should say "Let the decision maker be indifferent between x'' and the lottery."

In problems with only a few possible consequences--maybe even up to fifty--this direct assessment technique may be appropriate. However, we feel that in problems with many consequences, where there is a natural ordering for the underlying x's, an alternate approach is often better. The procedure involves fixing the utilities of a few consequences as above and then fitting a curve--that is a utilitiy function-- to these. As we will see in the next five sections, the shape and functional form of the utility function tells us very much about the basic attitudes toward risk of the decision maker. Hence, our general approach is to start with these basic attitudes toward risk, to establish functional forms of utility functions exhibiting these properties, and then to choose a specific utility function using a few assessed points. This will become clearer in the sequel.

4.3 UNIDIMENSIONAL UTILITY FUNCTIONS

Let us introduce some qualitative characteristics of utility functions. Each characteristic implies a certain attitude of the decision maker with regard to his preferences for consequences and lotteries. By expressing these attitudes mathematically, restrictions on the utility function implied by these attitudes can be analytically derived. Provided the decision maker subscribes to a certain attitude, his utility function is restricted to a degree, and thus, the actual assessment of his utility function is simplified. Furthermore it then becomes possible to do sensitivity and break-even analyses.

4.3.1 Monotonicity

Often a very reaonable characteristic is monotonicity. For example, when monetary asset position is appropriate to summarize consequences, most (if not all) decision makers prefer a greater amount to a lesser amount. If we let x respresent the amount of monetary assets and u a utility function for such, the above statement is expressed mathematically by

$$[x_1 > x_2] \Rightarrow [u(x_1) > u(x_2)] \quad . \tag{4.3a}$$

Note that the converse of this is also true due to the nature in which utility functions are assessed. That is,

$$[u(x_1) > u(x_2)] \Rightarrow [x_1 > x_2] \quad . \tag{4.3b}$$

Let us now consider the preferences for response time to calls for ambulance service. It seems quite reasonable to assume a smaller response time is always preferred to a larger one. In this case, if t is a secific response time and u again represents the utility function,

$$[t_1 > t_2] \Leftrightarrow [u(t_2) > u(t_1)] \quad . \tag{4.4}$$

All this is to say the utility function for response time is <u>monotonically decreasing</u>.

It is interesting to note that one can easily transform from a decreasing to an increasing utility function by simply changing the attribute. For example, suppose that instead of measuring ambulance service in terms of response time, we define a "standard response time" as fifteen minutes and use the attribute "standard minutes saved in response" to measure service. For a particular call for service, if we let y be the standard minutes saved in response and define it by

$$y \equiv 15 - t ,$$

where t is the previously defined response time, then clearly preferences are increasing in y. And thus, the utility function for our new attribute Y is increasing. This is the case whether or not the "standard minutes saved in response" is ever negative (i.e. response time is greater than fifteen minutes).

Certainly it is clear that one can just as easily change from an increasing to a decreasing utility function by switching the measure of effectiveness. Perhaps the most intuitive example of this involves measuring the operations of a business concern in terms of either profits or opportunity losses. It is safe to assume that preferences are increasing in profits and decreasing in opportunity losses.

Let us suggest a situation where the utility function is not monotonic. In a medical context, a patient may be having problems with sugar in his blood. The doctor in charge may have a variety of alternatives to try to solve the problem. The blood sugar count may be used as the measure of effectiveness. There is some "normal" blood sugar count that is desired. Below the normal, the less the blood sugar count, the worse the situation is; above the normal, larger blood sugar counts are less preferred than smaller ones. In this case, preferences are monotonically increasing up to the normal level and monotonically decreasing afterwards. Such a utility function is illustrated in Fig. 4.3.

4.3.2 The Certainty Equivalent and Strategic Equivalence

The concept of the certainty equivalent is basic to utility theory. It is introduced now, since it will be frequently used in the presentation of the various risk characteristics of utility functions in the following sections.

Let L be a lottery yielding consequences $x_1, x_2, \ldots, x_n$ with probabilities $p_1, p_2, \ldots, p_n$ respectively. We will denote the uncertain consequence (i.e., a random variable) of the lottery by $\tilde{x}$ and the expected consequence by $\bar{x}$, where of course,

$$\bar{x} \equiv E(\tilde{x}) = \sum_{i=1}^{n} p_i x_i \; . \tag{4.5}$$

The expected utility of this lottery is

$$E[u(\tilde{x})] = \sum_{i=1}^{n} p_i \, u(x_i) \; , \tag{4.6}$$

which is an appropriate index to maximize in choosing among lotteries.

Definition: A certainty equivalent of lottery L is an amount $\hat{x}$ such that the decision maker is indifferent be-

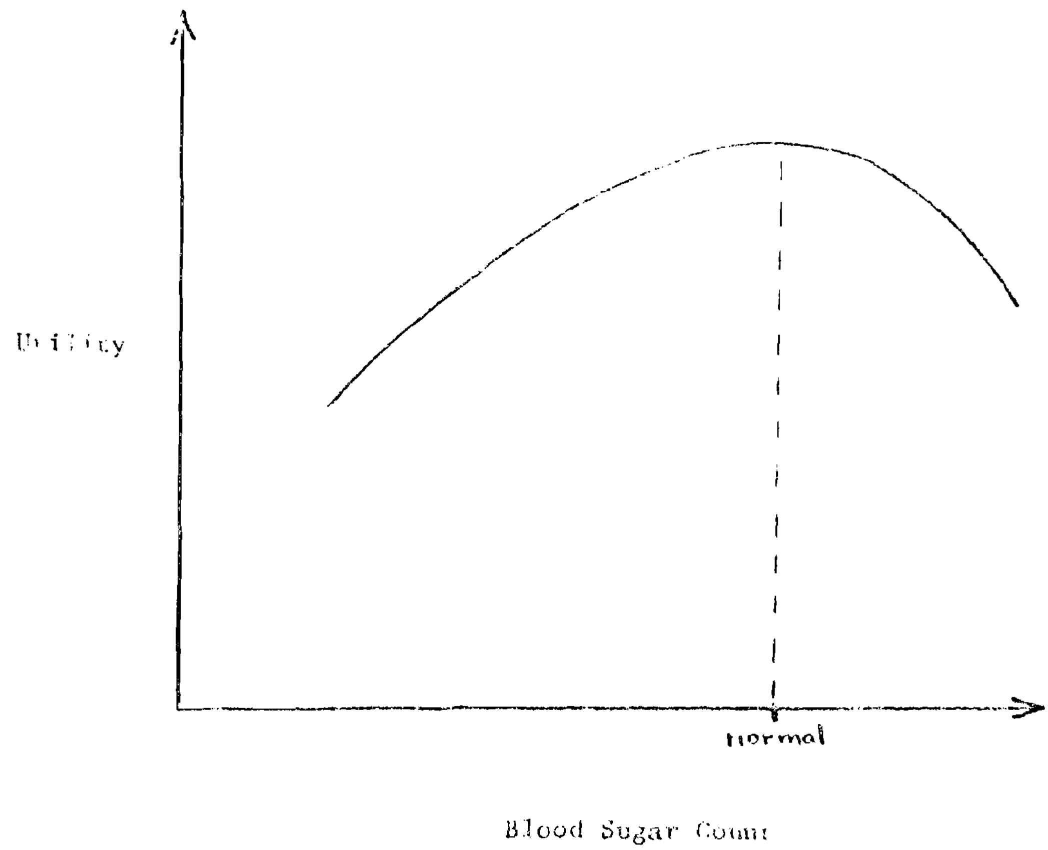


Figure 4.3 A Nonmonotonic Utility Function

tween L and the amount $\hat{x}$ for certain. Thus, $\hat{x}$ is defined by

$$u(\hat{x}) = E[u(\tilde{x})] \quad . \qquad (4.7)$$

Note that the certainty equivalent of any lottery is unique for monotonic utility functions.

When the attribute X of interest is monetary asset position, then a certainty equivalent of a lottery is referred to as a certainty monetary-asset equivalent. If X is a response time, the certainty equivalents are more appropriately called certainty response-time equivalents. However, since it will always be clear from the context of the discussion, we choose to use just the term "certainty equivalent" without further specification.

Historically, much of the development of unidimensional utility theory and thus, certainty equivalents, has been concerned with the utility for money. For this case, the terms <u>cash equivalent</u> and <u>selling price</u> of a lottery are often found in the literature. Both terms mean the certainty equivalent of a lottery with consequences representing monetary amounts*.

---

*The <u>buying price</u> of a lottery with monetary consequences is another term frequently found in the literature. It is defined as the largest amount of money the decision maker would pay for a lottery given his present asset position. Only in special circumstances is the buying price equal to the selling price of a lottery. See Chapter 4, Section 11 of Raiffa [1968].

Although it is perhaps obvious, the following point must be made. Notice that the expected consequence and certainty equivalent defined by (4.5) and (4.7) respectively were concerned with a lottery having a discrete number of possible consequences. When the possible consequences of a lottery are described by a probability density function f, then the expected conseqeuence $\bar{x}$ of that lottery is clearly

$$\bar{x} = E(\tilde{x}) = \int xf(x)dx \tag{4.8}$$

and a certainty equivalent $\hat{x}$ is the solution to

$$u(\hat{x}) = E[u(\tilde{x})] = \int u(x)f(x)dx. \tag{4.9}$$

Before presenting some examples, it is important to introduce the concept of strategic equivalence.

Definition: Two utility functions, u_1 and u_2, are strategically equivalent, written $u_1 \sim u_2$, if and only if there exisits constants h and k > 0 such that

$$u_1(x) = h + ku_2(x) \quad , \text{for all } x \, . \tag{4.10}$$

For example, $-e^{-2x}$ and $6 - 13e^{-2x}$ are strategically equivalent utility functions. It is easy to prove

Theorem 4.1: If $u_1 \sim u_2$, the certainty equivalents for any particular lottery implied by u_1 and u_2 are the same.

Proof. Assume (4.10) holds and let $\hat{x}$ be a certainty equivalent implied by u_1 for the lottery $\tilde{x}$. Using (4.9) and (4.10) in successive steps,

$$u_1(\hat{x}) = E[\,u_1(\tilde{x})\,] = E[h + ku_2(\tilde{x})]$$
$$= h + k\,E[u_2(\tilde{x})] \quad .$$

But from (4.10),

$$u_1(\hat{x}) = h + ku_2(\hat{x}) .$$

From the two previous equations,

$$u_2(\hat{x}) = E[u_2(\tilde{x})]$$

so that $\hat{x}$ is the certainty equivalent of $\tilde{x}$ using u_2. ◁

As an immediate consequence of the above assertion, we have the following corollary.

Corollary: <u>If lottery $\tilde{x}_1$ is preferred to lottery $\tilde{x}_2$ using utility function u_1, the same preference will hold using any strategically equivalent utility function u_2.</u>

<u>Proof:</u> Given $E[u_1(\tilde{x}_1)] > E[u_1(\tilde{x}_2)]$, it follows from (4.10) that $E[u_2(\tilde{x}_1)] > E[u_2(\tilde{x}_2)]$. ◁

This result means that strategically equivalent utility functions have identical implications for action. Let us present some examples.

<u>Example 4.1</u> Let $u(x) = a + bx \sim x$, $b > 0$. Suppose the decision maker is faced with a lottery described by the probability density function f. Then the expected consequence is

$$\bar{x} = E[\tilde{x}] = \int x \; f(x) \; dx \quad ,$$

and the certainty equivalent $\hat{x}$ is found from

$$u(\hat{x}) = E[u(\tilde{x})] = E[a + b\tilde{x}] = a + b\bar{x} .$$

Since $u(\hat{x}) = a + b\hat{x}$ it follows that $\hat{x} = \bar{x}$. This example shows that if the utility function is linear, the certainty equivalent for any lottery is equal to the expected consequence of that lottery. □

Example 4.2 Let $u(x) = a - be^{-cx} \sim -e^{-cx}$, where $b > 0$, and suppose the decision maker is faced with a 50-50 lottery yielding either x_1 or x_2, written $\langle x_1, x_2 \rangle$. The expected consequence $\bar{x}$ is $(x_1 + x_2)/2$. The certainty equivalent is the solution to

$$u(\hat{x}) = E[u(\tilde{x})] ,$$

or equivalently, the solution to

$$-e^{-c\hat{x}} = -\frac{e^{-cx_1} + e^{-cx_2}}{2} .$$

Table 4.1 exhibits some certainty equivalents and expected outcomes for a few $\langle x_1, x_2 \rangle$ lotteries given $u(x) = -e^{-cx}$.

Table 4.1 Certainty Equivalents for Lotteries $\langle x_1, x_2 \rangle$ Using $u(x) = -e^{-cx}$

c	x_1	x_2	$\bar{x}$	$\hat{x}$
1	0	10	5	0.69
1	10	20	15	10.69
1	20	30	25	20.69
0.2	0	10	5	2.85
0.2	10	20	15	12.85
0.2	20	30	25	22.85
0.1	0	10	5	3.8
0.1	10	20	15	13.8
0.1	20	30	25	23.8

Now suppose the lottery is described by the uniform probability density function

$$f(x) = \begin{cases} \dfrac{1}{x_2 - x_1} , & x_1 \le x \le x_2 \\ 0 , & \text{elsewhere} \end{cases} .$$

The expected consequence is clearly $(x_1 + x_2)/2$ and the certainty equivalent is found by solving

$$u(\hat{x}) = \int_{x_1}^{x_2} (-e^{-cx}) \left(\frac{1}{x_2 - x_1}\right) dx$$

or

$$-e^{-c\hat{x}} = \left(\frac{1}{x_2 - x_1}\right) \left(\frac{e^{-cx_2} - e^{-cx_1}}{c}\right) .$$

Completing the calculations for a few cases, we obtain Table 4.2. □

Table 4.2 Certainty Equivalents for Uniform Lotteries Using $u(x) = -e^{-cx}$

c	x_1	x_2	$\bar{x}$	$\hat{x}$
1	0	10	5	2.30
1	10	20	15	12.30
1	20	30	25	22.30
0.2	0	10	5	4.2
0.2	10	20	15	14.2
0.2	20	30	25	24.2
0.1	0	10	5	4.58
0.1	10	20	15	14.58
0.1	20	30	25	24.58

From Tables 4.1 and 4.2, it appears that if all the consequences of a lottery are increased by a fixed amount, the certainty equivalent is increased by that same amount. This is an important property of the exponential utility function.

Theorem 4.2 If $u(x) = -e^{-cx}$ and $\hat{x}$ is the certainty equivalent for a lottery $\tilde{x}$, then $\hat{x} + x_0$ is the certainty equivalent for the lottery $\tilde{x} + x_0$.

Proof. The certainty equivalent $\hat{x}'$ for the second lottery solves

$$-e^{-c\hat{x}'} = E[u(\tilde{x} + x_0)]$$
$$= E[-e^{-c(\tilde{x} + x_0)}]$$
$$= e^{-cx_0} E[-e^{-c\tilde{x}}].$$

But by definition,

$$-e^{-c\hat{x}} = E[-e^{-c\tilde{x}}],$$

so

$$-e^{-c\hat{x}'} = e^{-cx_0}(-e^{-c\hat{x}}) = -e^{-c(\hat{x} + x_0)}$$

from which it follows that $\hat{x}' = \hat{x} + x_0$. ◁

Example 4.3 Let $u(x) = \log(x + b)$, $x > -b$. The expected consequence for lottery $\langle x_1, x_2 \rangle$ is $(x_1 + x_2)/2$ as before. The certainty equivalent is the solution to

$$\log(\hat{x} + b) = \frac{\log(x_1 + b) + \log(x_2 + b)}{2},$$

which is

$$\hat{x} = \sqrt{(x_1 + b)(x_2 + b)} - b$$

A few cases are cataloged in Table 4.3. □

Table 4.3 Certainty Equivalents for Lotteries $\langle x_1, x_2 \rangle$ Using $u(x) = \log(x + b)$

b	x_1	x_2	$\bar{x}$	$\hat{x}$
1	0	10	5	2.32
1	10	20	15	14.2
1	20	30	25	24.5
11	0	10	5	4.2
11	10	20	15	14.5
11	20	30	25	24.7
21	0	10	5	4.5
21	10	20	15	14.7
21	20	30	25	24.8

One can see from this Table 4.3 for every lottery the certainty equivalent is always less than the expected consequence. However, for any particular value of b, this difference grows smaller as the consequences x_1 and x_2 are increased by a fixed amount. Later in this section, we shall devote a good deal of attention to utility functions which imply such a behavior for the certainty equivalents.

<u>Example 4.4</u> The first three examples have all concerned monotonically increasing utility functions. Let us consider the decreasing utility function i.e. $u(x) = -x^2$, $x \geqslant 0$, and calculate the expected consequences and certainty equivalents for $< 0,10 >$ and $< 10,20 >$. The expected consequences are clearly 5 and 15, respectively. The certainty equivalent for $< 0,10 >$ is the solution to

$$-\hat{x}^2 = -\frac{0^2 + 10^2}{2} = -50 \ .$$

Thus, $\hat{x} = 7.07$. Similarly the certainty equivalent for $< 10,20 >$ is found to be 15.8. This means that the decision maker is indifferent between obtaining $x = 7.07$ for certain and the lottery $< 0,10 >$, and that he is indifferent between obtaining $x = 15.8$ for certain and the lottery $< 10,20 >$. $\sqcup$

By now you should feel at ease with the certainty equivalent. The examples have illustrated calculation of the certainty equivalents for some representative lotteries.

However, they dealt only with monotonic utility functions. But what about the monotonic case? In this situation, the certainty equivalent may not even be unique. Refer to Fig. 4.4 and consider a 50-50 lottery between x_1 and x_2. A certainty equivalent is any consequence whose utility equals the expected utility of the lottery $[u(x_1) + u(x_2)]/2$. As we can see from Figure 4.4, both x_3 and x_4 are certainty equivalents for $\langle x_1, x_2 \rangle$, and in fact, one of these does not even fall between the two possible consequences of the lottery.

4.4 RISK AVERSION

In this and the next four sections, we introduce various basic attitudes toward risk and illustrate their implications on the functional form of the utility function. In order to maintain a continuity in the presentation and to help the reader develop an intuitive understanding for these concepts, the sections through 4.7 concern only monotonically increasing utility functions. And, for the same reasons, much of our discussion will concern the cases involving a monetary attribute, such as "net assets" or "incremental income." However, as we have stressed earlier, the concepts are equally valid for nonmonetary attributes. Section 4.8 then extends the risk concepts to situations with decreasing and nonmonotonic preferences.

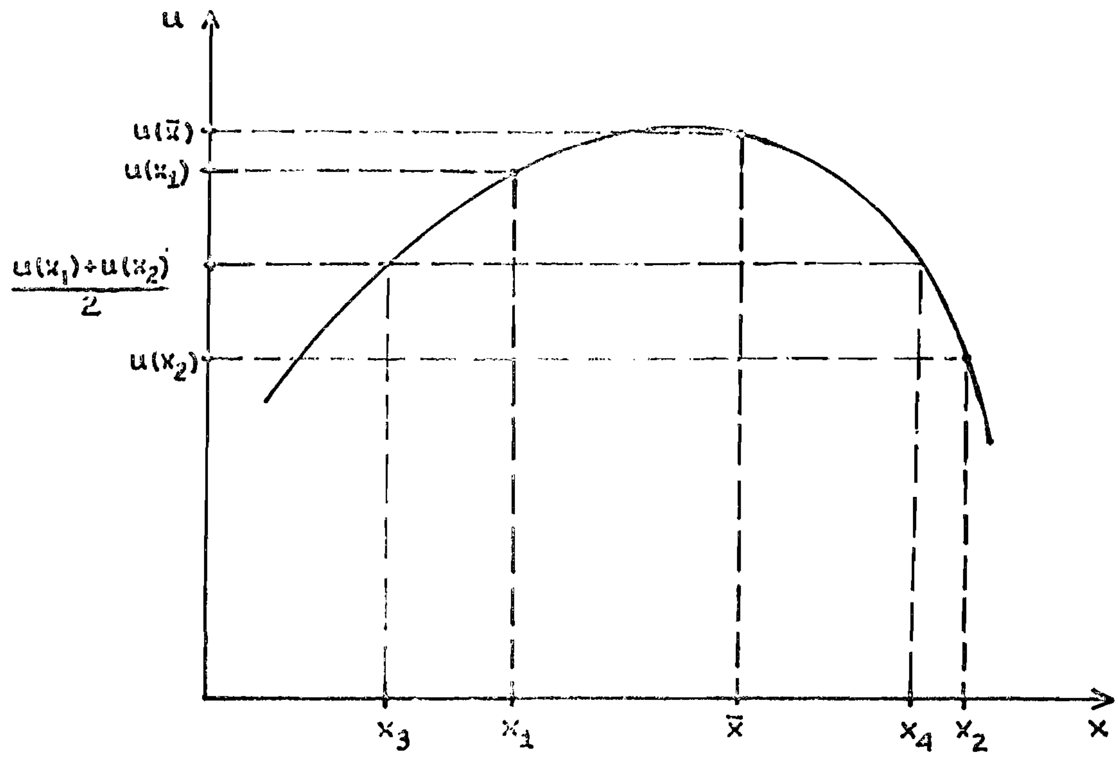


Fig 4.4. The Certainty Equivalents Using a Nonmonotonic Utility Function

4.4.1 A Definition of Risk Aversion

Intuitively, one thinks of a risk-averse person as one who prefers to behave conservatively. Consider a decision maker facing a lottery yielding either a consequence x' or a less prefereable consequence x'', with equal probability. Obviously, the expected consequence $\bar{x}$ of this lottery is $(x' + x'')/2$. Now suppose the decision maker is asked to state his preference between receiving $\bar{x}$ for certain and the lottery $\langle x', x'' \rangle$. If the decision maker prefers the certain consequence $\bar{x}$ to the lottery $\langle x', x'' \rangle$ with the same expected consequence, then the decision maker is actually saying he prefers to avoid the risks associated with the lottery. That is, although $\bar{x}$ and $\langle x', x'' \rangle$ have the same expected consequences, he prefers $\bar{x}$ since there is no risk associated with it, whereas there is risk associated with the outcome of the lottery. When a decision maker has this type of attitude toward all lotteries, we say he is risk-averse. Let us formalize this notion.

Definition: A decision maker is risk-averse if he prefers the expected consequence of any nondegenerate lottery to that lottery.*

---

*A nondegenerate lottery is one where no single consequence has a probability of one of occurring. It is an academic point, but had we allowed degenerate lotteries in definition (4.11), the "greater than" sign " $>$ ", would have to be replaced by the "greater than or equal" sign " $\geq$ ".

In such a situation, the utility of the expected consequence of any lottery must be greater than the expected utility of that lottery. So, if the possible consequences of a lottery are respresented by $\tilde{x}$, one is risk-averse if for all nondegenerate lotteries

$$u[E(\tilde{x})] > E[u(\tilde{x})] \quad . \tag{4.11}$$

It is easy to prove

Theorem 4.3 A decision maker is risk-averse if and only if his utility function is concave.

Proof: Consider a lotter which yields either x_1 with probability p or x_2 with probability $(1 - p)$, $0 < p < 1$. The expected consequence is $\bar{x} = px_1 + (1 - p)\, x_2$. For risk-averse utility functions, from (4.11)

$$u[px_1 + (1 - p)x_2] > pu(x_1) + (1-p)u(x_2) \quad ,$$
$$0 < p < 1 \quad ,$$

which is the definition of (strict) concavity.

To prove the converse (done only for the finite case), consider lottery $\tilde{x}$ yielding x_i with probability p_i, for $i = 1, \ldots, m$, where no $p_i = 1$. Since u is strictly concave, we know that

$$u[\sum_{i=1}^{m} p_i x_i] > \sum_{i=1}^{m} p_i u(x_i) \quad .$$

This inequality is just (4.11) for the finite case, so u is risk-averse. √

Operationally, it may be unrealistic to check condition (4.11) for all nondegenerate lotteries to determine whether or not one is risk-averse. To help matters, there is the related

Corollary: A decision maker who prefers the expected consequence of any 50-50 lottery $<x_1, x_2>$ to the lottery itself is risk-averse.

Proof. It follows from the premise that

$$u(\bar{x}) = u(\tfrac{1}{2}x_1 + \tfrac{1}{2}x_2) > \tfrac{1}{2}u(x_1) + \tfrac{1}{2}u(x_2), \text{all } x_1 \neq x_2$$

which implies u is concave. ◁

As one learns in every basic econmics course, the economist's concept of decreasing marginal utility implies the utility function is concave and conversely. Here, utility function is in italics because it is a completely different construction from the von Neumann-Morgenstern utility function which we are considereing in this chapter. The distinction seems important enough to deserve a brief digression.

When the econmist says "his marginal utility for attribute X is decreasing", he means that the increase in units of utility, called utiles--which are never explicitly defined--due to an incremental unit of X from x to x+1 decreases as x increases. No probabilistic notions are introduced and any expected utility calculated from such a utility function has no particular interpretation as it does in the case of von Neumann-Morgenstern utility functions.

As an example of our economist's utility function with decreasing marginal utility, suppose one considered 8 utiles as the utility of one day of skiing, 14 utiles for two days, 18 utiles for three days, etc. Then one could say the first day is worth 8 utiles, the second an additional 6, and the third another 4. The marginal utility of each additional day of skiing is decreasing. However, if this individual had a choice between two days of skiing for sure or a lottery yielding either 1 or 3 days with equal likelihood, one could not say which option should be preferred using the utility function. This is so even though the expected number of utiles for the lottery is 13, whereas it is 14 for the sure 2 days skiing. The concept of "expected utiles" has no meaning. The utility functions we are talking about in this chapter are completely different from the economist's utility function . Knowing one implies very little about the other. One can easily be convex and the other concave for the same attribute.

Let us return to our decision maker and suppose he did not wish to behave conservatively. In fact, suppose the decision maker preferred any lottery to the expected consequence of that lottery. That is to say he was more than willing to accept the risks associated with any lottery. This type of individual is said to be risk prone.

Definition: A decision maker is risk prone if he prefers any nondegenerate lottery to the expected consequence of that lottery.

For such an individual, the utility of the expected consequence must be less than the expected utility of the lottery, that is

$$u[E(\tilde{x})] < E[u(\tilde{x})]. \qquad (4.12)$$

Because of the similarity to the preceding assertion, we state without proof

Theorem 4.4: A decision maker is risk prone if and only if his utility function is convex.

There is an alternative way which we could have defined risk aversion for increasing utility functions. However, since this definition would not be valid for other cases, we chose (4.11) to define risk averse and to state the alternative as a fact. Let us illustrate this with

Theorem 4.5: For increasing utility functions, a decision maker is risk averse if and only if his certainty equivalent for any nondegenerate lottery is less than the expected consequence of that lottery.

Proof. Assume he is risk averse. Then from (4.11)

$$u[E(\tilde{x})] > E[u(\tilde{x})] \quad .$$

But by the definition of a certainty equivalent, we know

$$u(\hat{x}) = E[u(\tilde{x})] \quad ,$$

so

$$u[E(\tilde{x})] > u(\hat{x}) \quad .$$

Since the utility function is increasing, clearly

$$E(\tilde{x}) > \hat{x} \quad .$$

Now to go the other way, assume

$$E(\tilde{x}) > \hat{x} \quad .$$

Then, because the utility function is increasing,

$$u[E(\tilde{x})] > u(\hat{x}) = E[u(\tilde{x})] \quad ,$$

which completes the proof. $\surd$

For <u>increasing</u> utility functions we make the following

<u>Definition</u>: The <u>risk premium</u> RP of a lottery $\tilde{x}$ is its expected consequence minus its certainty equivalent. In symbols we have

$$RP(\tilde{x}) = \bar{x} - \hat{x} = E(\tilde{x}) - u^{-1}\{E[u(\tilde{x})]\} \quad ,$$

where u^{-1} is the inverse of u.

It is easy to show

<u>Theorem 4.6</u>: <u>For increasing utility functions, a decision maker is risk averse if and only if his risk premium is positive for all nondegenerate lotteries</u>.

The proof is omitted as it follows directly from the definition of the risk premium.

It may be illustrative to work through a couple of examples. Refer to Fig. 4.5 for an illustration of the certainty equivalent and risk premium for $<x_1, x_2>$ using a risk averse utility function.

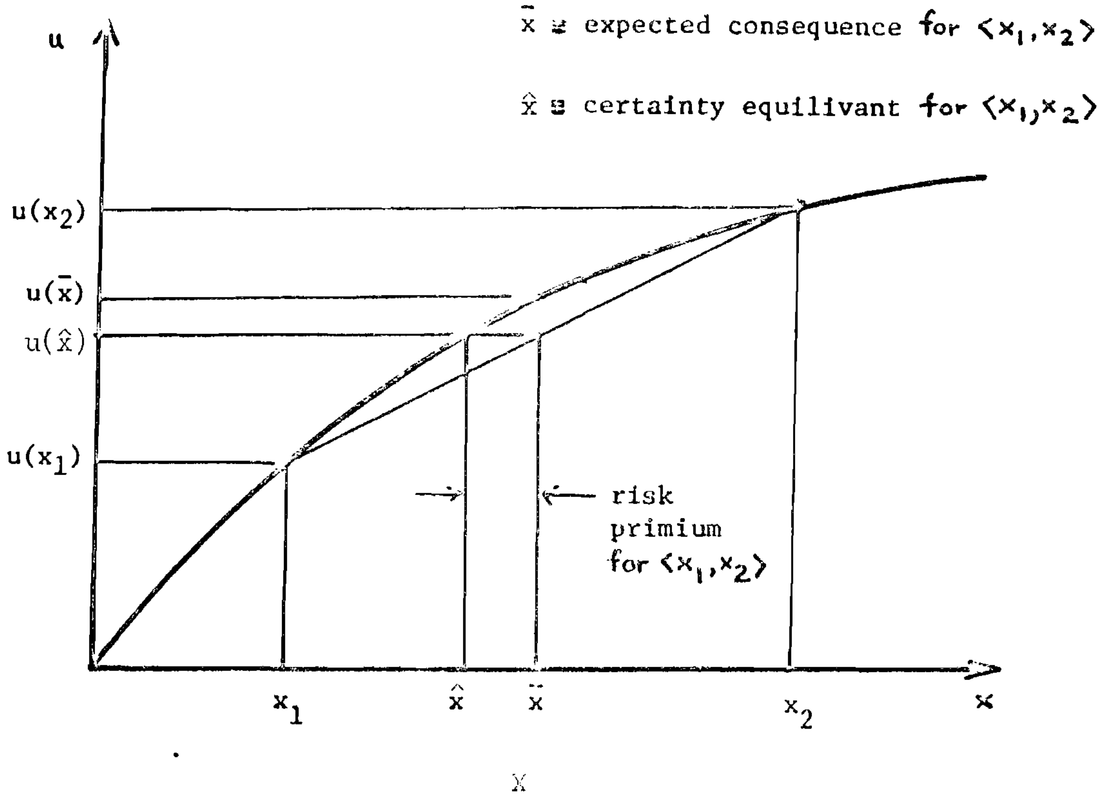


Figure 4.5. An Increasing Utility Function Exhibiting Risk Aversion

Example 4.5 From Table 4.1, we see that using the utility function $u(x) = -e^{-0.2x}$, the certainty equivalent for $\langle 0,10 \rangle$ is 2.85 and the expected consequence is 5.0. Thus, the risk premium is (5.0 - 2.85) or 2.15. Similarly, the certainty equivalent for $\langle 20,30 \rangle$ is 22.85 and the expected consequence is 25.0, so again the risk premium is 2.15. □

Example 4.6 Given $u(x) = -e^{0.2(20 - x)}$, $x \geq -20$, we find the certainty equivalents for $\langle -20,-10 \rangle$ and $\langle -10,0 \rangle$ to be -17.15 and -7.15 respectively. Their expected consequences are clearly -15.0 and -5.0. Thus, the risk premium for the first lottery is $[-15.0 -(-17.15)]$ or 2.15. Similarly, 2.15 is the risk premium for $\langle -10,0 \rangle$. □

Intuitively, the risk premium is the amount of the attribute that the decision maker is willing to "give up" from the average (i.e., the amount less than the expected consequence) to avoid the risks associated with the particular lottery.

When the decision maker is faced with an unfavorable lottery, that is, one which is less preferable than the status quo, it is natural to ask how much would he "pay" in terms of attribute X to avoid accepting this lottery. This leads us to make the following

Definition: The insurance premium IP for a lottery $\tilde{x}$ is the negative of the certainty equivalent of the lottery. In symbols

$$IP(\tilde{x}) = -\hat{x} = -u^{-1}\left(E[u(\tilde{x})]\right).$$

If, for example, the lottery $\tilde{x}$ has a certainty equivalent of -\$ 5,000 say, then the insurance premium is \$ 5,000. The decision maker should just be willing to give up \$ 5,000 to rid himself of the financial responsibility of the lottery.

In the last example, assume that x = 0 was equivalent to doing nothing, i.e. the status quo. Then <-20,-10> and <-10,0> are unfavorable lotteries since their expected utilities are less than the utility of the status quo. The decision maker was indifferent between <-20,-10> and its certainty equivalent -15.8. This means the decision maker should be willing to pay 15.8 to eliminate the responsibility of the lottery <-20,-10> . Thus, 15.8 is the insurance premium for <-20,-10>. Likewise, the decision maker should pay 5.8 to avoid <-10,0> so 5.8 is the insurance premium.

4.4.2 Restricting the Form of the Utility Function

Before going any deeper into the theory, let us illustrate how monotonicity and risk aversion can be exploited to greatly simplify the assessment of a utility function. Suppose we wish to assess a utility function u for attribute X, and the decision maker has indicated that his preferences increase monotonically in X and that he is risk averse.

To begin, we choose x_1 and x_2, where $x_2 > x_1$, and arbitrarily assign $u(x_1)$ and $u(x_2)$ subject to the restriction that $u(x_2) > u(x_1)$. This is permissible since utility functions are unique up to positive linear transformations. By plotting the points $[x_1, u(x_1)]$ and $[x_2, u(x_2)]$ on the graph in Figure 4.6, we can see the decision maker's utility function is limited to the nonshaded area. Consider point 3 in the figure. If the utility function passed through this point, then part of the function would necessarily not be concave. But since the decision maker is risk averse, his utility function must be concave therefore, it cannot pass through point 3. Similarly, if the decision maker's utility function passed through point 4, monotonicity would be violated since $x_4 > x_2$ and $u(x_4) < u(x_2)$.

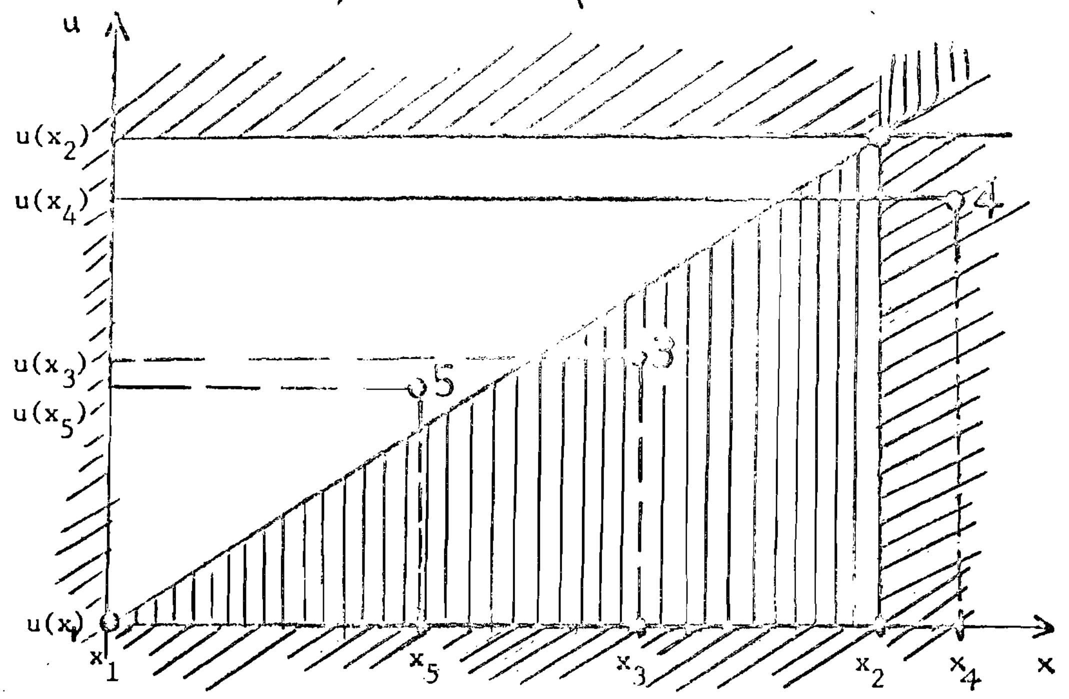
Now suppose we question the decision maker to find his certainty equivalent for the lottery yielding either x_1 or x_2, each with probability 1/2. Denoting this certainty equivalent by x_5, we have one additional point, $[x_5, u(x_5)]$, on the utility function, where

$$u(x_5) = \frac{u(x_1) + u(x_2)}{2} .$$

After plotting this point on the graph of Fig. 4.6A we use the same logic as before to restrict the decision maker's utility function to the nonshaded region of Fig. 4.6B. As can be seen from the figure, by empirically evaluating the utility of only one consequence, the shape of the utility function can be restricted quite a degree by exploiting the qualitative characteristics of monotonicity and risk aversion.

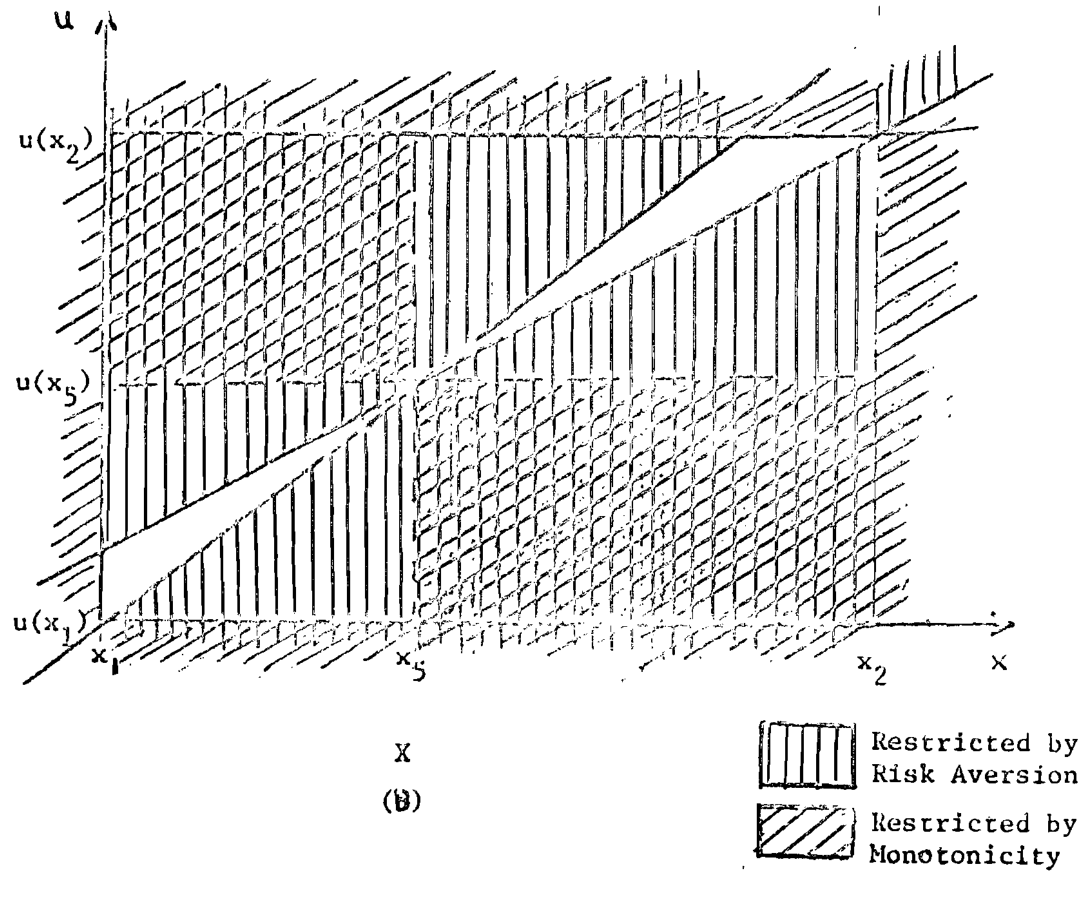
The same type of reasoning can be used to bound the

Figure 4.6. Restrictions Placed on a Utility Function by Monotonicity and Risk Aversion



X

(A)



X

(B)

certainty equivalents of a lottery. Perhaps this can best be illustrated with

Example 4.7 Let us say $x_1 = 0$, $x_2 = 100$, and $x_5 = 40$ in Fig. 4.6. Furthermore, assume we had arbitrarily set $u(0) = 0$ and $u(100) = 1$, so $u(40) = 0.5$. Then, as we can see in Fig. 4.7, by elementary geometric reasoning, any monotone, risk-averse utility function must lie between

$$u_1(x) = x/80$$

and

$$u_2(x) = 0.167 + x/120.$$

Suppose we want to bound the certainty equivalent for the lottery described by the probability density function f(x) where

$$f(x) = \begin{cases} 1/50 , & 25 \leq x \leq 75 , \\ 0 , & \text{elsewhere} . \end{cases}$$

To get an upper bound on the certainty equivalent for a lottery, in general we could obtain an upper bound on its expected utility and find the largest value of x which could possibly have this utility. Because of risk aversion, Theorem 4.5 implies the certainty equivalent can be no greater than 50. However, for the specific lottery, note from Fig. 4.7 that it is possible that the utility function be linear from x = 25 to x = 75. Since the probability density specifies possible outcomes only in this range, the certainty equivalent could be as high as 50, the expected outcome. Hence, the lowest upper bound on the certainty equivalent

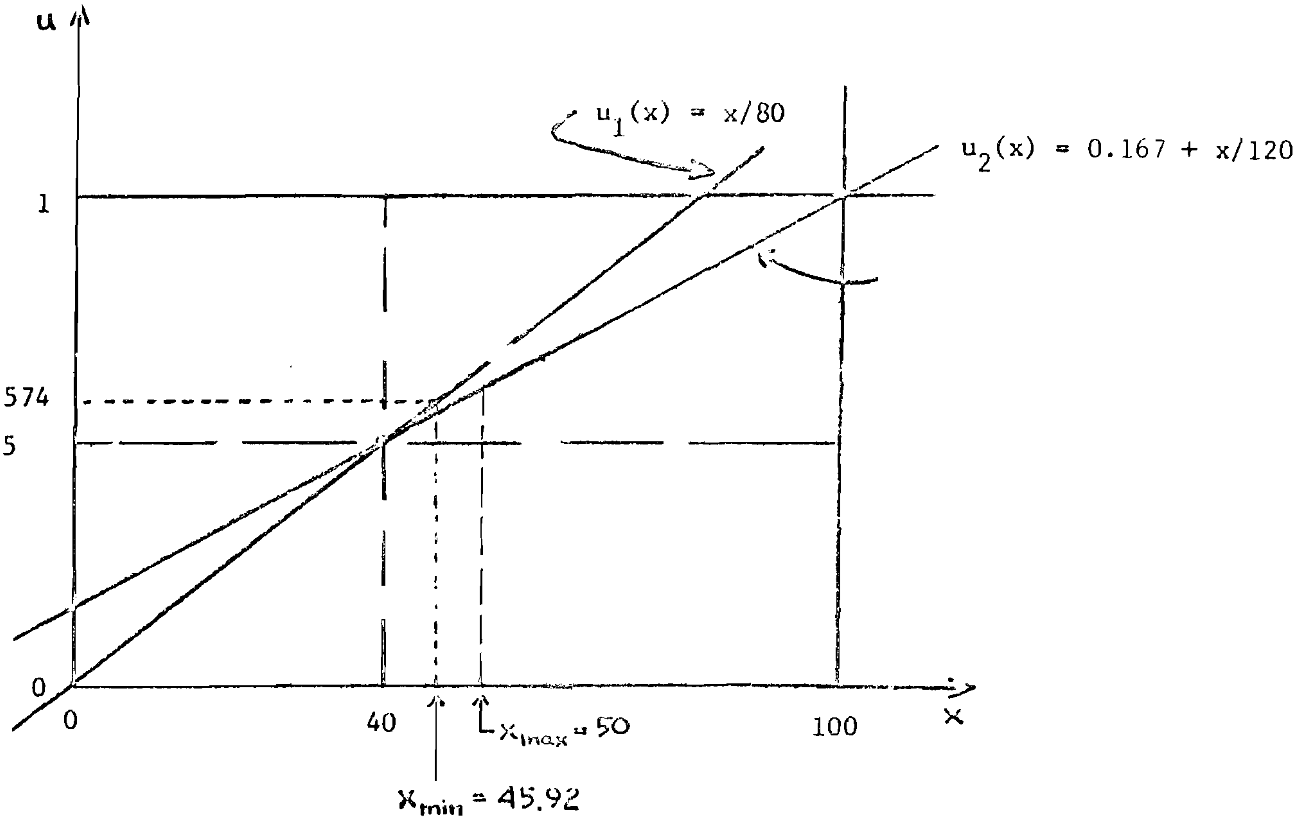


Figure 4.7 Bounding the Certainty Equivalent Using Risk Aversion and Monotonicity

for our lottery, call it x_{max}, is 50.0.

To find a lower bound on the certainty equivalent, we could first get a lower bound on the expected utility of this lottery, and then find the smallest value of x which could possibly have this utility. Clearly, regardless of what the actual utility function u is, the expected utility of the lottery

$$E[u(\tilde{x})] = \int_{25}^{75} u(x)\ f(x)\ dx$$

$$\geq \int_{25}^{40} u_1(x)\ f(x)\ dx + \int_{40}^{75} u_2(x)\ f(x)\ dx\ ,$$

so

$$E[u(\tilde{x})] \geq 0.122 + 0.452 = 0.574 \quad .$$

As can be seen from Fig. 4.7, the smallest possible amount x, call it x_{min}, which could have a utility equal to 0.574 results when $u(x) = u_1(x)$ and is found by solving

$$u_1(x_{min}) = x_{min}/80 = 0.574\ .$$

This gives us

$$x_{min} = 45.92 \quad .$$

and x_{min} is a lower bound on the "true" certainty equivalent of our lottery. It is not necessarily the greatest lower bound since x_{min} was calculated using $u = u_1$ in the range $x \geq 40$, whereas $u = u_2$ was used in this range to calculate the minimum utility for the given probability density. Hence, tighter bounds could probably be found. □

However, our purpose in this example was not to find the tightest possible bounds on the certainty equivalent but to illustrate how some rather powerful inferences can be made from a limited amount of information about a decision maker's preferences and to become more familiar with some of the concepts we will be using continually.

4.4.3 The Risk Prone Case*

Let us now take a look at the opposite of a risk-averse decision maker, that is, a risk-prone one. It is easy to prove

Theorem 4.7: For increasing utility functions, a decision maker is risk prone if and only if his certainty equivalent for any nondegenerate lottery is greater than the expected consequence of that lottery.

The proof is omitted because of its similarity to the corresponding proof for the risk averse case.

Recall that the risk premium was defined as the expected consequence minus the certainty equivalent for increasing utility functions. Following directly from this

---

*This subsection examines the risk prone case in a manner analogous to the risk averse case. It is included primarily for reference purposes and may be skipped.

definition, we have

Theorem 4.8: For increasing utility functions, a decision maker is risk prone if and only if his risk premium is negative for all nondegenerate lotteries.

The proof is omitted. However, let us illustrate this result with

Example 4.8 Consider a risk prone utility function of the form $u(x) = 0.2x^2$ as illustrated in Fig. 4.8 and let us calculate the expected consequence, certainty equivalent, and risk premium for the lottery $< 4, 12 >$. Clearly, the expected consequence is

$$\bar{x} = \frac{4 + 12}{2} = 8 \quad .$$

The expected utility of this lottery is

$$\frac{1}{2}\, u(4) + \frac{1}{2}\, u(12) = \frac{1}{2}\,(0.2 \times 16) + \frac{1}{2}(0.2 \times 144) = 16$$

so its certainty equivalent $\hat{x}$ is the solution to

$$0.2(\hat{x})^2 = 16 \quad .$$

Solving this, we find $\hat{x} = 8.94$. The risk problem, $\bar{x} - \hat{x}$ is then easily found to be -0.94. □

A risk prone individual is one who is "willing to gamble." In laboratory experiments and in operational situations in the "real-world," different researchers have found certain decision makers to to be risk prone. For instance, Grayson [1960], by measuring several oil wildcatters' utility functions for money, found some of them to have

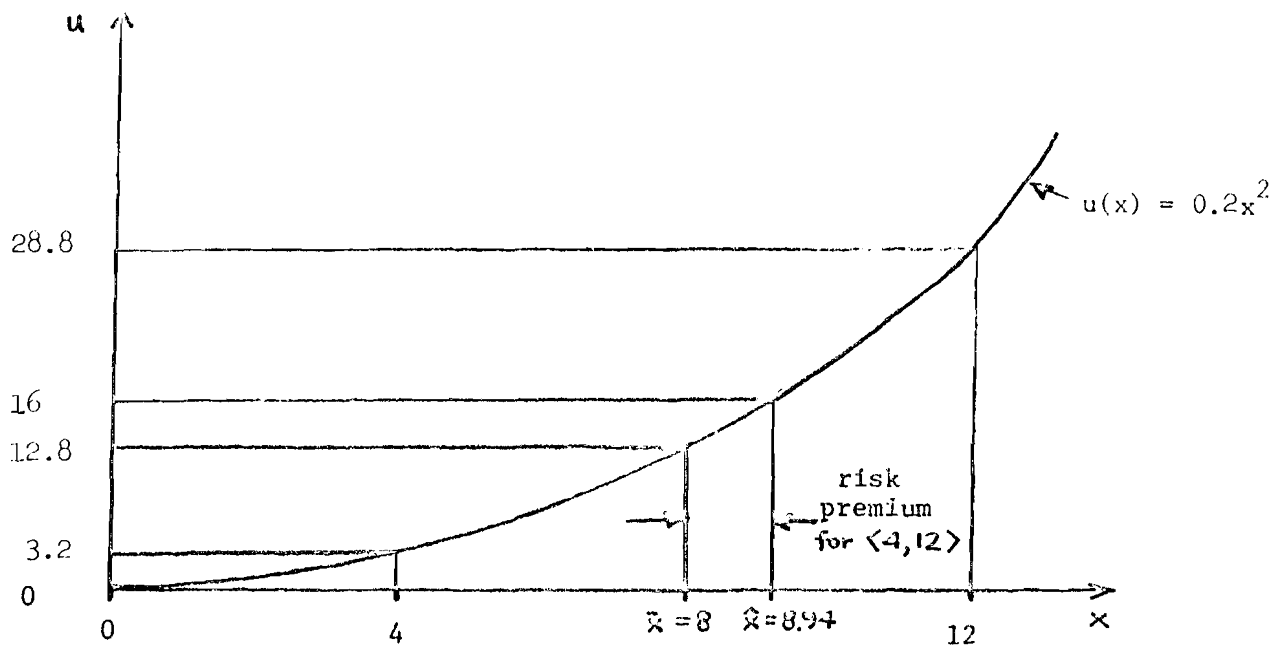


Figure 4.8 An Increasing Utility Function Illustrating Risk Proneness

this characteristic. In other words, these oil wildcatters were willing to risk their stakes on a lottery (i.e., drilling for oil) with an expected return less than their stakes, but which might result in a very large return (i.e., striking oil). This large return represented the opportunity for a "new way of life," and this made the gamble worth it to many wildcatters. Aspects of Grayson's work are discussed in Section 4.10.

Given that the decision maker's preferences are increasing, that he is risk-prone, and given the certainty equivalent for one 50-50 lottery, we could bound his utility function as we did for the risk-averse individual. Also, we could calculate bounds for the certainty equivalent of any other lotteries using the procedure illustrated in Example 4.7. However, since the ideas are similar to the previous case, another example would not be particularly illustrative, so we omit it.

4.5 A MEASURE OF RISK AVERSION*

Now that the usefulness of risk aversion is established, we will direct our attention toward a measure of this property for increasing utility functions. We would like a measure of risk aversion to indicate when one decision

---

*The reader is strongly urged to read Pratt [1964] which is the original source for much of what is discussed in this and the following two sections.

maker is more risk-averse than another in the sense that for any specified lottery, his risk premium is greater than that of the other decision maker.

Consider the lottery $\langle x + h, x - h \rangle$ where h is a specified amount of X. Intuitively, it seems the more concave the utility function u is about x, the larger the risk premium $\pi(x,h)$ for the lottery $\langle x + h, x - h \rangle$ will be. However, this notion is quickly dismissed by viewing Fig. 4.9. As can be seen, although u", the second derivative of u with respect to x, is different for the two utility functions, the risk premium is the same. Therefore, the magnitude of u" provides no insight into one's attitudes toward risk. With good hindsight, we can see that of course this is the case since utility functions which are positive linear transformations of each other are strategically equivalent.

The sign of u" does provide some information however. If u" is negative for all x, then u must be concave, and therefore risk-averse. On the other hand, if u" is positive for all x, then u is convex implying the decision maker is risk-prone. Thus it seems reasonable to take u" into account in some way in a measure of risk aversion.

Let us proceed in the same manner which led to the development of a measure of risk aversion. It seemed desirable that such a measure should, among other things,

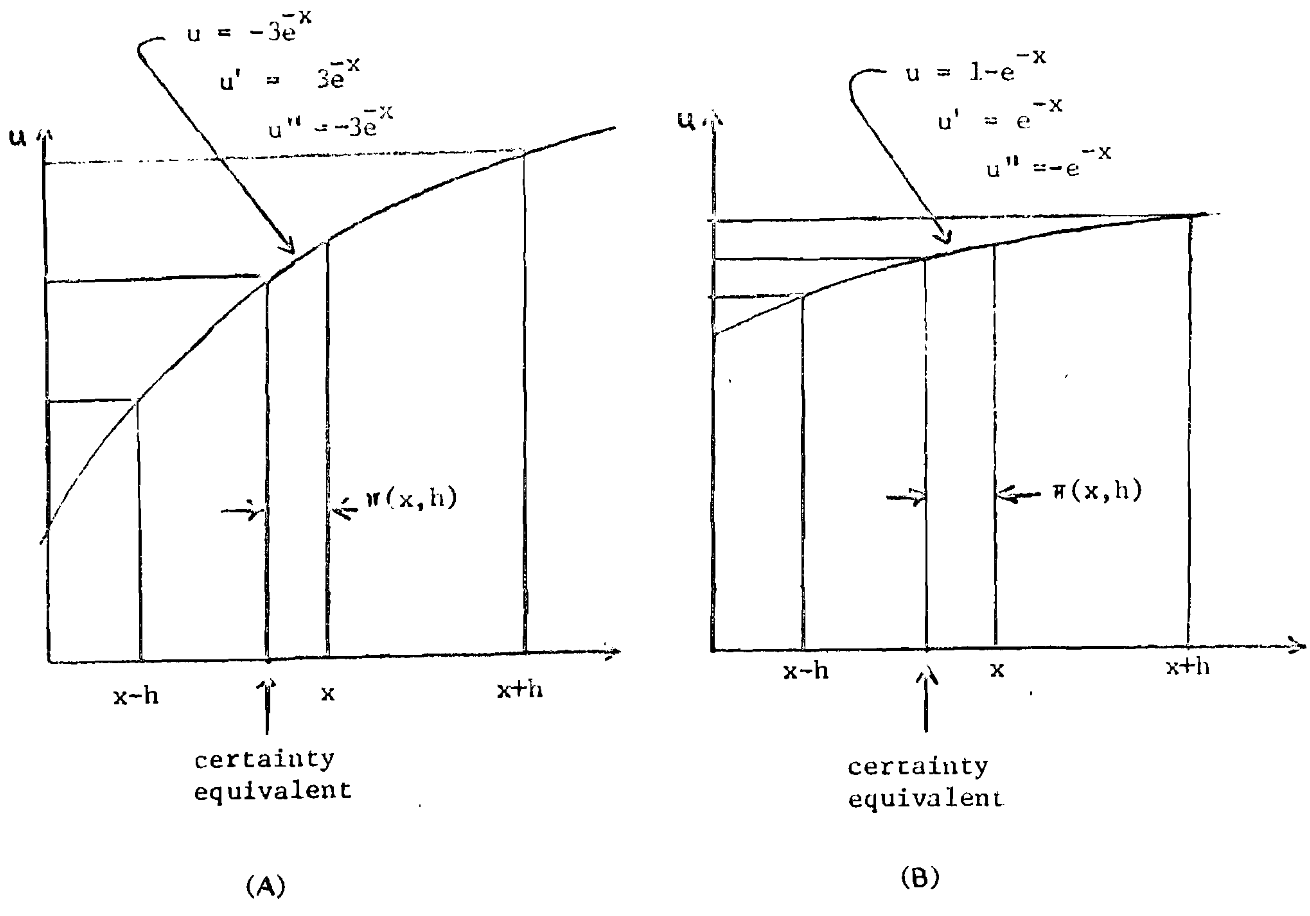


Figure 4.9

Two Utility Functions With The Same Risk Aversion Function

(1) indicate whether a utility function is risk-averse or risk-prone (which can be done with u'') and (2) be identical for strategically equivalent utility functions. Following this theme, for strategically equivalent utility functions u_1 and u_2, clearly $u_2 = d + ku_1$, so that $u_2' = ku_1'$, and $u_2'' = ku_1''$. From this, one can observe that $u_2''/u_2' = u_1''/u_1'$, and thus it seems that a relevant measure of one's aversion to risk might be the ratio of u'' and u'. This was tried and it was discovered that such a measure had many desirable properties. Many of these properties are stated in this section. With this motivation we introduce the following

Definition: The local risk aversion at x, written r(x), is defined by

$$r(x) \equiv - \frac{u''(x)}{u'(x)} \quad . \tag{4.13}$$

Operationally, it is useful to note that

$$r(x) = - \frac{d}{dx} [\log u'(x)] \quad . \tag{4.14}$$

The risk aversion function* preserves all that is essential concerning u while eliminating the arbitrariness. That means, more formally,

Theorem 4.9: Two utility functions which are strategically equivalent have the same risk aversion function and conversely.

---

*Whenever $r(\cdot)$ is discussed, we are assuming that $u(\cdot)$ is twice continously differentiable.

<u>Proof</u>. Let $u_1(x) = a + b\, u_2(x)$, $b > 0$. Clearly, $u_1'(x) = bu_2'(x)$ and $u_1''(x) = bu_2''(x)$, so

$$r_1(x) \equiv -\frac{u_1''(x)}{u_1'(x)} = -\frac{bu_2''(x)}{bu_2'(x)} \equiv r_2(x)$$

To prove the converse, notice from (4.14) that

$$-r(x) = \frac{d}{dx}\,[\log u'(x)] \,.$$

Integrating both sides gives us

$$\int -r(x)\,dx = \log u'(x) + c \,,$$

where c is an integration constant. Exponentiating this, we find

$$e^{-\int r(x)\,dx} = e^{\log u'(x)+c} = e^c\, u'(x) \,.$$

And finally integrating again,

$$\int e^{-\int r(x)\,dx} = \int e^c\, u'(x)\,dx = e^c\, u(x) + d \,.$$

Since $e^c > 0$ and d are constants, r(x) specifies u(x) up to positive linear transformations. ◁

4.5.1 <u>Interpreting the Risk Aversion Function</u>

Let us try to build up an intuitive interpretation for the risk aversion function. Let x_0 denote the decision maker's initial endowment of a given attribute X, and now consider adding to x_0 a lottery $\tilde{x}$ involving only a small range of X with an expected consequence $E(\tilde{x})$ equal to zero. Also, let $\pi(x_0,\tilde{x})$ be the decision maker's risk premium*

---

*A cautionary word about a possible notational confusion is in order. We use the notation $\pi(x,\tilde{x})$ as the risk premium for the lottery $(x+\tilde{x})$. When $\tilde{x}$ is the special lottery $\langle -h,h\rangle$, we use the symbolism $\pi(x,h)$ instead of $\pi(x,\langle -h,h\rangle)$ for the risk premium of the lottery $x + \langle -h,h\rangle$, or equivalently of the lottery $\langle x-h,x+h\rangle$.

for $x_0 + \tilde{x}$. By definition of the certainty equivalent

$$u(x_0 - \pi) = E[u(x_0 + \tilde{x})] . \tag{4.15}$$

Using Taylor's formula to expand both sides of (4.15), we find

$$u(x_0 - \pi) = u(x_0) - \pi u'(x_0) + \frac{\pi^2}{2!} u''(x_0) + \dots \tag{4.16}$$

and

$$\begin{aligned} E[u(x_0 + \tilde{x})] &= E[u(x_0) + \tilde{x}u'(x_0) + \frac{1}{2!}\tilde{x}^2 u''(x_0) \\ &\quad + \frac{1}{3!}\tilde{x}^3 u'''(x_0) + \dots] \\ &= u(x_0) + \frac{1}{2} E[\tilde{x}^2]u''(x_0) + \frac{1}{3!} E[\tilde{x}^3]u'''(x_0) + \dots \end{aligned} \tag{4.17}$$

Equating (4.16) and (4.17) and neglecting the higher order terms gives us

$$-\pi u'(x_0) \approx \frac{1}{2} E[\tilde{x}^2]u''(x_0) . \tag{4.18}$$

Realizing that $E[\tilde{x}^2]$ is the variance σ_x^2 of the lottery $\tilde{x}$, since $E(\tilde{x}) = 0$, and rearranging (4.18), we find

$$\pi(x_0, \tilde{x}) \approx \frac{1}{2} \sigma_x^2 \; r(x_0) , \tag{4.19}$$

where $r(x_0)$ is defined by (4.13). Thus, starting with an initial level x_0, the decision maker's risk premium for a small-ranged lottery with $E(\tilde{x}) = 0$ is $r(x_0)$ times half the variance of $\tilde{x}$ to a first approximation. Stated another way, the risk aversion $r(x_0)$ is twice the risk premium per unit variance for such lotteries.

Let us now work through a couple of examples to gain a better feeling for the risk aversion function.

Example 4.9 To find the risk aversion function for $u(x) = a - be^{-cx}$, $b > 0$, we calculate $u'(x) = cbe^{-cx}$ and $u''(x) = -c^2be^{-cx}$, so from (4.13),

$$r(x) \equiv -\frac{u''(x)}{u'(x)} = -\frac{-c^2be^{-cx}}{cbe^{-cx}} = c \quad .$$

Using the same utility function, in Table 4.1 we displayed the expected consequence $\bar{x}$ and certainty equivalents $\hat{x}$ for three different lotteries of the form $\langle x_1, x_2 \rangle$ for three different values of c. Using this, it is a simple matter to calculate the risk premium π for all these lotteries. This is done in Table 4.4. Notice that for any particular value of c, the risk premium for lotteries of the form $\langle x, x + 10 \rangle$ are the same. Also, notice that as c gets smaller, the risk premiums for the same lottery get smaller, and that all the risk premiums are positive. ◻

Table 4.4 The Risk Aversion Function for $u(x) = a - be^{-cx}$

c	x_1	x_2	$\bar{x}$	$\hat{x}$	π	$r(x)$
1	0	10	5	0.69	4.31	1
1	10	20	15	10.69	4.31	1
1	20	30	25	20.69	4.31	1
0.2	0	10	5	2.85	2.15	0.2
0.2	10	20	15	12.85	2.15	0.2
0.2	20	30	25	22.85	2.15	0.2
0.1	0	10	5	3.8	1.2	0.1
0.1	10	20	15	13.8	1.2	0.1
0.1	20	30	25	23.8	1.2	0.1

Obervations such as these might lead one to wonder what kind of general statements can be implied about the decision maker's preferences from a knowledge of the risk aversion function. One simple result is

<u>Theorem 4.10</u> <u>If r is positive for all x, then u is concave and the decision maker is risk-averse.</u>

<u>Proof</u>. Assume r is positive. Now since u' is always positive (u is increasing), u"(x) must be negative. This implies u is concave which in turn implies the decision maker is risk averse. ◁

And as you might expect, the analog is

<u>Theorem 4.11</u> <u>If r is negative for all x, then u is convex and the decision maker is risk-prone</u>.

Let u_1 and u_2 be utility functions with risk aversion r_1 and r_2 respectively. Then, from (4.19) one can see that if $r_1(x_0) > r_2(x_0)$ at a particular point x_0, the risk premium $\pi_1(x_0,\tilde{x})$ for a small range lottery $\tilde{x}$ with $E(\tilde{x}) = 0$ is larger than the corresponding risk premium $\pi_2(x_0,\tilde{x})$. However, a more important result which holds for any lottery is

<u>Theorem 4.12</u> <u>If $r_1(x) > r_2(x)$ for all x, then $\pi_1(x,\tilde{x}) > \pi_2(x,\tilde{x})$ for all x and $\tilde{x}$.</u>

(In other words if u_1 has a uniformly larger <u>local</u> risk aversion than u_2, then the risk premium for any lottery $x + \tilde{x}$ is larger with u_1 than u_2. This means that a uniform local condition has a natural global implication.)

Proof.* Assume $r_1(x) > r_2(x)$. Therefore,

$$r_2(x) - r_1(x) = -\frac{d}{dx}\left[\log u_2'(x)\right] + \frac{d}{dx}\left[\log u_1'(x)\right]$$

$$= \frac{d}{dx}\left[\log \frac{u_1'(x)}{u_2'(x)}\right] .$$

is negative. It follows that $\log[u_1'(x)/u_2'(x)]$ is decreasing. Note that

$$\frac{d}{dt} u_1(u_2^{-1}(t)) = \frac{u_1'(u_2^{-1}(t))}{u_2'(u_2^{-1}(t))} ,$$

which is also decreasing in t since $\log[u_1'(x)/u_2'(x)]$ is. Therefore $u_1(u_2^{-1}(t))$ is a concave function of t.

Working the other direction, by definition

$$\pi_i(x,\tilde{x}) = x + E[\tilde{x}] - u_i^{-1}(E[u_i(x + \tilde{x})]), i=1,2.$$

Then, simply subtracting, we find

$$\pi_1(x,\tilde{x}) - \pi_2(x,\tilde{x}) = u_2^{-1}(E[u_2(x + \tilde{x})]) - u_1^{-1}(E[u_1(x + \tilde{x})])$$

$$= u_2^{-1}(E[\tilde{t}]) - u_1^{-1}(E[u_1(u_2^{-1}(\tilde{t}))])$$

where $\tilde{t} = u_2(x + \tilde{x})$. Since $u_1(u_2^{-1}(t))$ is concave, from Jensen's Inequality†, we have

$$E[u_1(u_2^{-1}(\tilde{t}))] < u_1(u_2^{-1}[E(\tilde{t})]).$$

---

*This proof, which is given in Pratt [1964], is mathematically more involved than the rest of this section. The details of the proof are not required in later discussions.

†See William Feller, An Introduction to Probability Theory and Its Applications, vol. 2., Wiley, New York, 1966.

Substituting this into the previous expression, we find

$$\begin{aligned}\pi_1(x,\tilde{x}) - \pi_2(x,\tilde{x}) &> u_2^{-1}(E[\tilde{t}]) - u_1^{-1}[u_1(u_2^{-1}[E(\tilde{t})])] \\ &> u_2^{-1}[E(\tilde{t})] - u_2^{-1}[E(\tilde{t})] \\ &> 0\end{aligned}$$

which is the desired result. ◁

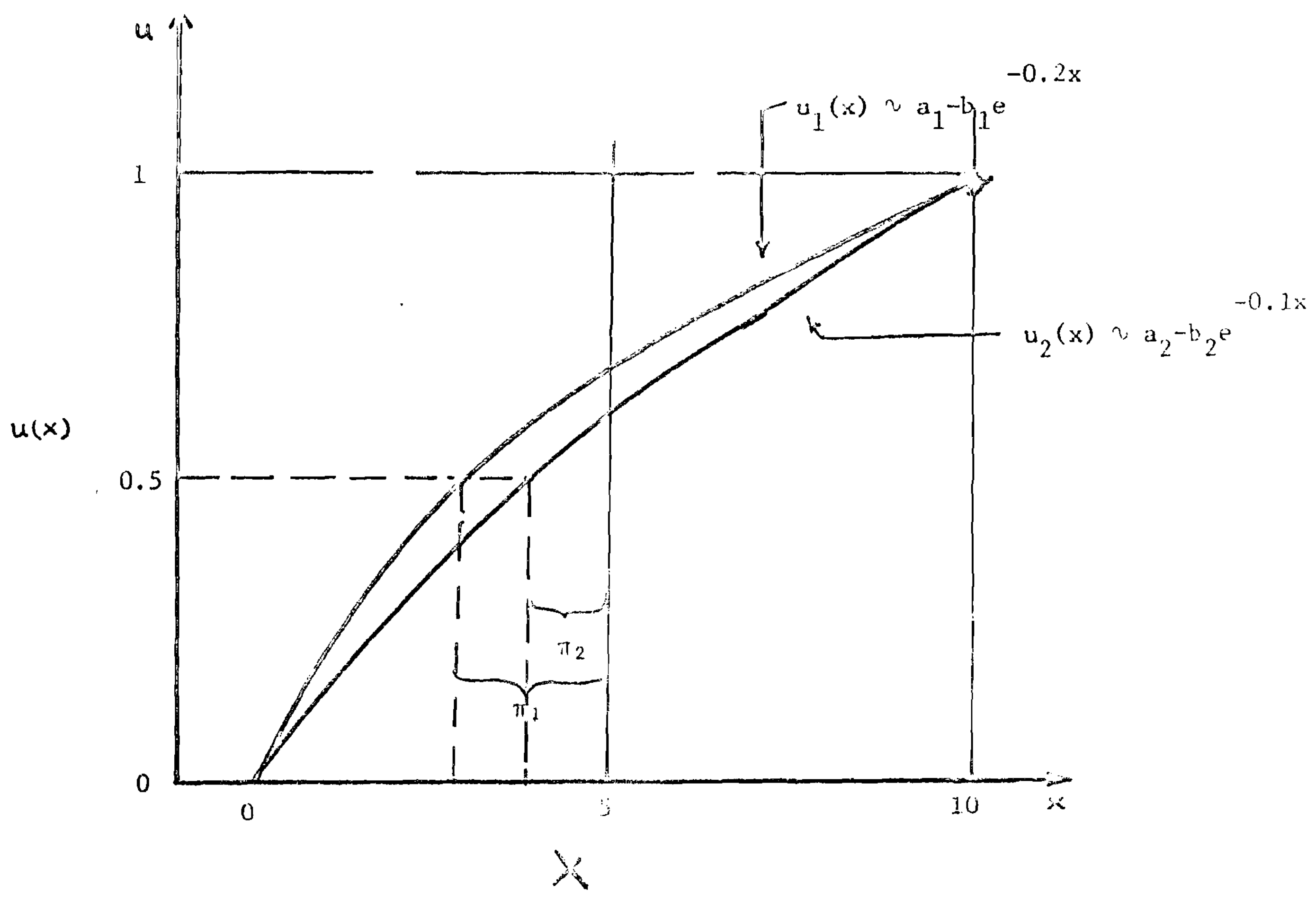
It should be pointed out that the above result required no restrictions on the sign of r_1 or r_2. Thus, the statement is valid for both risk-averse and risk-prone decision makers.

An illustration of the implications of the preceding result seems appropriate. In example 4.9, we showed the risk aversion function for $u(x) = a - be^{-cx}$ was c. Table 4.4 indicated the risk premium for <0,10> was 2.15 when $c = 0.2$ and 1.2 when $c = 0.1$. This is illustrated in Fig. 4.10 where we let $u_1(x) = a_1 - b_1e^{-0.2x}$, $u_2(x) = a_2 - b_2e^{-0.1x}$, and set $u_1(0) = u_2(0) = 0$ and $u_1(10) = u_2(10) = 1$ for reference. Our result states that since $r_1(x) > r_2(x)$ for all x, then π_1 for <0,10> must be greater than π_2 for <0,10>. That this is the case is clear from the figure.

In Figure 4.11, we take things one step further and plot the risk premium and the certainty equivalent for < 0,10 > using $u(x) = -e^{-cx}$ as a function of c (the risk aversion function). As we expected, the risk premium for the lottery increases and the certainty equivalent decreases as the risk aversion increases. For all values of c, the risk premium plus the certainty equivalent must equal the expected consequence, which is always 5.

Figure 4 .10

The Relationship Between The Risk Premium and The Risk Aversion Function



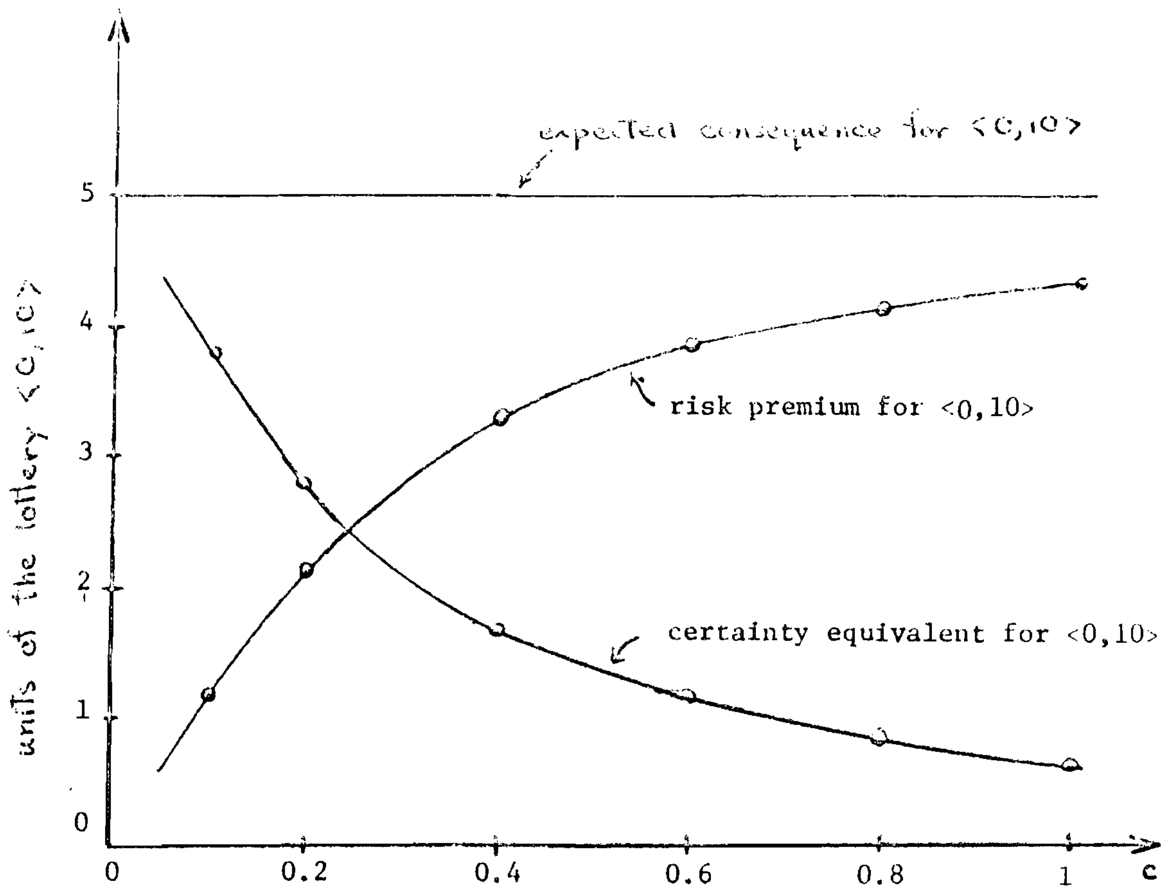


Figure 4.11

Relationship Between The Risk Aversion Function and The Risk Premium with $u(x) = -e^{-cx}$

4.6. CONSTANT, DECREASING, AND INCREASING RISK AVERSION

In previous sections we have spoken of a risk premium $\pi(x,\tilde{x})$ for lottery $\tilde{x}$ given reference point x--i.e., for lottery $(x + \tilde{x})$. A very interesting question is what happens to $\pi(x,\tilde{x})$ as x increases. For greater amounts of x, is the decision maker's risk premium larger or smaller? Often a decision maker may be able to state that as x increases, the risk premium he would be willing to pay for $\tilde{x}$ decreases, for example. As we will show in this section, such attitudes put strong restrictions on the functional form of the utility function. Also, by working directly with the utility function u, it is difficult to determine whether or not such preference attitudes are implied. However, they are very apparent from the risk aversion function r.

For an increasing utility function, let us consider the risk premium $\pi(x,h)$ for the lottery $\langle x + h, x - h\rangle$ for a risk averse individual. Clearly, π is positive for all amounts of x. However, it might be reasonable that one's risk premium for this lottery should decrease as x increases. As an illustration of a situation where such an attitude might be relevant, suppose x represents a specific monetary asset positition of a decision maker and h is some monetary amount. It seems to be empirically true for many people that as their assets increase, they are only willing to pay a smaller risk premium for a given risk. Their reasoning is that as they become richer, they can better afford to take a specific risk, and therefore will forgo less to avoid it. The same reasoning implies that the insurance premium associated with an unfavorable lottery (i.e. one less preferable than the status quo) decreases as one gets richer and increases as one gets poorer.

Let us formalize this attitude which is intuitively appealing for many decision makers.

Definition: An individual is said to be decreasingly risk averse if (1) he is risk averse, and (2) his risk premium $\pi(x,\tilde{x})$ for any lottery $\tilde{x}$ decreases as the reference amount x increases.

However, with our present tools, it would be all but impossible to determine whether or not a specific utility function implied such an attitude. To accomplish such a task would require an exhaustive check for all possible lotteries $\tilde{x}$. Fortunately, Pratt proves an important result which gets us out of this difficulty and makes the concept of decreasing risk aversion operationally significant. That is,

Theorem 4.13. The risk aversion function r for utility function u is decreasing if and only if the risk premium $\pi(x,\tilde{x})$ is a decreasing function of x for all $\tilde{x}$.

Proof. Theorem 4.12 states if $r_1(x) > r_2(x)$, then $\pi_1(x,\tilde{x}) > \pi_2(x,\tilde{x})$ for all $\tilde{x}$. Applying this to $u_1(x) \equiv u(x)$ and $u_2(x) \equiv u(x+k)$ for positive and negative k proves the "if" and "only if" parts of this assertion, respectively. ◁

As we will soon see, many of the "traditional" candidates for a utility function, such as the exponential and quadratic utility functions, are not appropriate for a decreasingly risk-averse decision maker. Thus, the characteristic of decreasing risk aversion places strong restrictions on the shape (i.e., functional form) of one's utility function. If we know that the decision maker desires his utility function to be decreasingly risk averse, then this constraint significantly simplifies the assessment of his utility function. Some examples seem appropriate.

<u>Example 4.10</u> Consider the exponential utility function $u(x) = -e^{-cx}$, $c > 0$. In example 4.2 we illustrated and later proved that the risk premium $\pi(x,\tilde{x})$ associated with any lottery $\tilde{x}$ does not depend on x when $u(x) = -e^{-cx}$. Thus, although this utility function implies risk aversion, it clearly does not imply decreasing risk aversion since $\pi(x,\tilde{x})$ is constant, not decreasing, for any $\tilde{x}$. □

Let us consider such an attitude in a little more detail. A fact related to the previous assertion, which we state without proof, is

<u>Theorem 4.14.</u> <u>The risk aversion r is constant if and only if $\pi(x,\tilde{x})$ is a constant function of x for</u> all $\tilde{x}$.

<u>Definition:</u> A decision maker is <u>constantly risk averse</u>

if r is a positive constant, constantly risk neutral if r is zero, and constantly risk prone if r is a negative constant. To indicate the strong restrictions these conditions place on the shape of a utility function, we can show

Theorem 4.15 $u(x) \sim -e^{-cx} \Leftrightarrow r(x) \equiv c > 0$, (constant risk aversion) (4.20)

$u(x) \sim x \quad \Leftrightarrow r(x) \equiv 0$, (risk neutrality), (4.21)

$u(x) \sim e^{-cx} \Leftrightarrow r(x) \equiv c < 0$, (constant risk proneness). (4.22)

Proof. If $u(x) \sim -e^{-cx}$, using definition (4.13), $r(x) = c$. Now, if $r(x) = c > 0$, from (4.14)

$$\frac{d}{dx}[\log u'(x)] = -c \ .$$

Integrating and exponentiating both sides gives

$$e^{-cx} = e^{-c\int dx} = e^{\log u'(x) + d} = e^{d}u'(x) ,$$

where d is a constant of integration. Integrating again yields

$$\int e^{-cx}dx = -\frac{e^{-cx}}{c} = e^{d}\, u(x) + h \ ,$$

where h is another constant of integration. From this, clearly, $u(x) \sim -e^{-cx}$.

The other proofs are similar. $\triangleleft$

This result says, for instance, that if the decision maker is constantly risk averse, his utility function must be of the form (4.20). Knowing this, one needs only to determine the value of parameter c in order to completely specify his utility function. This can easily be done by determining the certainty equivalent of any one lottery. However, the sophisticated analyst would employ consistency checks in his assessments, so the procedure may not be as simple as it appears. The problem of assessing utility functions is considered in Section 4.9.

Since we're still interested in finding a decreasingly risk averse family of utility functions, let us examine the following.

Example 4.11. Consider the quadratic utility function

$$u(x) = a + bx - cx^2 , \qquad (4.23)$$

where $b > 0$, $c > 0$, and x is constrained to amounts less than $b/2c$, since the utility function is decreasing beyond this point. Taking derivatives, we find $u'(x) = b - 2cx$ and $u''(x) = -2c$, so the risk aversion function

$$r(x) \equiv -\frac{u''(x)}{u'(x)} = \frac{2c}{b - 2cx} . \qquad (4.24)$$

Since $r > 0$ for all x, clearly u is risk averse, but r increases as x increases, so u is certainly not decreasingly risk averse. Hence we see that the quadratic utility function is not appropriate to use when decreasing risk aversion is a compelling desideratum. $\sqcup$

Aside from possessing the property of risk aversion, the quadratic utility function is often used in the literature because the expected utility of a lottery yielding uncertain consequence $\tilde{x}$ depends only on the mean and variance of $\tilde{x}$. That is,

$$\begin{aligned} E[u(\tilde{x})] &= E[a + b\tilde{x} - c\tilde{x}^2] \\ &= a + b\bar{x} - c(\sigma^2 + \bar{x}^2) \\ &= u(\bar{x}) - c\sigma^2 , \end{aligned}$$

where $E(\tilde{x}) = \bar{x}$ and Var $(\tilde{x}) = \sigma^2$. As indicated in the first section of this chapter, in general, we do not think it is reasonable to base one's decisions solely on the mean and variance of the expected consequence. This example gives us the motivation for a

Definition: A decision maker is increasingly risk averse if (1) he is risk averse, and (2) his risk premium $\pi(x,\tilde{x})$ increases in x for any specific lottery $\tilde{x}$.

To relate such an attitude to the risk aversion function, we have

Theorem 4.16. The risk aversion function r is increasing if and only if $\pi(x,\tilde{x})$ is increasing in x for any $\tilde{x}$.

The proof, which is similar to preceding ones, is omitted.

Recall from (4.24) that r for the quadratic utility function is increasing in x. Since that utility function is also risk averse, it follows that it is increasingly risk averse. Since this attitude implies, for instance, that a person should be willing to pay higher insurance premiums to avoid certain unfavorable lotteries when he becomes richer, we would normally not expect many decision makers to subscribe to it. However, provided this condition held, it could, and should, be exploited to simplify the assessment of the utility function.

Finally, we get to a decreasingly risk averse utility function.

Example 4.12 Consider the logarithmic utility function $u(x) = \log (x + b)$, discussed in Example 4.3. Taking derivatives, we find $u'(x) = \frac{1}{x + b}$ and $u''(x) = -\frac{1}{(x + b)^2}$, so

$$r(x) = -\frac{u''(x)}{u'(x)} = \frac{1}{x + b} .$$

Clearly, $r(x)$ is positive and decreasing in x for all $x > -b$. Thus $u(x)$ is a decreasingly risk averse utility function over this range of x. □

Let us digress and see where we stand. We have looked at increasing risk aversion, constant risk aversion, and decreasing risk aversion. Intuitive arguments and experience tell us the increasing case is of little interest, and we have essentially covered what is important concerning the constant case. However, more must be said about the decreasing case. From the example, the reader may gather that few utility functions of simple form are, in fact, decreasingly risk averse. This is unfortunate as we would like a few simple families of utility functions with a rich variety of specific members. Then provided an individual wished to be decreasingly risk averse, we could hypothesize a particular family of utility functions and concentrate on evaluating the specific member appropriate to the situation in question. This defines the problem.

A useful result which allows us to construct such classes of utility functions is as follows.

<u>Theorem 4.17.</u> <u>A utility function, which is the weighted sum of two or more utility functions which are decreasingly or constantly risk averse on the interval $[x^o, x^*]$, is itself decreasingly or constantly risk averse on $[x^o, x^*]$, and except on subintervals where the weighted utility functions have equal and constant risk aversion, it is decreasingly risk averse.</u>

<u>Proof</u>. Let $u = u_1 + ku_2$, $k > 0$. Then

$$r = -\frac{u''}{u'} = -\frac{u_1'' + ku_2''}{u_1' + ku_2'}$$

$$= \frac{u_1'}{u_1' + ku_2'} r_1 + \frac{ku_2'}{u_1' + ku_2'} r_2$$

which is differentiated to yield

$$r' = \frac{u_1'}{u_1' + ku_2'} r_1' + r_1 \frac{(u_1' + ku_2')u_1'' - u_1'(u_1'' + ku_2'')}{(u_1' + ku_2')^2}$$

$$+ \frac{ku_2'}{u_1' + ku_2'} r_2' + r_2 \frac{(u_1' + ku_2')ku_2'' - ku_2'(u_1'' + ku_2'')}{(u_1' + ku_2')^2}$$

$$= \frac{u_1'r_1' + ku_2'r_2'}{u_1' + ku_2'} + \frac{r_1[k(u_2'u_1'' - u_1'u_2'')] + r_2[k(u_1'u_2'' - u_2'u_1'')]}{(u_1' + ku_2')^2}$$

$$= \frac{u_1'r_1' + ku_2'r_2'}{u_1' + ku_2'} - \frac{k(r_1 - r_2)^2 u_1'u_2'}{(u_1' + ku_2')^2} .$$

Since $u_1' > 0$, $u_2' > 0$, $r_1' \leq 0$, and $r_2' \leq 0$, we see that $r' < 0$ and therefore the assertion is true for the case $u = u_1 + ku_2$. The general case,

$u = \sum_{i=1}^{n} c_i u_i$, $c_i > 0$, follows from repeated application of the proof. ◁

Let us illustrate the applicability of this result with

Example 4.13 What is the risk aversion for $u(x) = -e^{-ax} - be^{-cx}$, where a , b, and c are positive constants? If we define $u_1(x) = -e^{-ax}$ and $u_2(x) = -e^{-cx}$, then $u(x) = u_1(x) + bu_2(x)$. Also, we know $r_1(x) = a$ and $r_2(x) = c$. Thus, from Theorem 4.17, it follows that u(x) must be constantly risk averse if a = c and decreasingly risk averse if $a \neq c$. This can be validated directly. Suppose a = c, then $u(x) = -e^{-ax} - be^{-ax} = -(1 + b)e^{-ax}$, which we know is constantly risk averse. If $a \neq c$, $u'(x) = ae^{-ax} + bce^{-cx}$ and $u''(x) = -a^2e^{-ax} - bc^2e^{-cx}$, so

$$r(x) = \frac{a^2e^{-ax} + bc^2e^{-cx}}{ae^{-ax} + bce^{-cx}} , \qquad (4.25)$$

whose derivative is negative. Thus u(x) is indeed decreasingly risk averse. □

The utility function of the preceding example is frequently used in actual assessment of preferences. Let us consider it in more detail to further our intuitive understanding of decreasing risk aversion. To develop a feeling for the risk aversion r(x) in (4.25) as a function of x, we must first look at the behavior of e^{-ax} and e^{-cx}. Without loss of generality, let us assume a > c. Both e^{-ax} and e^{-cx} are graphed in Fig. 4.12 with $a = 1$ and $c = 0.25$. Both terms have large positive values for large negative amounts of x, decrease but remain positive for all x, are less than one for positive x, and asymptotically approach zero as x grows large. Their ratio, that is $e^{-ax}/e^{-cx} = e^{-(a-c)x}$ is perhaps more revealing. It, too, is plotted in Fig. 4.12 and is clearly of the same shape as the original two functions. Thus e^{-cx} is very small compared to e^{-ax} for large negative values of x, they are equal at $x = 0$, and e^{-ax} is small compared to e^{-cx} for large positive values of x.

With this background, let's look at the risk aversion for $u(x) = -e^{-ax} - be^{-cx}$, $a > c$, in more detail. From (4.25), $r(0) = (a^2 + bc^2)/(a + bc)$ which is less than $\underline{a}$ but greater than c. For large negative amounts of x, since e^{-cx} is small compared to e^{-ax}, we find

$$r(x) = \frac{a^2e^{-ax} + bc^2e^{-cx}}{ae^{-ax} + bce^{-cx}} \approx \frac{a^2e^{-ax}}{ae^{-ax}} = a \quad .$$

The limit of r(x) as x goes to minus infinity is $\underline{a}$. For large positive amounts of x, we know e^{-ax} is small compared e^{-cx}, so

$$r(x) \approx \frac{bc^2e^{-cx}}{bce^{-cx}} = c \quad .$$

The limit of r(x) as x approaches plus infinity is c.

A graph of r(x) as a function of x for $a = 1.0$, $c = 0.25$ and two different values of b is shown in Fig. 4.13. The general shape of each curve is as we just described. The

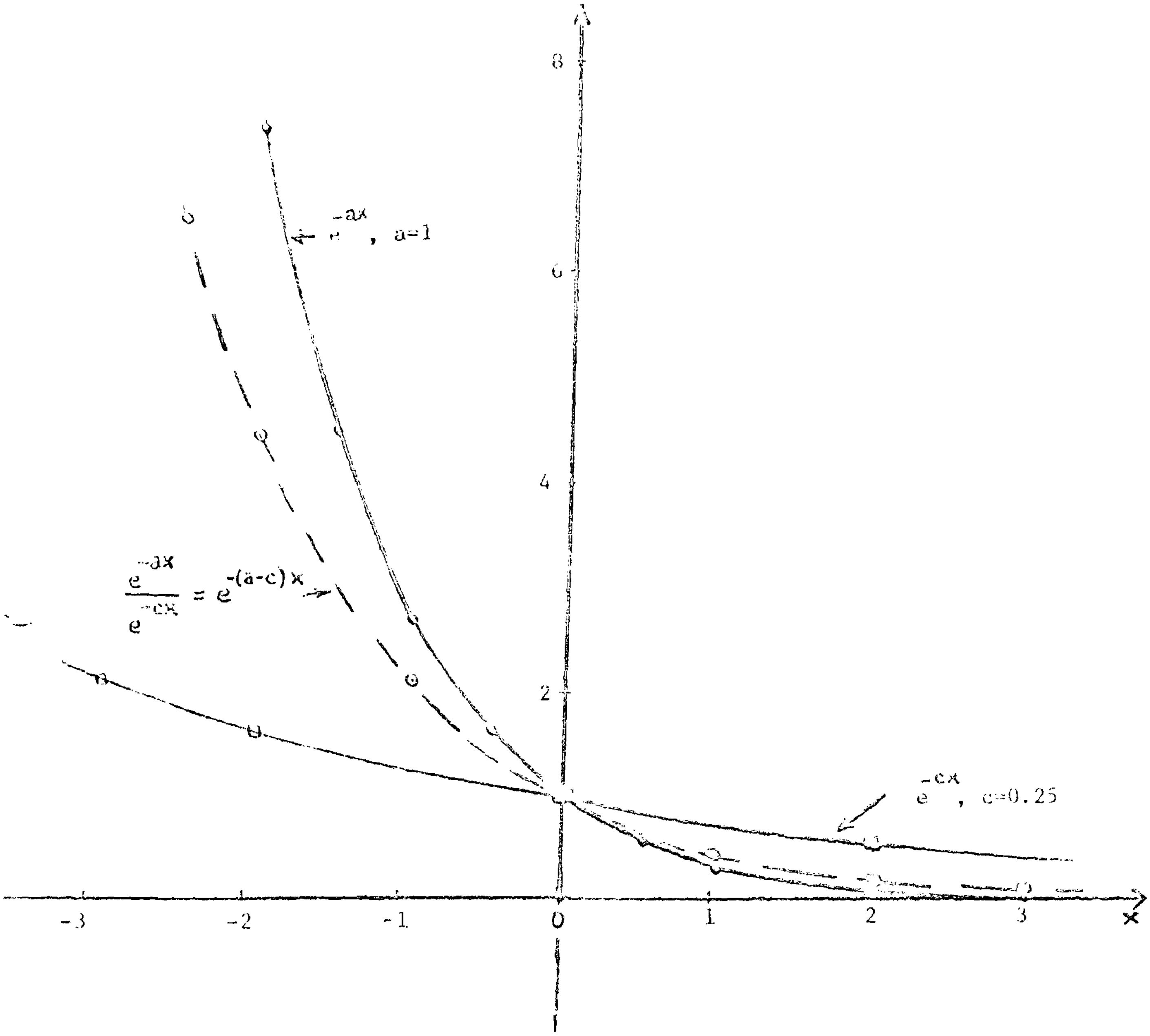


Figure 4.12

The Behavior of the Component Exponentials of the Utility Function $u(x) = -e^{-ax} - be^{-cx}$

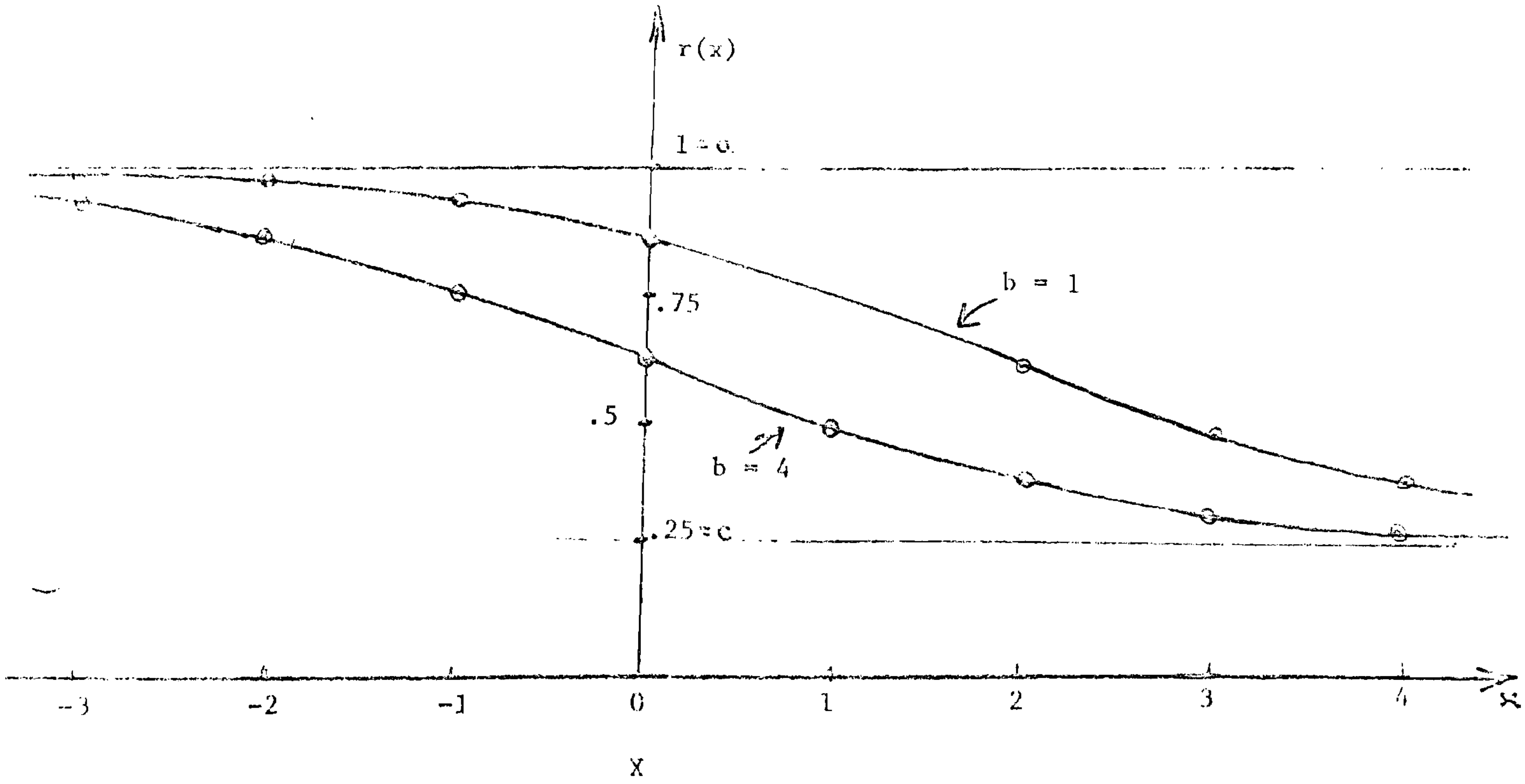


Figure 4.13. The risk aversion $r(x)$ for the utility function $u(x) = -e^{-x} - be^{-0.25x}$

risk aversion $r(x)$ is decreasing in x and always between $r_1(x) = a$ and $r_2(x) = c$, where $u(x) = u_1(x) + bu_2(x)$. The weighting factor b determines the amounts of x for which $r(x)$ is essentially $r_2(x)$. That is, for larger values of b, the term $r(x)$ is closely approximated by $r_2(x)$ for smaller values of x. Notice that $r(x)$ given $b = 1$ is larger than $r(x)$ given $b = 4$ for all amounts of x. From Theorem 4.12, we know this implies the risk premium for any lottery $\tilde{x}$ found using $u(x) = -e^{-ax} - b_1e^{-cx}$ will be larger than the risk premium for $\tilde{x}$ found with $u(x) = -e^{-ax} - b_2e^{-cx}$ if and only if $b_2 > b_1$.

Another example of a decreasingly risk averse utility function seems appropriate.

<u>Example 4.14</u> What is the risk aversion for $u(x) = -e^{-ax} + bx$, where a and b are positive. If we let $u_1(x) = -e^{-ax}$ and $u_2(x) = x$, then $r_1(x) = a$ and $r_2(x) = 0$, so from Theorem 4.17, $u(x)$ must be decreasingly risk averse. To prove this directly, we have $u'(x) = ae^{-ax} + b$ and $u''(x) = -a^2e^{-ax}$, giving

$$r(x) = \frac{a^2e^{-ax}}{ae^{-ax} + b} \quad .$$

Experience from the preceding example tells us $r(x)$ is approximately equal to a for large negative amounts of x, decreases to $a^2/(a + b)$ at $x = 0$, and asymptotically approaches zero as x grows larger. ☐

Because of the fundamental importance of decreasingly risk averse utility functions in utility theory, we catalogue some of the more common ones in Table 4.5. This list, of course, is not exhaustive.

Table 4.5

Some Common Decreasingly Risk Averse Utility Functions

$u(x)$	restrictions	$r(x)$	decreasing risk averse range
$\log(x+b)$	-	$\frac{1}{x+b}$	$x \geq -b$
$(x+b)^c$	$0 < c < 1$	$-\frac{(c-1)}{x+b}$	$x \geq -b$
$(x+b)^{-c}$	$c > 0$	$\frac{c+1}{x+b}$	$x > -b$
$x + c\log(x+b)$	$c > 0$	$\frac{c}{(x+b)(x+c+b)}$	$x > -b$
$-e^{-ax} - be^{-cx}$	$a,b,c, > 0$	$\frac{a^2e^{-ax} + bc^2e^{-cx}}{ae^{-ax} + bce^{-cx}}$	all x
$-e^{-ax} + bx$	$a,b > 0$	$\frac{a^2e^{-ax}}{ae^{-ax} + b}$	all x

4.6.1 Decreasing Risk Proneness*

It is probably evident by now that one could categorize risk prone utility functions as either decreasingly risk prone, constantly risk prone, or increasingly risk prone. We have already mentioned the second of these, but let us discuss the first to make sure the concepts are clear.

*This subsection is included only for completeness' sake and for reference purposes, and it can be omitted without impairing continuity of the development.

Definition: An individual is decreasingly risk prone if (1) he is risk prone, and (2) his risk premium $\pi(x,\tilde{x})$ for any lottery $\tilde{x}$ increases as the reference amount x increases. Recall that $\pi(x,\tilde{x})$ for risk prone utility functions is always negative.

To provide an operational method for implementing the concept of a decreasingly risk prone utility function, we have

Theorem 4.18. The utility function u is decreasingly risk prone if and only if the associated risk aversion r is negative and increasing.

The proof is omitted because of its similarity to previous ones. Let us illustrate the result with a simple example.

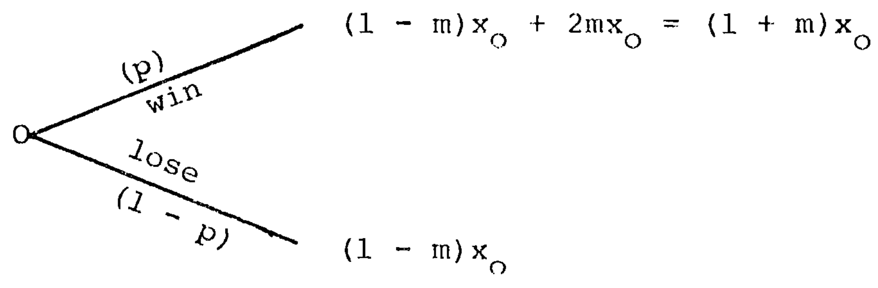
Example 4.15 Consider the utility function $u(x) = x^2$. Since $u'(x) = 2x$ and $u''(x) = 2$, the associated risk aversion $r(x) = -1/x$. This is clearly negative and increasing for positive x so u(x) is decreasingly risk prone for that range of x. The expected utility of <1,3> is 5, from which the certainty equivalent for <1,3> is found to be 2.24. The associated risk premium is -0.24. Likewise, the risk premiums for <2,4> and <3,5> are -0.17 and -0.12, respectively. As expected, they are increasing. ☐

4.7 PROPORTIONAL RISK AVERSION

In this section, another concept concerning risk is examined--namely, proportional risk aversion. And as we have often done earlier, the ideas will be introduced in the context of preferences for monetary consequences. Again however, the theory is relevant in other contexts.

Consider the situation of an investor who has an amount x_0 he can invest in any one of a set of investment plans $\{I_\alpha\}$. If he chooses investment I_α his resulting asset position (his gross payoff) will be $x_0 \tilde{z}_\alpha$ where z_α is a nonnegative random variable. Thus if the investor has utility function u defined on assets rather than on incremental monetary values, so that 0 now refers to "ruin" rather than the status quo--then he will choose that investment I_α to maximize $E[u(x_0 \tilde{z}_\alpha)]$. Throughout this section, we assume increasing preferences for assets.

As an illustration, consider the class of investments where the investor puts up a proportion m, say, of his assets on a double or nothing bet where the probability of winning is p and of losing is 1 - p. The outcome of his investment can then be depicted as follows:



Hence this investment leads to a payoff of $x_0 \tilde{z}_m$

where

$$\tilde{z}_m = \begin{cases} 1 + m \text{ with probability } p \\ 1 - m \text{ with probability } 1 - p. \end{cases}$$

4.7.1. Investments Independent of Asset Position

We will now consider four special classes of utility functions for which the optimal investment plan does not depend on the initial asset position x_o. These are shown to be the only utility functions that enjoy this property. As motivation, consider two examples.

Example 4.16. Let $u(x) \sim x$, the linear utility function. The decision maker should choose I_α to maximize his expected utility. In this case

$$\begin{aligned}\max_\alpha \; Eu(x_o \cdot \tilde{z}_\alpha) &= \max_\alpha \; E(x_o \cdot \tilde{z}_\alpha) \\ &= x_o \max_\alpha \; E(\tilde{z}_\alpha) \;,\end{aligned}$$

so that the optimal investment does not depend on the amount x_o to be invested. For later purposes we observe that for $u(x) \sim x$, we have

$$x r(x) \equiv -x \frac{u''(x)}{u'(x)} = 0 \;, \quad \text{for all } x. \quad \square$$

Example 4.17. Suppose $u(x) = x^{1-c}$ for $o \neq c < 1$. Then the expected utility of the optimal investment is

$$\begin{aligned}\max_\alpha \; E\, u(x_o \cdot \tilde{z}_\alpha) &= \max_\alpha \; E(x_o \tilde{z}_\alpha)^{1-c} \\ &= x_o^{1-c} \max_\alpha \; E(\tilde{z}_\alpha^{1-c})\end{aligned}$$

so that again the optimal investment does not depend on the amount x_o to be invested. For this case observe that

$$x r(x) \equiv -x \frac{u''(x)}{u'(x)} = c \;.$$

Note that when $c < o$, since x is nonnegative, then r is negative so u is risk prone. When $c > o$, u is risk averse. $\square$

With this as background we shall now prove the following Result and Corollary.

Theorem 4.19.: If in any class of investments the optimal investment plan does not depend on the amount to be invested and if a risk averse u is "well-behaved", then

$$-x\frac{u''(x)}{u'(x)} \text{ is constant.*}$$

Proof. Suppose p is some fixed number where $\frac{1}{2} < p < 1$. Consider the class of investments described earlier where

$$\tilde{z}_m = \begin{cases} 1 + m & \text{with probability} \quad p \\ 1 - m & \qquad\qquad\qquad 1 - p \end{cases}$$

and $0 \leq m \leq 1$. Now

$$E\, u(x_0 \tilde{z}_m) = p\, u[x_0(1 + m)] + (1 - p)\, u[x_0(1 - m)] .$$

To find the maximum m to invest (assuming it is an internal maximum) we differentiate with respect to m and set the result equal to zero, getting

$$pu'[x_0(1 + m)] = (1 - p)\, u'[x_0(1 - m)],\ \text{all } x_0 .$$

Now by the hypothesis, the value of m that satisfies the above is constant for all x_0. Letting

---

*By "well-behaved" we mean u is twice differentiable, and

$$\lim_{x \to 0} -x\,\frac{u''(x)}{u'(x)}$$

exists.

$K = (1 - p)/p$, $x = x_0(1 - m)$, and $\lambda = (1 + m)/(1 - m)$ we get

$$u'(\lambda x) = K\, u'(x) \quad , \text{ for all } x \ .$$

But then

$$\lambda u''(\lambda x) = K\, u''(x) \quad ,$$

and dividing the above two equations, we get

$$-\lambda x \frac{u''(\lambda x)}{u'(\lambda x)} = -x \frac{u''(x)}{u'(x)}$$

or

$$\lambda x\, r(\lambda x) = x\, r(x) \quad \text{for all } x \ .$$

Now using the existence of lim $x\, r(x)$ as $x \to 0$ we must prove $x\, r(x)$ is constant. Suppose, to the contrary, that

$$x_1\, r(x_1) \neq x_2\, r(x_2) \ .$$

Then we have

$$\frac{x_1}{\lambda^n}\, r\left(\frac{x_1}{\lambda^n}\right) = x_1\, r(x_1) \neq x_2\, r(x_2) = \frac{x_2}{\lambda^n}\, r\left(\frac{x_2}{\lambda^n}\right) \ .$$

Taking the limit as $n \to \infty$ in the above (observing that $\lambda > 1$), we contradict the existence of lim $x\, r(x)$ as $x \to 0$.

If the optimal m is not an internal maximum, then the optimal $m = 0$ or 1. But both these cases can be ruled out: the case $m = 0$ by observing that u behaves like a linear function in the small and $E(\tilde{z}_m) > 1$; the case $m = 1$ by observing that since u is risk averse there is an asset amount $\bar{x}_0$ which is preferred to a gamble giving a p chance at x_0 and a complementary chance at 0. (For this last argument think of p as some value such as .51.) This completes the proof. ◁

<u>Corollary:</u> The following are equivalent:

(i) $xr(x)$ is constant

(ii) $u(x) \sim \log x$, or x^{1-c} for $0 \neq c < 1$, or $-x^{-(c-1)}$ for $c > 1$, or $u(x) = x$.

(iii) the optimum investment plan is independent of assets.

By examples 4.16 and 4.17 (plus analogous examples for $u \sim x$ and $u \sim -x^{-(c-1)}$, $c > 1$), it is shown that

(ii) → (iii) and (ii) → (i). The above theorem demonstrates (iii) → (i). It remains to show (i) → (ii).

<u>Proof:</u> From

$$x\, r(x) = -x\frac{d}{dx}\left[\log u'(x)\right] = c \ ,$$

we have

$$\frac{d}{dx} \log u'(x) = -\frac{c}{x} \ ,$$

or

$$\log u'(x) = -c \log x + \text{constant, for } c \neq 0,$$

$$\sim \log x^{-c}, \text{ for } c \neq 0.$$

It follows that

$$u'(x) \sim x^{-c} \ , \text{ for } c \neq 0.$$

For $c = 0$, it is easy to show that $u'(x) \sim k$, where $k > 0$.

Hence, we have

$$u(x) \sim \begin{cases} x & \text{for } c = 0 \ , \\ x^{1-c} & \text{for } c < 1 \ , \ c \neq 0 \ , \\ \log x & \text{for } c = 1 \ , \\ -x^{-(c-1)} & \text{for } c > 1 \ . \end{cases} \qquad (4.26)$$

This completes the proof of the corollary. ◁

Definition: The expression

$$xr(x) \equiv -x\frac{u''(x)}{u'(x)} ,$$

is called the proportional local risk aversion at x.

To interpret this, consider the following two options:

1. Certainty Option: Receive asset position $x(1 + \pi^*_{x,m})$ for certain.
2. Risky Option: Receive with equal probabilities asset positions $x(1 + m)$ or $x(1 - m)$.

If the decision maker is indifferent between these options, the expression $\pi^*_{x,m}$ can be thought of as the proportional risk premium. Now using (4.18), and noting the risk premium $\pi = x\pi^*_{x,m}$ we get

$$\lim_{xm \to 0} \frac{x\pi^*_{x,m}}{x^2 m^2} = -\frac{1}{2}\frac{u''(x)}{u'(x)} = \frac{1}{2} r(x)$$

or

$$\lim_{m \to 0} \frac{\pi^*_{x,m}}{m^2} = \frac{1}{2} xr(x) ,$$

and hence we get the term proportional local risk aversion for $x\,r(x)$.

4.7.2 Specifying the Parameter in Utility Functions Exhibiting Constant Proportional Risk Aversion

We now turn to the question: Given that the decision maker ascertains that he wants to use a utility function with constant proportional risk aversion, how can he operationally determine the appropriate parameter c?

Let the decision maker's current endowment of the given attribute be x_o. We ask him to compare the two options:

1 the status quo, i.e., x_o for certain

2 a 50-50 lottery which will either double his endowment to $2x_o$ or reduce it to ρx_o.

If he is indifferent between options 1 and 2 when $\rho = 1/2$, then $c = 1$ or $u(x) \sim \log x$. If we keep $\rho = 1/2$ and he prefers option 1, then $c > 1$; if he prefers option 2, then $c < 1$. Suppose the decision maker is indifferent between the two options for $\rho > 1/2$, the case where $c > 1$. Then c can be evaluated using $u(x) = -x^{-(c-1)}$ from (4.26) by solving the equation

$$-x_o^{-(c-1)} = \frac{1}{2}\left[-(\rho x_o)^{-(c-1)} - (2x_o)^{-(c-1)}\right]$$

or

$$2 = \rho^{-(c-1)} + 2^{-(c-1)} .$$

For $\rho < 1/2$, the case where $c < 1$, we must solve the equation

$$x_o^{1-c} = \frac{1}{2}\left[(\rho x_o)^{1-c} + (2x_o)^{1-c}\right]$$

or

$$2 = \rho^{1-c} + 2^{1-c} .$$

A plot of c against ρ is illustrated in Figure 4.14. Thus if ρ were .8 for example, c could be read as equal to 4 and then $u(x) \sim -x^{-3}$.

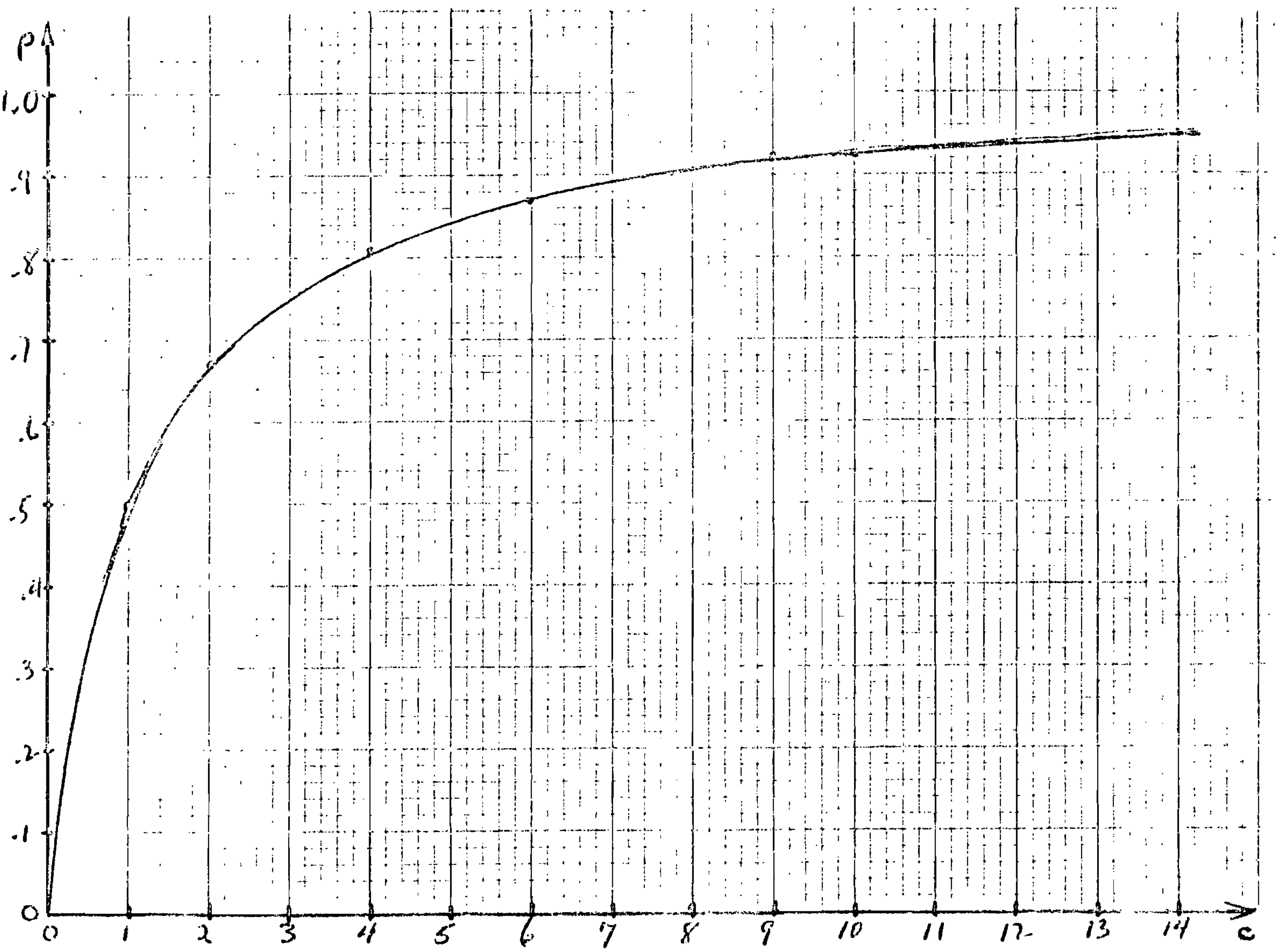


Figure 4.14. Determining the Parameter c in Utility Functions Exhibiting Constant Proportional Risk Aversion

4.8 MONOTONICALLY DECREASING AND NONMONOTONIC UTILITY FUNCTIONS*

In this section, we will extend the concepts of risk introduced in the last four sections to monotonically decreasing and nonmonotonic utility functions. The former case will be considered first, and the order of presentation will be the same as for monotonically increasing utility functions. The concepts of risk averse and risk prone are defined, then a measure of risk aversion introduced, and increasing, decreasing, and constant risk aversion discussed. The last subsection concerns the nonmonotonic case. Proofs of results which are analagous to those presented in earlier sections will be omitted here.

4.8.1 Risk Aversion

For monotonically decreasing preferences, one will be referred to as risk averse if he prefers the expected consequence of any nondegenerate lottery to that lottery. Then of course, if the utility function u represents such preferences, the utility of the expected consequence must be greater than the expected utility of the lottery. If one prefers [is indifferent to] every nondegenerate lottery to its expected consequence, then he is said to be risk prone [risk neutral] . As with the increasing case, one

---

*Once again this section is included primarily for reference purposes. It can be skipped without impairing the continuity of the development.

need not try to verify the property of risk aversion, for example, by checking every possible degenerate lottery. A necessary and sufficient condition for this to hold is that it holds for all 50-50 lotteries. It is not difficult to prove the following

<u>Theorem 4.20.</u> A decision maker is risk averse [risk prone; risk neutral] if and only if his monotonically decreasing utility function is concave [convex; linear].

Figure 4.15 illustrates these cases.

Before preceding further, let us suggest a couple of problems which involve monotonically decreasing preferences. First, consider the response times to calls for ambulance service. Because of the manner in which response time relates to the patient's condition, it may be reasonable to assume that for any response time t, the certainty of t would be preferred to the 50-50 chance at t - 1 or t + 1. Hence, $u(t) > [u(t-1) + (t+1)]/2$, from which it follows that the decision maker's utility function is concave.

A second illustration concerns response times to calls for police service. In this situation, the decision maker may prefer a sure response time t to a lottery $<t + 1, t - 1>$ for any t. The reasoning might be that the probability of apprehending a criminal decreases very fast as the response time increases. This means that $u(t) < [u(t-1) + u(t+1)]/2$, which implies u is convex and risk prone. The decision maker is willing to gamble in this situation in order to have a reasonable chance of obtaining a small response time.

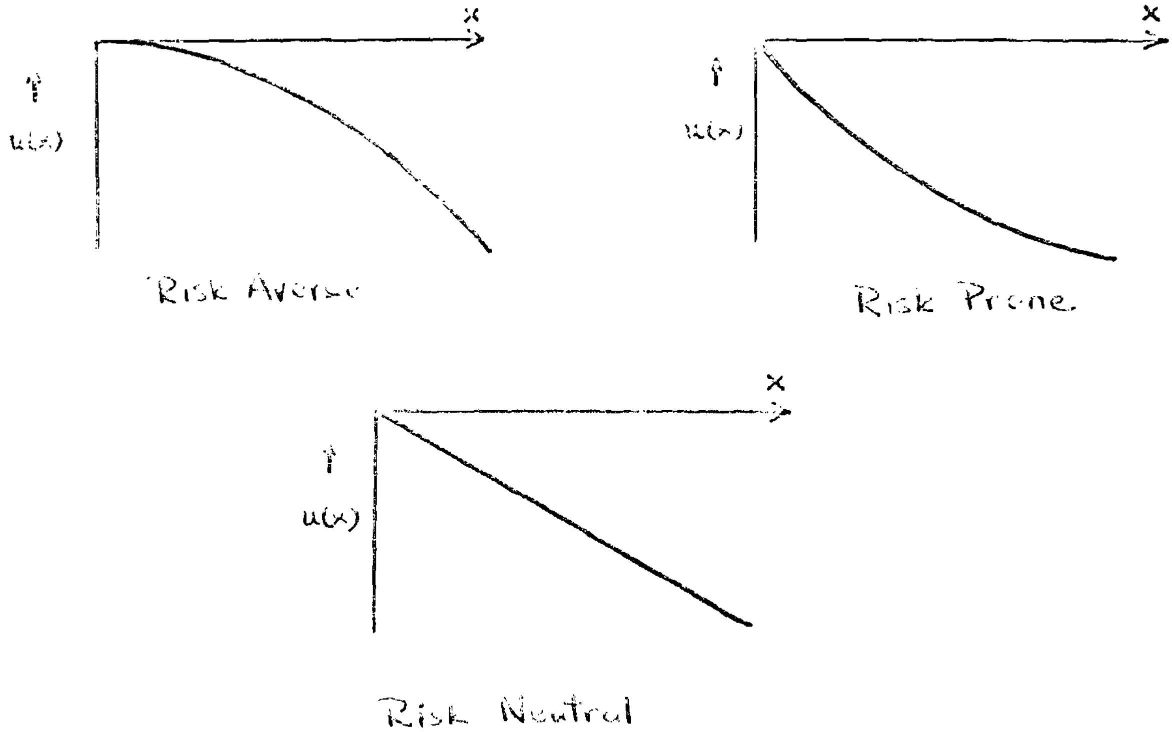


Figure 4.15. Risk Properties of Monotonically Decreasing Utility Functions

So far, the definitions and results stated in this section have been identical to those given for the monotonically increasing case. Now a few differences will come out. Recall that for increasing utility functions, the certainty equivalent had to be less than the expected consequence of a lottery for a risk averse individual. Just the reverse is true for risk averse decreasing utility functions. Furthermore, in the context of increasing utility functions, the risk premium, defined as the expected consequence minus the certainty equivalent represented the amount the decision maker would be willing to give up (from the expected consequence) in order to avoid the risks associated with a particular lottery. To keep this connotation for decreasing utility functions, we are forced to change the definition of the risk premium for the decreasing case. In this context, we define the *risk premium* of a lottery as the certainty equivalent minus the expected consequence of that lottery. Then, it follows that the risk premium is again the amount the decision maker is willing to give up (from the expected consequence) in order to free himself of the responsibilities of a particular lottery. Now, we can state

Theorem 4.21. *For decreasing utility functions, a decision maker is risk averse if and only if his risk premium is positive for all nondegenerate lotteries.*

An example may be helpful.

Example 4.18 Consider the risk averse, decreasing utility function of the form $u(x) = -e^{0.1x}$ illustrated in Fig. 4.16. Let us find the expected consequence, certainty equivalent, and risk premium for a lottery yielding either $x = 2$, $x = 3$, or $x = 7$, each with a probability of 1/3. The expected consequence is

$$\bar{x} = \frac{1}{3}(2 + 3 + 7) = 4 ,$$

and the expected utility is

$$E[u(\tilde{x})] = \frac{1}{3}(-e^{0.1(2)} - e^{0.1(3)} - e^{0.1(7)}) = -1.528.$$

Therefore, the certainty equivalent $\hat{x}$ is such that

$$-e^{0.1\hat{x}} = -1.528 .$$

Solving this, we find $\hat{x} = 4.24$. The risk premium, $\hat{x} - \bar{x}$, is then 0.24. ■

Now let us consider risk proneness.

Theorem 4.22. For decreasing utility functions, the following are equivalent:

1. a decision maker is risk prone,
2. the certainty equivalent is less than the expected consequence of any nondegenerate lottery,
3. the risk premium for all nondegenerate lotteries is negative.

To help illustrate this result, consider

Example 4.19 Suppose $u(x) = e^{-0.2x}$, and we are interested in the certainty equivalent and risk premium for <0,10> .

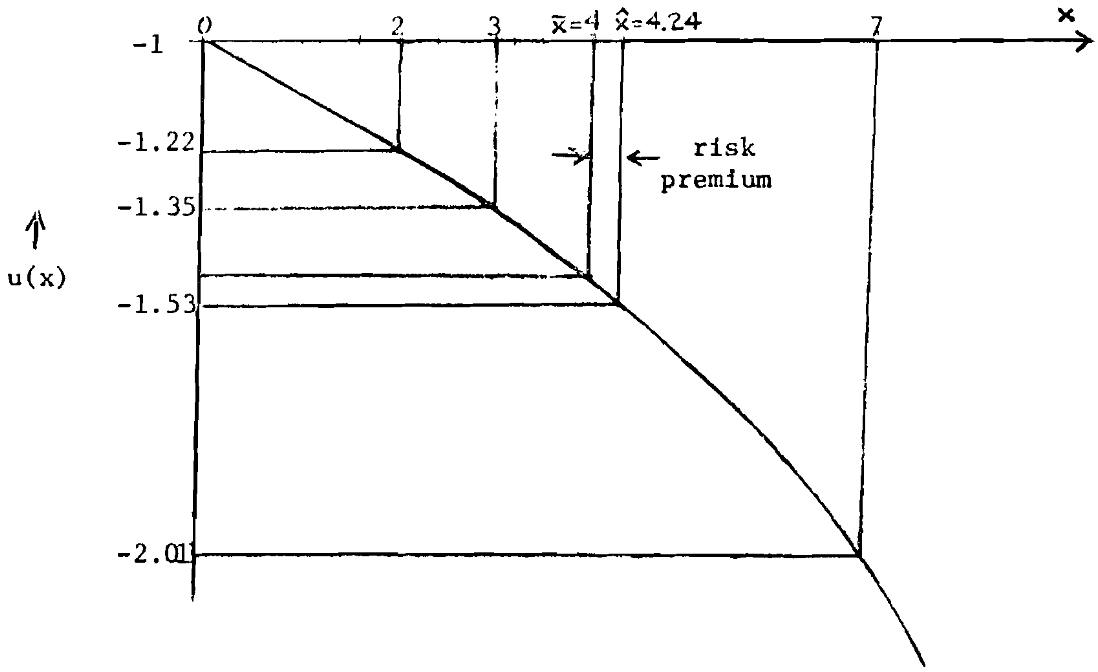


Figure 4.16. A Decreasing Utility Function Illustrating Risk Aversion

The expected utility of this lottery is

$$E[u(\tilde{x})] = \frac{1}{2}\left(e^{-0.2(0)} + e^{-0.2(10)}\right) = 0.568.$$

Calculating the certainty equivalent $\hat{x}$ from

$$e^{-0.2\hat{x}} = 0.568,$$

we find $\hat{x} = 2.83$. Since the expected consequence, $\bar{x} = 5$, the risk premium, $\hat{x} - \bar{x}$, is -2.17. This is all illustrated in Fig. 4.17.

4.8.2. A Measure of Risk Aversion

By a development similar to that for increasing utility functions, we can show that a relevant measure of risk aversion for decreasing utility functions is

$$q(x) \equiv \frac{u''(x)}{u'(x)} = \frac{d}{dx}\left[\log\ (u'(x))\right] . \qquad (4.27)$$

Notice that q(x) is defined almost the same as r(x) in Section 4.5; only a minus sign is different. The reason for this is, as you will see in the examples, is motivated by

Theorem 4.23. If q is positive for all x, then u is concave and the decision maker is risk averse.

Proof. Assume q(x) is positive. Then since u'(x) is negative for decreasing utility functions, u''(x) must be negative implying u(x) is concave. This in turn implies the decision maker is risk averse. ◁

The idea is then consistent with the case of increasing utility functions; positive risk aversion means the decision

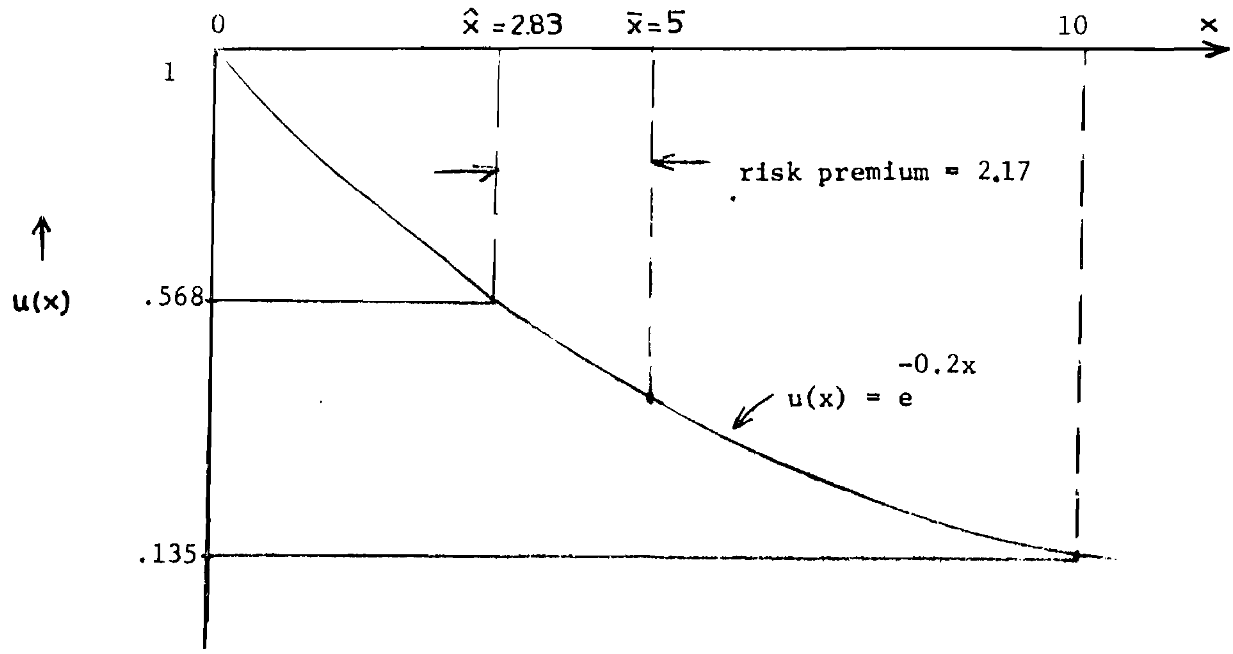


Figure 4.17 A Decreasing Utility Function Illustrating Risk Proneness

maker is risk averse. Also, analogous to the previous case, we have

Theorem 4.24. Two utility functions are strategically equivalent if and only if they have the same risk aversion function.

This says the arbitrariness of a utility function as to scale and origin is eliminated by the risk aversion function although one's attitudes toward risk are retained.

To link this risk aversion function q, which represents the decision maker's risk attitude toward small lotteries with a zero expected consequence, to his risk attitudes toward any lottery we have

Theorem 4.25. If $q_1(x) > q_2(x)$ for all x, then $\pi_1(x,\tilde{x})$, the risk premium for any lottery $\tilde{x}$ given reference x and a utility function with risk function $q_1(x)$, is larger than $\pi_2(x,\tilde{x})$.

Some examples seem appropriate to illustrate these results.

Example 4.20 In example 4.4, we showed using $u(x) = -x^2$, the certainty equivalents for <0,10> and <10,20> were 7.07 and 15.8, respectively. The risk premiums are then 2.07 for <0,10> and 0.8 for <10,20>. Using (4.27) we find the risk aversion function for $u(x) = -x^2$ to be $q(x) = 1/x$. This is positive for $x \geq 0$, so we expect risk premiums for lotteries with consequences in this range to be positive. Our results follow this pattern.

Notice that q is decreasing. Thus the risk aversion in the range 0-10 is greater than the risk aversion in the range 10-20. Hence, you would expect that risk premiums for a particular lottery $\tilde{x}$ in the range 0-10 would be greater than these for an equivalent lottery, $\tilde{x} + 10$, in the range 10-20. The risk premiums for $<0,10>$ and $<10,20>$ bear this relationship. □

Example 4.21 What is the risk aversion for $u(x) = -e^{0.1x}$? Working directly for the definition (4.27),

$$q(x) = \frac{u''(x)}{u'(x)} = \frac{-(0.1)^2 e^{0.1x}}{-0.1e^{0.1x}} = 0.1 .$$

In example 4.18, we used this utility function and found the risk premium for the lottery yielding either x = 2, x = 3, or x = 7 with probability 1/3 was 0.24. Since q is positive, we expected this risk premium to be positive. □

Example 4.22 Suppose $u(x) = e^{-0.2x}$, and we are interested in the risk aversion function. From the definition

$$q(x) = \frac{u''(x)}{u'(x)} = \frac{(0.2)^2 e^{-0.2x}}{-0.2e^{-0.2x}} = -0.2 .$$

Notice this is negative. In example 4.10, we used the same utility function and found the risk premium for $<0,10>$ to be -2.66, also negative. □

This example is an indication of

Theorem 4.26. If q(x) is negative for all x, then u(x) is convex and the decision maker is risk prone.

In Section 4.3, we discussed the possibility of changing attributes in a manner such that the utility function for a new attribute is increasing whereas the utility function for the present attribute was decreasing. Let us consider the effects of such a transformation on the risk aversion of the decision maker. Suppose Y is the attribute of concern and $u(y) = -e^{cy}$ where c is positive. Note u(y) is decreasing and risk averse with $q(y) = c$. Let us define $x = y^* - y$, for all y, where y^* is some standard amount of Y. Let $u_1(x)$ be the utility for x and define it by

$$u_1(x) \equiv u(y^* - x) = -e^{c(y^* - x)} = -(e^{cy^*})e^{-cx}.$$

Since e^{cy^*} is just a positive constant,

$$u_1(x) \sim -e^{-cx}$$

which is increasing with risk aversion $r(x) = c$. The conclusion is that although a transformation was used to change from a decreasing to an increasing utility function, the decision maker's attitudes towards risk were not effected by this change.

Let us try to generalize this notion.

Theorem 4.27. <u>If a transformation of the form $x = y^* - y$ is made to change from a decreasing utility function u(y) to an increasing utility function w(x), the risk aversion q(y) associated with u(y) and the risk aversion r(x) associated with w(x) must be such that $r(x) = q(y^* - x)$, or equivalently, $q(y) = r(y^* - y)$.</u>

In other words, the risk aversion function associated with a particular consequence, either x or y, is not changed by the stated transformation.

Proof. By definition, $q(x) = u''_y(y)/u'_y(y)$ where the subscript designates differentiation with respect to y. An appropriate utility function for x is $w(x) = u(y^* - x)$. Taking derivatives of w(x) with respect to x, we find

$$w'_x(x) = \frac{du(y^* - x)}{dy}\left(\frac{dy}{dx}\right) = u'_y(y)(-1)$$

and

$$w''_x(x) = \frac{dw'_x(x)}{dx} = -\frac{d[-u'_y(y)]}{dy}\left(\frac{dy}{dx}\right) = -u''_y(y)(-1) \; .$$

Substituting these into $r(x) = -w''_x(x)/w'_x(x)$, we have

$$r(x) = -\frac{u''_y(y)}{-u'_y(y)} = \frac{u''_y(y)}{u'_y(y)} = \frac{u''_y(y^* - x)}{u'_y(y^* - x)} \; .$$

Thus $r(x) = q(y^* - x)$. Substituting variables, $q(y) = r(y^* - y)$. ◁

4.8.3 Increasing, Constant, and Decreasing Risk Aversion

The most important category of decreasing utility functions is probably those which are increasingly risk averse. Let us formally define what we mean by this category, and then argue its importance. Concerning decreasing utility functions, we will say an individual is increasingly risk averse if (a) he is risk averse, and (b) his risk premium

$\pi(x,\tilde{x})$ for any lottery $\tilde{x}$ increases as the reference amount x increases. Notice that the words used to define increasingly risk averse in this case are the same as those used to define this concept for monotonically increasing utility functions. However, since the risk premium is defined differently for these cases, the definitions of increasingly risk averse are different.

To be increasingly risk averse implies that the risk premium that a decision maker would be willing to pay to avoid the lottery $\langle x - h, x + h \rangle$ would increase as x increased. This might be quite reasonable if X represented costs, for example. For smaller amounts of X, the decision maker could afford to take the lottery, but as x increased, he might be forced to avoid the same lottery since the potential high cost might cause severe financial problems.

The same reasoning would apply to decision problems within fire departments, where X represents the response time to a fire. A chief may prefer $\langle 1,3 \rangle$ to a response time of 2.2 minutes, and also prefer 4.2 minutes to $\langle 3,5 \rangle$. In other words, he would not be willing to pay a 0.2 minute risk premium to avoid $\langle 1,3 \rangle$, but he would pay this premium to avoid $\langle 3,5 \rangle$. The chief wants to behave more conservatively when dealing with larger response times, so his utility function must be increasingly risk averse.

Another consideration is as follows. Suppose the decision maker's utility function for X was decreasing and

and increasingly risk averse. Then, if we transform to an attribute Y, where a specific value $y = x^* - x$, the decision maker's utility function for Y will be increasing and decreasingly risk averse. That is to say, the increasingly risk averse category of decreasing utility functions corresponds to the decreasingly risk averse category of increasing utility functions. More formally, we have

Theorem 4.28. *If decreasing utility function $u(x)$ is increasingly risk averse and if $y = x^* - x$, the utility function $w(y)$ is increasing and decreasingly risk averse.*

Proof. If $q(x)$ is the risk aversion for $u(x)$ and $r(y)$ is the risk aversion for $w(y)$, the result follows directly from Theorem 4.27. ◀

And thus, the intuitive reasoning given for decreasing risk aversion concerning increasing utility functions is relevant to the current case in point.

All of the important results of Section 4.6 have analogs for decreasing utility functions. For instance,

Theorem 4.29. *The risk aversion $q(x)$ for utility function $u(x)$ is increasing [constant, decreasing] if and only if the risk premium $\pi(x,\tilde{x})$ is an increasing [constant, decreasing] function of x for all $\tilde{x}$.*

Let's try to find some simple examples of increasingly risk averse utility functions.

Example 4.23 Suppose $u(x) = -e^{cx}$, $c > 0$. Then clearly $u'(x) = -ce^{cx}$ and $u''(x) = -c^2e^{cx}$, so the risk aversion $q(x) = c$. Certainly $u(x)$ is decreasing and risk averse, but $q(x)$ is constant, not increasing. □

This example motivates some definitions and a generalization of the result. We will say a decision maker is constantly risk averse if $q(x)$ is a positive constant, constantly risk neutral if $q(x)$ is zero, and constantly risk prone if $q(x)$ is a negative constant. As with increasing utility functions, these conditions place strong restricitons on the shape of the utility function. More specifically, we can show

Theorem 4.30
$$u(x) \sim -e^{cx} \Leftrightarrow q(x) = c > 0, \text{ (constant risk aversion)},$$
$$u(x) \sim -x \Leftrightarrow q(x) = 0, \text{ (risk neutrality)},$$
$$u(x) \sim e^{cx} \Leftrightarrow q(x) = c < 0, \text{ (constant risk proneness)}.$$

Provided the assumptions implying such a utility function were valid, one would only need to determine the certainty equivalent of one simple lottery in order to specify the entire utility function.

Example 4.24 Consider the quadratic utility function of the form
$$u(x) = a - bx - cx^2,$$
where $b > 0$, $c > 0$, and $c \geq -\frac{b}{2c}$. This last condition is necessary

as u is only decreasing in this range. It is a simple matter to calculate

$$q(x) \equiv \frac{u''(x)}{u'(x)} = \frac{2c}{b + 2cx} \quad ,$$

from which one can see $q(x)$ is positive but decreases as x increases. □

In example 4.24, u is decreasingly risk averse. To define this notion more precisely, we will say one is <u>decreasingly risk averse</u> if (a) he is risk averse, and (b) his risk premium $\pi(x,\tilde{x})$ decreases in x for any lottery $\tilde{x}$. Such an attitude is, by definition, opposite of increasingly risk averse.

<u>Example 4.25</u> Suppose $u(x) = \log(b-x)$. Then, $u'(x) = -1/(b-x)$ and $u''(x) = -1/(b-x)^2$, so $q(x) = 1/(b-x)$. Clearly $q(x)$ is positive and increasing in x for $x < b$. This implies $u(x)$ is increasingly risk averse for $x < b$. □

<u>Example 4.26</u> Let $u(x) = -e^{ax} - be^{cx}$, where $a > 0$, $b > 0$, and $c > 0$. If $a = c$, then $u(x) = -(1+b)e^{cx}$ which is constantly risk averse as we have shown. If $a \neq c$, then $u'(x) = -ae^{ax} - bce^{cx}$ and $u''(x) = -a^2e^{ax} - bc^2e^{cx}$, so

$$q(x) = \frac{a^2e^{ax} + bc^2e^{cx}}{ae^{ax} + bce^{cx}} \quad .$$

In this case, the risk aversion $q(x)$ is always positive and increasing in x. Thus $u(x)$ is increasingly risk averse if $a \neq c$. Assuming $a < c$, which can be done without loss of generality, the risk aversion is slightly larger than a for large negative amounts of x, increases to $(a^2 + bc^2)/(a + bc)$ at x = o, and approaches c as x becomes positively large. ◻

In this example, we used a general result analogous to one for increasing utility functions. That is,

Theorem 4.31 A utility function, which is the weighted sum of two or more utility functions which are increasingly or constantly risk averse on the interval $[x^o, x^*]$, is increasingly risk averse on $[x^o, x^*]$ except on subintervals where the weighted utility functions have equal and constant risk aversion. Then it is constantly risk averse.

Note that in example 4.26, if we set $u_1(x) = -e^{ax}$ and $u_2(x) = -e^{cx}$, then u(x) is a weighted sum, namely $u(x) = u_1(x) + bu_2(x)$. Now u_1 and u_2 are each constantly risk averse. If they don't have equal risk aversion functions, that is if $a \neq c$, then u must be increasingly risk averse, and if they do have equal risk aversion functions, then clearly u must be constantly risk averse.

As we did with increasing utility functions, we could categorize the monotonically decreasing utility functions which are risk prone as increasingly risk prone, constantly risk prone, or decreasingly risk prone. Also we could define and investigate proportional risk aversion in the context of monotonically decreasing utility functions. However, at this point, we feel the exercise would provide little, if any, insight, so it is omitted.

4.8.4 Nonmonotonic Utility Functions

Our definitions for risk averse and risk prone are the same for nonmonotonic preferences as they were for the monotonic cases. Specifically, one is said to be risk averse if he prefers the expected consequences of any nondegenerate lottery to the lottery itself, and one is said to be risk prone if he prefers any nondegenerate lottery to its expected consequence. From these definitions one could prove

Theorem 4.32 Concerning nonmonotonic preferences, a decision maker is risk averse [risk prone] if and only if his utility function is concave [convex].

Examples of risk averse and risk prone nonmonotonic utility functions are given in Fig. 4.18.

As we illustrated earlier in Section 4.3, the certainty equivalent for nonmonotonic utility functions is not necessarily unique. Because of this, there are no alternate definitions of risk averse and risk prone in terms of the certainty equivalent as there was for monotonic utility functions. Also, the risk premium cannot be usefully defined for nonmonotonic utility functions. In addition,

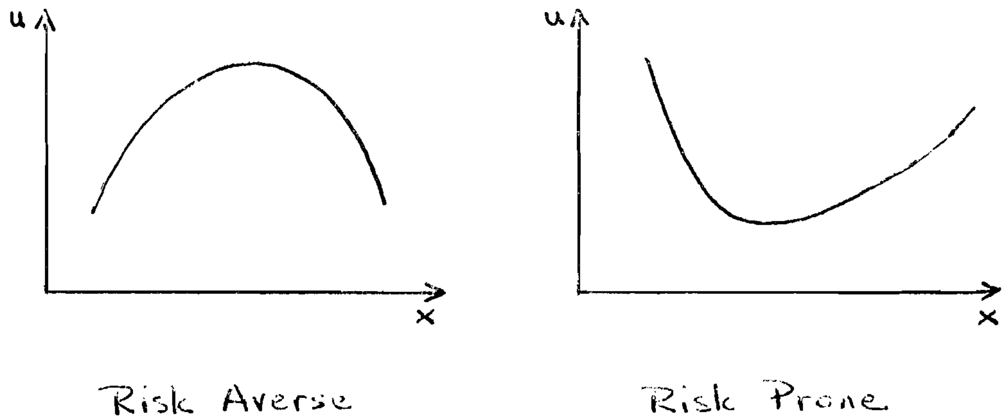


Fig.4.18. Risk Properties of Nonmonotonic Utility Functions

for nonmonotonic utility functions, the first derivative of u(x) is either undefined or zero for at least one value of x. Hence, a measure of risk aversion similar to the r(x) and q(x) in the monotonic cases would not be defined for all x. Perhaps an alternate definition of a local risk aversion exists for this case, but this seems to be an academic point. For operational problems, a reasonable approach would be to divide the range of the attribute into intervals such that preferences are monotonic in each interval, and then, treat each interval separately using the theory relevant to the respective cases.

4.9 A PROCEDURE FOR ASSESSING UTILITY FUNCTIONS

From the heading of this section, one might think it contains "a procedure for assessing utility functions" applicable to anyone at any time, that is, a general procedure. But in fact, it contains "a procedure for assessing utility functions" applicable to some of the people some of the time, maybe. So, clearly the question that must be addressed before we begin the main topic of this section is 'Given the situation as stated, what is the relevance of this material?'

To make sure there is no misunderstanding, note that we did not state that one cannot evaluate a utility function for the decision maker in most problems. It was stated, however, that the procedure to be discussed now is not necessarily appropriate in many cases. The main reason for this is that assessment of utility functions is as much of an art as it is a science, and as such no single set of rules can be laid down which invariably result in a utility function. In fact, there are not only many different techniques* for evaluating utility functions, but numerous variants of each of these. Also, which technique might be best in a certain situation would be very hard to predict beforehand and would depend on the particular decision maker and the context of the problem in addition to many less obvious factors. Thus, this section does not contain a generally applicable procedure simply because there isn't one.

The <u>basic</u> ideas, however, which one uses in assessing a utility function remain more or less the same for all the procedures. That is, regardless of the technique being used to assess a utility function, the specific points or objectives which must be considered and accomplished by any assessment procedure are essentially the same. To help clarify this, let us divide procedures into the following five steps:

(1) preliminaries to actual assessment,

(2) specifying the relevant qualitative characteristics,

---

*For example, see Mosteller and Nogee [1951]; Davidson, Suppes, and Siegel [1957]; Becker, DeGroot, and Marschak [1964]; and Schlaifer [1969].

(3) specifying quantitative restrictions,

(4) choosing a utility function,

(5) checking for consistency.

The different assessment procedures result from the numerous ways of carrying out each of these five steps. Although, this division allows us to emphasize exactly what goes into the assessment of a utility function, in practice the distinctions between certain steps may not be so clear.

Before beginning the main discussion, it should be mentioned that the level of detail given here is much greater than would be required for the assessment of a given decision maker's utility function. The analyst, being aware of all the small points described, will no doubt find it convenient to skip explicitly many of them in most circumstances. For example, the preliminaries to assessment may be omitted when dealing with someone familiar with decision analysis, since this step is to insure that analyst and decision maker are speaking the same language.

4.9.1 Preliminaries to Actual Assessment

Recall from chapter one that the paradigm of decision analysis is divided into five steps: pre-analysis, structuring the problem, assessing the judgmental probability distributions, assessing preferences for consequences, and maximizing exected utility. Before assessing the preferences, we would have explained the concept of decision analysis to the decision maker and with his help, structured the problem. Thus, we can assume that the decision maker realizes the purpose in assessing his preferences and is sufficiently motivated to think hard about his feelings for the various consequences. It is at this point that we begin to assess his utility function.

Before any assessments are made, it should be clear to the decision maker that the preferences we are interested in are his. It must be understood that there are no objectively correct preferences; that the preferences of importance represent the subjective feelings of the decision maker. At any time if the decision maker feels uncomfortable with any of the information he has offered about his subjective feelings, it is perfectly all right, in fact, necessary for a correct analysis, for him to change his mind. This is one of the purposes of decision analysis, to require the decision maker to reflect on his preferences and hopefully straighten them out in his own mind.

Let us hasten to add one caveat at this point. Experience has shown that in assessing utility functions for the first time, many individuals fall into certain standard traps. They respond to certain hypothetical questions and perhaps even feel comfortable about their responses. But then they are aghast at some of the implications of their judgmental inputs. The experienced analyst may wish to point out these implications to the assessor and by various compromises help guide him over these troublesome rough spots. Now, of course, there is a danger in doing this since we are seeking the decision maker's preferences and not the analyst's, but some healthy tensions might force the decision maker to think a bit deeper about his problem. If the intervention of the analyst is crude and overpowering, then, of course, one subverts the whole process of trying to organize the decision maker's preferences into a coherent whole.

In this chapter on unidimensional utility theory we are concerned with the case where each possible consequence of any act can be adequately described in terms of a single attribute. Let X be the evaluator function, which associates to any consequence Q say, the real number $x = X(Q)$. It is crucial that the decision maker understands the orientation of the scale: Are higher x numbers more or less desirable? Do preferences increase with x up to a point and then decrease?

In some contexts the attribute X may be quite natural and the x-scale can be given in natural physical units like monetary assets, share of the market, lives saved, or time elapsed. In other contexts the values on the x-scale may involve subjective appraisals, such as an index for comfort, for aesthetics, for functionality. No matter how we find the x-values we assume that it is meaningful to ask whether we prefer a consequence x_1 to consequence x_2.

Next, we want to limit the region over which we must assess preferences to as small a region as reasonable. From the problem structure, the decision maker should be able to bound the possible amount which x could assume. Then we would choose x^o and x^* such that any possible x is bounded by x^o below and x^* above. These values should be chosen for convenience and meaningfulness to the decision maker. For instance, if x ranged from 0 to 8.75 in the specific units, we might define $x^o = 0$ and $x^* = 10$. A value of $x^* = 10{,}000$, for example, probably would have little meaning to the decision maker. The preferences which we eventually assess must only be those for consequences x with $x^o \leq x \leq x^*$.

As a final check on the decision maker's understanding of how consequences are represented as real numbers, we might ask him whether or not he prefers consequence T to consequence S in Figure 4.19, where the points S and T should be chosen such that it is clear to us, the analyst, that the decision maker would almost for sure prefer a particular one. If the decision maker's preference in this case agreed with the expected result, we could proceed to assess the utility function. If not, it would seem desirable to ask the decision maker to clarify his reasoning and perhaps then to repeat some or all of the familiarization process.

Enough has been said about the preliminaries. The basic idea is to acquaint the decision maker with the framework which we use in assessing his utility function.

All these preliminaries are theoretically trivial and you might feel that we are talking down to you by emphasizing the self-evident. However, we have made many errors ourselves

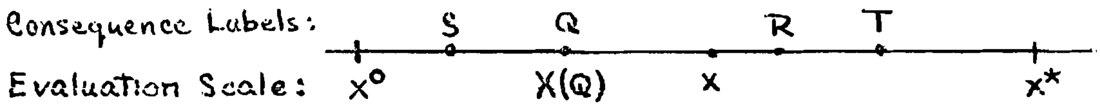


Figure 4.19. The Evaluation Space for a Single Attribute

in helping others assess utility functions and it is often these simplistic preliminaries that foul up the procedure.

4.9.2 Specifying the Relevant Qualitative Characteristics

At an early stage in the assessment process we should determine whether or not the utility function u is monotonic. Referring to Figure 4.19, we ask the decision maker if S or Q is more preferable. Suppose Q is preferred to S. Then we might ask whether T is preferred to R; and again, assume it is. A few more questions of this nature may be appropriate, but finally we ask: "If x_k is greater than x_j, is x_k always preferred to x_j?" For example, from the previous responses, we would probably expect an answer of yes, implying that $u(\cdot)$ is monotonically increasing in x. If this did not agree with our own understanding of the consequence, we should offer our reasoning to the decision maker and recheck his preferences. This serves to educate the decision maker, not to bias him, and hopefully, it forces him to think hard about his preferences.

Next, we want to determine whether u is risk averse, risk neutral, or risk prone. First we ask the decision maker if he prefers $\langle x + h, x - h \rangle$ or x for some arbitrarily chosen amounts of x and h. If he prefers the lottery, we have reason to believe he might be risk prone, whereas if he prefers the expected consequence x, this indicates he might be risk averse. The same question should be repeated using many different amounts for either x or h while holding the other amount fixed. If the lotteries are chosen to cover the entire range of possible consequences and if the expected consequence is always preferred, it is reasonable to assume the decision maker is risk averse. In similar circumstances, if the lottery is always preferred, he is risk prone. And, of course, indifference between each lottery and its expected consequence indicates risk neutrality. For a mathematically sophisticated decision maker who preferrred x to the initial lottery $\langle x + h, x - h\rangle$, we might simply ask "If x and h are allowed to vary over the range of possible consequences, would you always prefer x to $\langle x + h, x - h\rangle$?" An affirmative response is a sufficient condition for risk aversion.

The less sophisticated decision maker may require a more specific version of this procedure. For example, we might divide the range of attribute X into ten equal segments, where division points are denoted by x_0, $x_1, \ldots,$ and x_{10} respectively. This notation is illustrated in Figure 4.20. Now we ask the decision maker whether or not he prefers $\langle x_2, x_0\rangle$ or x_1. For risk aversion, x_1 should be preferred. Similarly we ask for preferences between $\langle x_{i+1}, x_{i-1}\rangle$ and x_i for $i = 2, 3, \ldots, 9$. If u is risk averse, the certain consequence (which is the expected consequence) should be preferred to the lottery in all these cases. Given the decision maker answered all the questions in this manner, we would be justified in assuming he is risk averse. If he always preferred the lottery, we would assume he is risk prone.

It would now be useful to determine if u is increasingly,

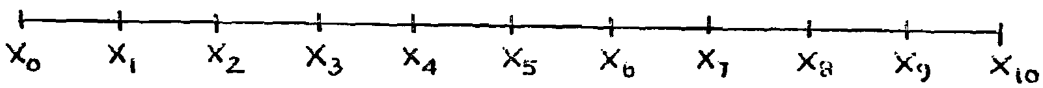


Figure 4.20 The Notation Used in Ascertaining Risk Properties

decreasingly, or constantly risk averse. One method to do this involves finding the certainty equivalent $\hat{x}_1$, such that the decision maker is indifferent between $\hat{x}_1$ and $\langle x_2, x_0 \rangle$. A procedure for evaluating such a cerainty equivalent is given in the next subsection. Also, we would like to determine certainty equivalent $\hat{x}_i$ which is indifferent to $\langle x_{i+1}, x_{i-1} \rangle$ for $i = 2, 3, \ldots, 9$. For increasing utility functions, if the risk premium $(x_i - \hat{x}_i)$ decreases [increases, is constant] as i increases, then u is decreasingly [increasingly, constantly] risk averse. It may be rather difficult to determine the $\hat{x}_i$'s exactly, but the decision maker should be able to qualitatively answer whether $(x_i - \hat{x}_i)$ is increasing, decreasing, or remaining constant as i increases without actually specifying the amounts of the x_i's. It is possible that $(x_i - \hat{x}_i)$ may be increasing in certain regions of X and decreasing in other regions. This information is also valuable.

For the more sophisticated subject the analyst might ask him for his risk premium for a lottery of the form $\langle x - h, x + h \rangle$ for a specific x and h. Then he would be asked how this risk premium would behave as x is increased with h held fixed. If, as is often the case for monetary assets, this risk premium decreases as x increases there is a strong presumption of decreasing risk aversion. In implementing this procedure one can often ascertain that the subject is decreasingly risk averse without ever forcing him to give a specific numerical value for the risk premium of any specific lottery $\langle x - h, x + h \rangle$. It is encouraging to note how often subjects feel comfortable with these qualitative type questions.

We've just illustrated a few ways of determining some possible qualitative characteristics of u, namely, monotonicity, risk aversion, decreasing risk aversion, etc. These methods have proven to be important in many decision problems. In other problems, however, a characteristic of main interest may be propotional risk aversion.

In a style similar to that just illustrated, the analyst should be able to devise a simple technique to ascertain which proportional risk characteristics apply. Such a technique should take into consideration the problem context and the abilities of the decision maker.

After the qualitative characteristics have been specified, one needs to assess quantitive utility values for a few points on X. The analyst could either then fair in a "smooth" utility function satisfying the qualitative characteristics and quantitative assessments or perhaps assess appropriate parameter values for an appropriate family of utility functions that exhibit the qualitative specifications already elicited from the subject. Let's consider these quantitative assessments.

4.9.3 Specifying Quantitative Restrictions

Our step three in assessing a utility function is determining some quantitative restrictions. That is, we want to fix the utilities of a few particular points on the utility function. This usually involves determining the certainty equivalents for a few fifty-fifty lotteries. Refer to Fig. 4.21 for the meaning of the consequences x_a, x_b, etc., and assume we want to determine the certainty equivalent for $\langle x', x'' \rangle$.

We begin by asking the decision maker if he prefers $\langle x', x'' \rangle$ or x_a. The consequence x_a is chosen such that a particular answer is expected. Suppose the decision maker prefers the lottery to x_a and this agrees with our expectation. Then we ask the decision maker whether he prefers $\langle x', x'' \rangle$ or x_b, where x_b is chosen so that we expect x_b would be preferred. Assume this is the case. Next, we inquire about the preferences of $\langle x', x'' \rangle$ relative to x_c. Since x_c is "near" x_a, we somewhat expect that the lottery will be preferred to x_c, but perhaps not. We continue with this convergence procedure until a consequence $\hat{x}$ is reached such that $\langle x', x'' \rangle$ and $\hat{x}$ are equally desirable (or undesirable) to the decision maker.*

If the decision maker indicates any preferences which we do not feel represent his "true" preferences, this should be pointed out and discussed again. Provided the assessments are correct in the sense that the decision maker really is

*The questions should be in a framework that the decision maker understands and finds reasonable. For a good example of this, see the work of Grayson [1960], which is briefly discussed in Section 4.10.

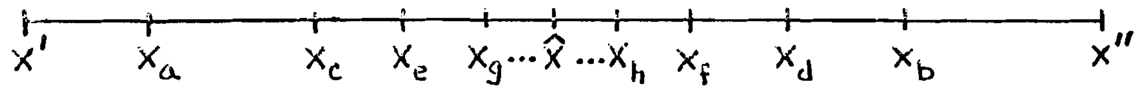


Figure 4.21.

A Convergence Technique for Assessing a Certainty Equivalent indifferent between $\hat{x}$ and $\langle x', x''\rangle$, then $\hat{x}$ is the certainty equivalent for that lottery. And of course, the utility assigned to $\hat{x}$ must equal the expected utility of $\langle x', x''\rangle$. More specifically, we set

$$u(\hat{x}) = \frac{1}{2}\, u(x') + \frac{1}{2}\, u(x'') \ .$$

Using this procedure, we can determine the certainty equivalents for some lotteries which will help us specify the decision maker's utility function. In particular, suppose we are interested in a utility function $u(\cdot)$ for all x such that $x_0 \leq x \leq x_1$. The reason for this notational change will soon be clear.

A reasonable first step would be to assess the certainty equivalent $x_{.5}$ for the lottery $\langle x_1, x_0\rangle$. Then, clearly

$$u(x_{.5}) = \frac{1}{2}\, u(x_1) + \frac{1}{2}\, u(x_0) \qquad (4.28)$$

Next, we assess the certainty equivalents for $\langle x_1, x_{.5}\rangle$ and $\langle x_{.5}, x_0\rangle$, which we will designate as $x_{.75}$ and $x_{.25}$ respectively. And, obviously,

$$u(x_{.75}) = \frac{1}{2}\, u(x_1) + \frac{1}{2}\, u(x_{.5}) \qquad (4.29)$$

and

$$u(x_{.25}) = \frac{1}{2}\, u(x_{.5}) + \frac{1}{2}\, u(x_0) \qquad (4.30)$$

Suppose the decision maker's preferences are increasing in x and that $x_1 > x_0$, then we can arbitrarily set

$$u(x_0) = 0 \qquad (4.31)$$

and

$$u(x_1) = 1 . \tag{4.32}$$

Substituting, these into (4.28), (4.29), and (4.30), we easily obtain

$$u(x_{.5}) = 0.5, \tag{4.33}$$

$$u(x_{.75}) = 0.75, \tag{4.34}$$

and

$$u(x_{.25}) = 0.25. \tag{4.35}$$

Equations (4.31) through (4.35) fix five points on the utility function for X as shown in Figure 4.22. A utility function with the previously specified qualitative characteristics can be faired through these points.

Before this is done, however, some simple consistency checks should be included in the procedure. For instance, we can assess the decision maker's certainty equivalent $\hat{x}$ for $\langle x_{.75}, x_{.25}\rangle$. To be consistent, $\hat{x}$ should equal $x_{.5}$ since $u(x_{.5}) = 0.5$ and

$$u(\hat{x}) = \frac{1}{2}u(x_{.75}) + \frac{1}{2}u(x_{.25}) = 0.5.$$

Also, we now have the necessary information for a simple check on whether the utility function is risk averse or risk prone. For u increasing, recall that the certainty equivalents $x_{.25}$, $x_{.5}$, and $x_{.75}$ are less than the expected consequences of their respective lotteries if u is risk averse. These certainty equivalents must be larger than the expected consequences if u is risk prone. For monotonically decreasing utility functions, as previously

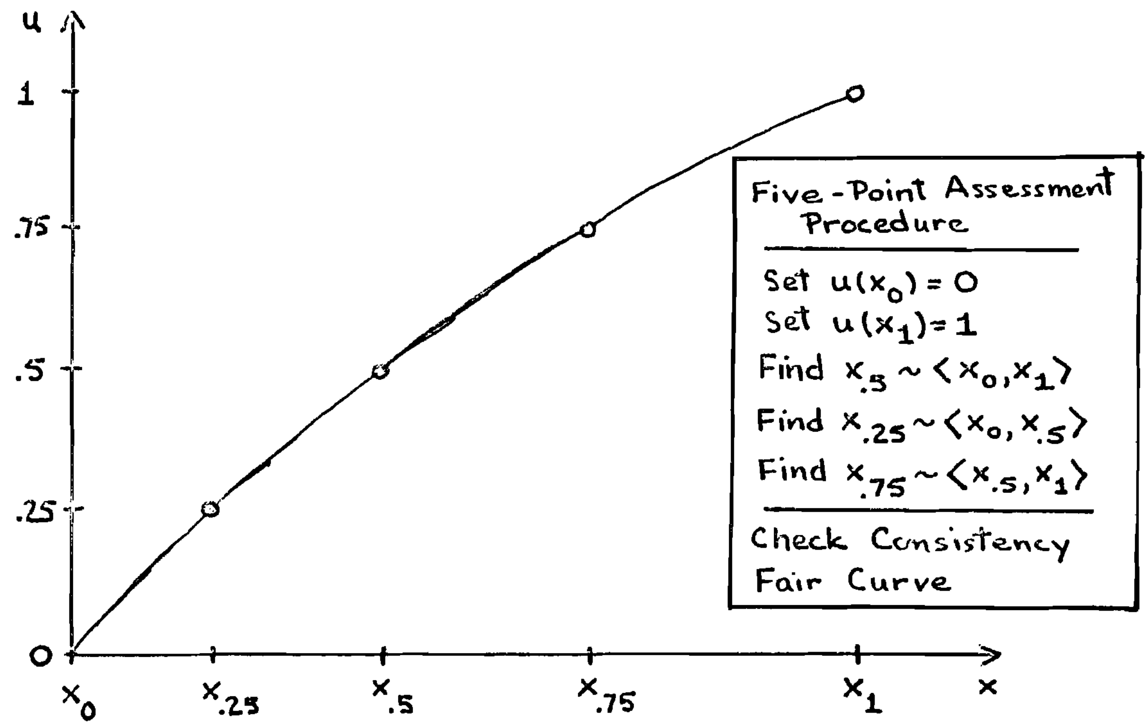


Figure 4.22. A Five-Point Assessment Procedure for Utility Functions

discussed, just the reverse is true.

When these consistency checks reveal inconsistent preferences, the discrepancies should be pointed out to the decision maker, and part of the assessment procedure must be repeated to iron out the differences and obtain consistent preferences. This iterative procedure hopefully results in a "better" statement of the decision maker's preferences.

Before proceeding any further, the great amount of overlap between determining the qualitative characteristics of a utility function and specifying qualitative restrictions should be explicitly mentioned. To take a simple example, suppose that in checking the risk aversion of utility function $u(\cdot)$ for $0 \leq x \leq 1000$, the decision maker stated 400 was the certainty equivalent for the lottery $<1000,0>$. We noted this and then asked "Is the expected consequence always preferred to a lottery?" A positive response indicated the decision maker was risk averse. Next suppose it was determined that he was constantly risk averse, so his preferences could be represented with the utility function $u(x) = -e^{-cx}$. Since this function has only one parameter, namely c, we do not need to get any more quantitative restrictions since we already know

$$u(400) = \frac{1}{2}\, u(1000) + \frac{1}{2}\, u(0) .$$

From this we can calculate a value for c. Of course, it will often be prudent to make consistency checks on this value. In Section 4.7, we indicated how the one parameter families of constant proportional risk averse utility functions could be assessed with the answer to one question. This also illustrated the interaction among the steps of a utility assessment--steps that we have identified mainly for discussion purposes.

We now raise two points about assessments that are discussed by Schlaifer [1969]. First, the consequences used to assess utility functions must be psychologically real to this decision maker. As an example, if we are interested in assessing someone's utility function for monetary amounts between zero and twenty-thousand dollars, he should not be asked to consider consequences like a million dollars. This consequence might be inconceivable to him and inconsistent assessments would likely result. The second point is that the differences between consequences must be psychologically real to the decision maker. Again for the same monetary utility function, assessing the certainty equivalent for <\$0, \$10> would likely not provide very useful information, since any extrapolation of the result to the range of interest would have little relevance. In terms of the total range of money considered, \$0, \$10, and the certainty equivalent would probably be thought of as essentially equal in preference for all practical purposes.

4.9.4 Choosing the Utility Function

After we have qualitatively determined the characteristics of the utility function and quantitatively assessed the preferences of approximately five consequences and satisfied ourselves that the results represent the true feelings of the decision maker, we next fair in a smooth utility function. However, having obtained this information, the analyst is faced with several questions. First, are the qualitative and quantitative assessments consistent, that is, does a utility function exist which simultaneously satisfies all of them? If there is such a utility function, how restrictive are these assessments, and how should an appropriate utility function be determined? If there is not such a utility function, how should one proceed to obtain a consistent set of assessments?

A method for addressing these questions involves first finding a parametric family of utility functions which possesses the relevant characteristics, such as risk aversion, etc., previously specified for the decision maker. Then using the quantitative assessments, that is, the certainty equivalents, we try to find a specific member of that family which is appropriate for the decision maker. The information on certainty equivalents is used to specify values for the parameters of the original family of utility functions. If we are lucky, we will find a utility function satisfying all the qualitative and quantitative assessments simultaneously. Unfortunately, no general procedure exists for either determining whether a given set of qualitative and quantitative assessments are consistent or indicating an appropriate functional form of the utility function when the assessments are consistent. To our knowledge, the most advanced work on these problems is that of Meyer and Pratt [1968], who have answered these questions for some important cases.*

The first situation concerns the case where certainty equivalents for some simple lotteries are given and regions of risk aversion and risk proneness specified. Increasing and decreasing risk aversion are not considered. They prove a utility exists satisfying these assessments provided certain linear constraints are satisfied. Finding bounds for the acceptable utility function is essentially a linear programming problem.

The second important case is when the decision maker is decreasingly risk averse and an arbitrary number of certainty equivalents is given. Meyer and Pratt develop and illustrate an algorithm which checks the consistency of these assessments and bounds the possible utility functions satisfying the constraints.

As a simple illustration of a couple points, suppose the decision maker's utility function was monotonically increasing in x and decreasingly risk averse. From Section 4.6, we know a family of utility functions which satisfies these characteristics is

$$u(x) = h + k(-e^{-ax} - be^{-cx}), \qquad (4.36)$$

---

*In their article, Meyer and Pratt [1968] address consistency questions in two situations concerning increasing utility functions. Using their methods, it would be a straight forward exercise to obtain results analogous to theirs for decreasing utility functions.

where a,b,c, and k are positive constants. Using (4.36) to evaluate the utilities of the consequences in (4.31) through (4.35) will give us five equations with five unknowns. Then, provided these equations can be solved subject to the restrictions on the parameters, they will give us the specific member of (4.36) which represents the decision maker's preferences*. If they have no solution the analyst is faced with implictly weighing the disadvantages of choosing an "almost appropriate" utility function against the disadvantages of further search for a "more appropriate" utility function, with a knowledge that further search might not improve matters. Thus, in many situations, choosing a utility function subject to the given constraints is somewhat of a heuristic search process. Unfortunately, we can't offer any clear-cut procedures for solving such a problem. However, if we have obtained a utility function which satisfies almost all of the constraints and which is not grossly incompatible with any of the others, then due to the subjectiveness of utility assessments, it would seem appropriate for the decision maker to operate with this utility function**.

---

*See Section 4.10.3 for a brief description of a computer program that addresses this problem.

**See Hammond [1974], where he indicates that in some situations, an easy-to-use simple utility function can be substituted for a more complex utility function which is not precisely known.

The final point we wish to address in this subsection concerns utility functions which are not monotonic. The theory for this case is not so nice, but operationally, the problem is only a little more difficult than in instances where the utility function is monotonic. Suppose one's preferences for X increase up to x_m and then decrease. A reasonable way to quantify these preferences is to assess one utility function $u_1(x)$ for $x \leq x_m$ and another $u_2(x)$ for $x \geq x_m$. Obviously $u_1(x)$ is monotonically increasing in x and $u_2(x)$ is monotonically decreasing, and the theory previously discussed is applicable to those cases. The only remaining problem would be to correctly scale u_1 and u_2. First we would fix one point on each utility function by setting $u_1(x_m) = u_2(x_m)$. Secondly, we could determine $x' < x_m$ and $x'' > x_m$ such that the decision maker is idifferent between x' and x". Then, of course, we set $u_1(x') = u_2(x'')$, which fixes a second point on each utility function. Having completed this, a utility function valid for all x is

$$u(x) = \begin{cases} u_1(x) & , \ x \leq x_m, \\ u_2(x) & , \ x \geq x_m. \end{cases}$$

4.9.5 Consistency Checks

There are many different consistency checks which can be used to detect errors in the decision maker's utility function. By an error, we mean that the utility function which we have assessed for him does not represent his true preferences. We will discuss two consistency checks in this subsection. With these, as well as those discussed throughout this section*, as a guide, the decision analyst should have no trouble developing other checks designed to uncover discrepancies in a utility function.

One generally useful check involves asking the decision maker his preference between any lottery and any consequences, or between two lotteries. In both cases, the expected utility of the preferred situation must be greater in order to be consistent.

A more "subtle" consistency check is illustrated by the following example. Suppose the decision maker's utility function is being assessed over the attribute 'incremental monetary assets' so zero is the status quo. And let us suppose we want $u(x)$ for $-100 \leq x \leq 100$. Experience has indicated that often in practice, the decision maker may seem to be risk averse in the entire range except for small negative amounts, say for $-10 \leq x \leq 0$, where he has indicated that he would rather face the lottery $A \equiv \langle -10, 0 \rangle$ than take the sure consequence $B \equiv -4$. Note that consequence B is essentially <u>payment</u> of 4 units. The analyst may be a bit skeptical about the appropriateness of the risk behavior and probe its implications with the decision maker.

---

*For instance, earlier in this section, two techniques for determining whether or not one is risk averse were described: one concerned preferences between lotteries and their expected consequences and the other involved evaluating certainty equivalents of some lotteries. Either of those can be used as a consistency check of the other.

Suppose an option C, defined as the decision maker pays 4 and then immediately must face the lottery <-6,4>, is displayed, along with A and B, for the decision maker. The options A, B, and C are illustrated in Figure 4.23.

We already know the decision maker indicated $A \succ B$. Then he is asked for his preference between B and C. He responds "In both situations, I must first pay four units. Then with B, I am finished. However, with C, I must face the additional lottery <-6,4> which has a negative expected value of -1. My preference is clear, I prefer B." Therefore $B \succ C$.

But now, the analyst asks "Compare A and C, and give consideration to the total impact to yourself." Thinking out loud, the decision maker says "Lottery A is clear, I either get -10 or 0 with a fifty-fifty chance. For C, I lose 4 and then either gain it right back or lose 6 more. I guess with C, I also either get -10 or 0 with a fifty-fifty chance, so I should be indifferent between A and C."

The punch line should be clear, the decision maker has said A is preferred to B and B is preferred to C, but then C is indifferent to A. An intransitivity has been created. When this is pointed out, most subjects are a little surprised and indicate they do not want such inconsistencies in their preference structure. On reflection, subjects often will feel comfortable maintaining that $B \succ C$ and $C \sim A$. Hence they are forced to conclude that $B \succ A$. This can lead to a removal of the risk prone segment of the utility function

Compare A, B, and C starting at the Status Quo SQ = 0

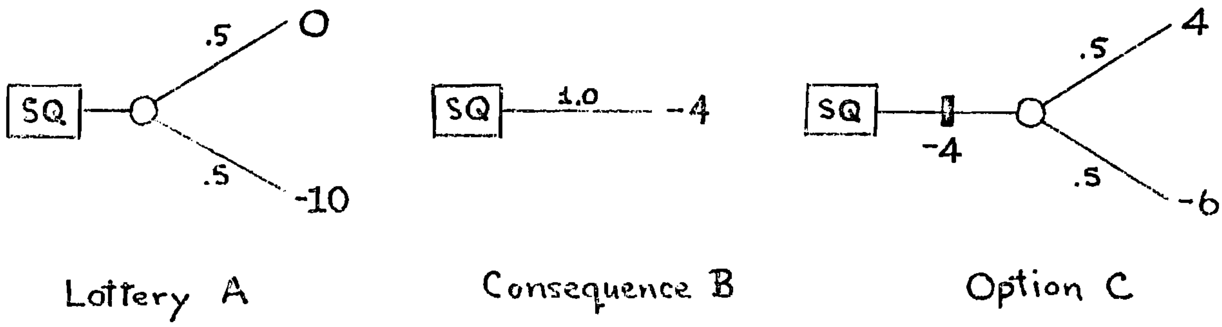


Figure 4.23. A Consistency Check which Generates Intransitivities To Promote Reconsideration

in the range $-10 \leq x \leq 0$.

An important part about this example is that through the facility of the analyst, the decision maker ends up teaching himself his preferences and in the process, help himself to 'straighten out his head.'

Obviously, for utility functions implying a complex preference structure, both the need and opportunity for meaningful consistency checks increase. As has been mentioned before, if the checks produce discrepancies with the previous prefernces indicated by the decison maker, these discrepancies must be called to his attention and parts of the assessment procedure should be repeated to acquire consistent preferences. Once a utility function is obtained which the decision maker and the analyst feel represents the true preferences of the decision maker, one can proceed with the analysis.

4.9.6 Using the Utility Function

In this subsection, we will consider two practical topics which are useful in sensitivity analysis. This ties in with the consistency checks and with the entire assessment procedure since it helps indicate how precise our assessments need to be.

Simplifying the Expected Utility Calculations. Often one deals with utility functions which have exponential terms. For instance, a common example is the constantly risk averse utility function for X of the form

$$u(x) = -e^{-cx} \tag{4.37}$$

where c is a positive constant. Another very important example is the decreasingly risk averse utility function

$$u(x) = -e^{-ax} - be^{-cx}, \qquad (4.38)$$

where a, b, and c are positive constants. There is a simple method to calculate expected utility when such utility functions are valid and when the possible consequences are described with a probability distribution function.

The exponential transform $T_x(s)$ for a probability distribution function f(x) which is defined by

$$T_x(s) \equiv E[e^{-sx}] = \int_{-\infty}^{\infty} e^{-sx} f(x)\, dx \qquad (4.39)$$

where $\int$ indicates summation for discrete distributions, has been calculated for most common probability distribution functions. Table 4.6 gives a partial list. Given a utility function of the form (4.37) and a course of action resulting in a random outcome x described by probability distribution f, the expected utility of this course of action can easily be calculated by observing from (4.39) that

$$E[u(\tilde{x})] = \int_{-\infty}^{\infty} u(x)\, f(x)\, dx = \int_{-\infty}^{\infty} -e^{-cx}\, f(x)\, dx = -T_x(c). \qquad (4.40)$$

If the utility function was of the form (4.38), the expected utility could be calculated from

$$E[u(\tilde{x})] = \int_{-\infty}^{\infty} (-e^{-ax} - be^{-cx})\, f(x)\, dx = -T_x(a) - b\, T_x(c) \qquad (4.41)$$

In a similar manner, the Mellin transform $M_x(s)$ for a probability distribution f(x) is defined by

Table 4.6

Exponential Transforms of Some Common Probability Distributions

Probability Distribution	$f(x)$	$T_x(s) \equiv E[e^{-sx}]$
Beta	$\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1}, \; 0 \le x \le 1; a>0, b>0$	$1+\sum_{n=1}^{\infty} \frac{(-s)^n (a)(a+1)\cdots(a+n)}{n!(a+b)(a+b+1)\cdots(a+b+n)}$
Binomial	$\frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}, \; x=0,1,\ldots,n; \; 0<p<1$	$(pe^{-s}+1-p)^n$
Cauchy	$\frac{1}{\pi}\frac{a}{a^2+(x-b)^2}, \; -\infty<x<\infty; \; a>0$	$e^{-bs-a\lvert s\rvert}$
Exponential	$\lambda e^{-\lambda x}, \; x \ge 0; \; \lambda>0$	$\frac{\lambda}{\lambda+s}$
Gamma	$\frac{\lambda}{\Gamma(r)}(\lambda x)^{r-1} e^{-\lambda x}, \; x \ge 0; \; \lambda>0, r>0$	$\left(\frac{\lambda}{\lambda+s}\right)^r$
Geometric	$p(1-p)^{x-1}, \; x=1,2,\ldots; \; 0<p<1$	$\frac{pe^{-s}}{1-(1-p)e^{-s}}$
Normal	$\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\bar{x})^2}{2\sigma^2}}, \; -\infty<x<\infty; \; \sigma>0$	$e^{-s\bar{x}+s^2\sigma^2/2}$
Poisson	$\frac{\lambda^x e^{-\lambda}}{x!}, \; x=0,1,\ldots; \; \lambda>0$	$e^{\lambda(e^{-s}-1)}$
Uniform	$\frac{1}{b-a}, \; a \le x \le b; \; a<b$	$\frac{e^{-sa}-e^{-sb}}{s(b-a)}$

$$M_X(s) \equiv E[x^s] = \int_{-\infty}^{\infty} x^s \; f(x) \; dx. \qquad (4.42)$$

This transform has also been tabulated for many common probability distributions and could be used in expected utility calculations where the utility function contains powers of x.

Parametric Analyses. The sophisticated analyst would usually include a sensitivity analysis in his work. For decision problems, this might mean that the sensitivity of the best decision to parameters of the utility function be determined . For example, suppose from his characteristics, we found a decision maker's preferences would be quantified by the utility function

$$u(x) = 1 - e^{-cx}. \qquad (4.43)$$

However, further suppose he had difficulty in specifying certainty equivalents for lotteries, and thus our confidence in the value of parameter c might not be too great. His certainty equivalents for different lotteries may have led to quite different values of c between one-third and one, for example.

In such a case, the appropriateness of a sensitivity analysis is clear. First, we would evaluate the expected utilities of each course of action as a function of parameter c. If there were three possible courses of action, a plot of this might be shown as in Figure 4.24. With such information, alternative 3 would immediately be eliminated from further consideration since it is dominated

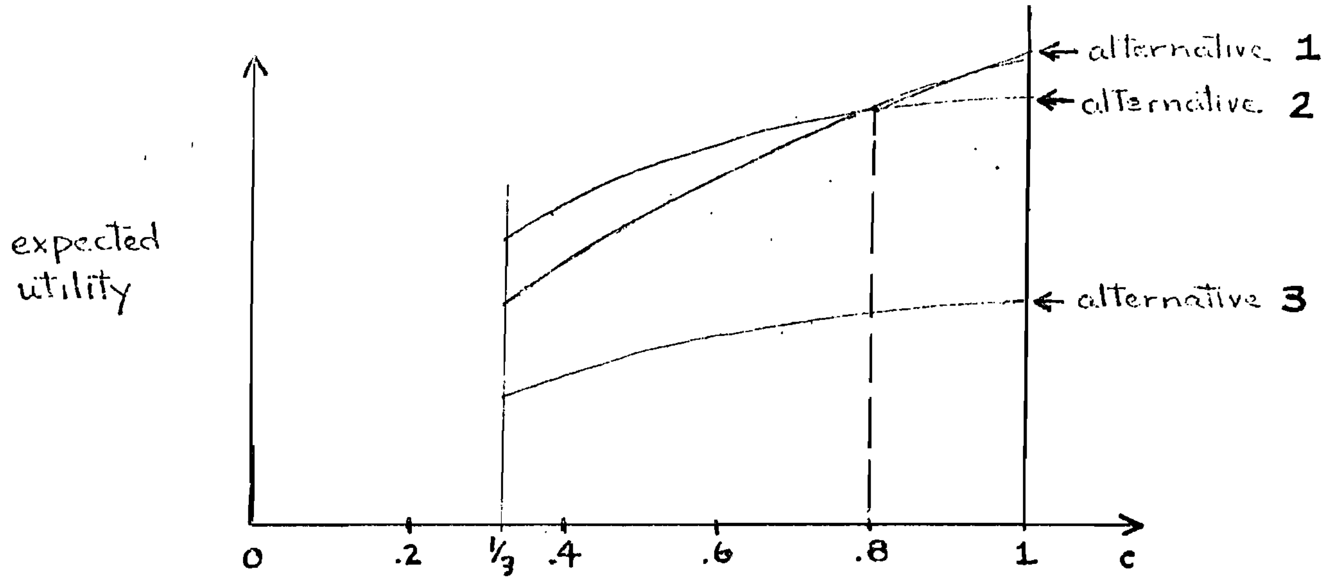


Fig. 4.24 Parametric Analysis Using $u(x) = -e^{-cx}$

by both alternatives 1 and 2. If $c < 0.8$, alternative 2 is best; otherwise alternative 1 should be chosen. Now rather than specify the exact value of c to solve the decision problem, we only need to determine whether c is larger or smaller than 0.8. This should be an easier assessment task than our former one.

4.10 ILLUSTRATIONS OF THE ASSESSMENT OF UNIDIMENSIONAL UTILITY FUNCTIONS

The purpose of this section is to illustrate by example the assessment of some unidimensional utility functions. It is by no means meant to be a catalog of the work on this problem. In fact, to illustrate the state of the art, so to speak, we emphasize more recent work at the expense of some earlier investigations which helped pave the way to our present status. However, let us briefly mention two of these initial efforts.

One of the pioneering attempts to measure utility functions was that of Mosteller and Nogee [1951]. In a laboratory setting, an individual was offered his choice between accepting a monetary lottery $\langle h,p,k \rangle$, a lottery yielding h cents with probability p or costing k cents with probability 1-p. During the course of the experiment, the same lottery was offered several times. From this the proportion of times the lottery was accepted was calculated. By using this procedure and varying h while holding p and k constant, the amount of money h_0 where the acceptance proportion was one-half determined. Then zero (not accepting the lottery) was taken as the certainty equivalent for $\langle h_0, p, k \rangle$, so

$$u(0) = pu(h_0) + (1-p)u(k) \qquad (4.44)$$

where u is the subject's utility function. The experimenters arbitrarily set $u(0) = 0$ and $u(k) = -1$ and used (4.44) to calculate the realtive preference of h_0. By repeating the above procedure for seven different values of p, the utilities of seven experimental points were specified from which the subject's utility function was graphed.

Another important contribution to the measurement of utility functions was that of Davidson, Suppes, and Siegel [1957] who attempted to improve upon the work just described. One of their major criticisms of Mosteller and Nogee's experiment was that almost every choice offered to the subjected involved choosing between accepting or rejecting a lottery. Thus, one alternative had uncertainty and participation in the experiment associated with it, while the other alternative involved no uncertainty or participation. If a subject were biased either toward or against gambling or participation, this procedure could have led to distorted results. A second criticism concerned the fact that Mosteller and Nogee used objective probabilities as if they were the subjective probabilities perceived by the subjects. To deal with these problems, Davidson, Suppes, and Siegel offered their subjects choices between lotteries, which hopefully canceled out distortion due to preferences for gambling and participation, and experimentally determined the subject's subjective probabilities.

The stated purpose of both of these experiments was to test the appropriateness of the expected utility decision model with regard to small sums of money. In both cases, their results established that utility functions could be measured in laboratory settings, at least for small sums of money. They also pointed out some of the "do's and dont's" in assessing utility functions. An important remaining problem was demonstrating that meaningful utility functions could be assessed for decision makers faced with real-world decision problems.

4.10.1 Preferences of Oil Wildcatters

One of the first major attempts to assess utility functions in an operational situation was that of Grayson [1960]. He spent a considerable amount of time quantifying the preferences for money of a number of oil wildcatters engaged in exploratory search for gas and oil. His approach was as follows. A hypothetical drilling venture was offered to a wildcatter, and he was asked to accept or reject this on the basis of the investment required, potential payoff, and probability of success.

For instance, the operator would be asked whether he would invest $ 20,000 in a venture which had a potential gross payoff of $ 100,000 if it were successful and had a 0.4 probability of success. If the answer was yes, the probability of success was lowered until the operator was indifferent between accepting and rejecting the venture. If the investment was originally rejected, the probability of success was raised to the indifference probability.

If this indifference probability is p, then

$$u(0) = pu(\$80,000) + (1-p)\, u(-\$20,000). \quad (4.45)$$

By arbirarily setting two points on the utility function u, a third was empirically evaluated using (4.45). This procedure was repeated for a large number of ventures, thus providing many points on the wildcatter's preference curve. Finally a "best fit" curve (determined visually) was drawn through these points.

Before presenting a specific example of Grayson's work, two comments on his work are in order. First of all, no attempt was made to exploit the general characteristics of utility functions, such as risk aversion, in assessing the wildcatter's preferences. Of course, seminal research in this area did not appear until after Grayson's work. Secondly, as pointed out by Grayson, inconsistences in an operator's preferences were not brought to his attention for possible modification except in one case. For this operator, William Beard of Beard Oil Company, these inconsistencies were reduced to a nominal level.

Mr. Beard's utility function for money on October 23,1957 is illustrated in Fig. 4.25. The points marked by an "o" on the figure are those which were empirically determined by

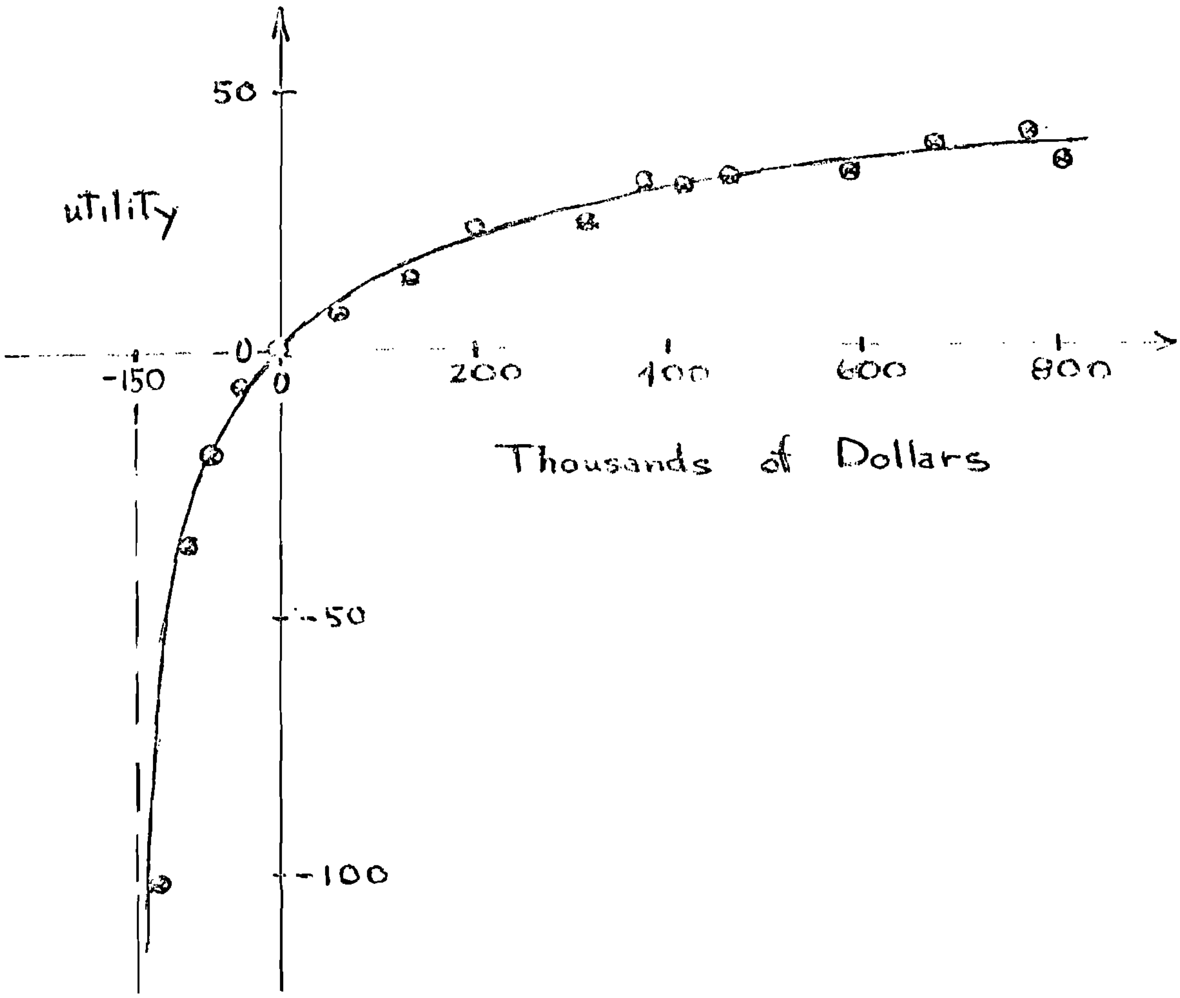


Fig. 4.25 William Beard's Utility Function for Change in Monetary Assets

Grayson. Kaufman [1963] later found that an analytical function which is an "astoundingly good" fit to this empirical data is the logarithmic utility function

$$u(x) = -263.31 + 22.093 \log (x + 150{,}000), \; x > -150{,}000 \qquad (4.64)$$

where x reprsents the change in Mr. Beard's asset position in dollars.

It is evident from Fig. 4.25 that u is monotonically increasing and risk averse. Also, by calculating the risk aversion

$$r(x) = -\frac{u''(x)}{u'(x)} = \frac{1}{x + 150{,}000}, \qquad (4.47)$$

it is clear that r decreases as x increases, so u is decreasingly risk averse. If it had been possible to determine beforehand that Mr. Beard subscribed to these characteristics, the number of empirical evaluations required to accurately assess his utility function would have been considerably smaller.

4.10.2 Preferences of Business Executives

Another large effort to assess utility functions was conducted by Swalm [1966]. He interviewed approximately one hundred people from various corporations in an attempt to evaluate experimentally their corporate utility function for money. That is, he was interested in the utility function they used to make corporation decisions as differentiated from personal decisions. The intent of this work was to describe, not prescribe, how these people made corporate decisions.

The first step in each interview was to familiarize the decision maker with the concept of utility theory. Then his 'planning horizon', defined as twice the maximum amount he might recommend be spent in any one year, was determined. The utility functions were assessed for consequences up to the planning horizon, as it was felt amounts greater than this would not be meaningful to the decision maker.

The type of questioning used to empirically evaluate points on the utility curve involved choices between simple 50-50 lotteries with two consequences and another consequence for certain. The sure-thing consequence was then adjusted in succeeding questions until the decision maker was indifferent between it and the lottery (i.e., the certainty equivalent for the lottery was found). By arbitrarily setting the utility for the consequences of this lottery, the utility assigned to the certainty equivalent, which had to be equal to expected utility of the lottery to be consistent, was found easily. This provided an empirical point on the utility function. Now, the certainty equivalent could be used in new lotteries to fix the utilities of other consequences. A number of points on the utility function involving both gains and losses were specified in this manner. Finally a smooth curve was fitted to the data.

Throughout the questioning, the alternatives available to the decision maker were made as realistic to him as possible.

As an example quoted from Swalm:

"Suppose your company is being sued for patent infringement. Your lawyer's best judgment is that your chances of winning the suit are 50-50; if you win, you will lose nothing, but if you lose, it will cost the company $1,000,000. Your opponent has offered to settle out of court for $200,000. Would you fight or settle?"

Two of Swalm's conclusions were particularly interesting. First, he found businessmen did not attempt to maximize expected dollar income in situations involving risk, and that cardinal utility was "at least a step in the right direction." Secondly, most junior executives made company decisions in a manner that put their own interests before the company's interests. From our point of view, that is, from one mainly interested in the normative implications of utility theory, perhaps the most important result was that many people's utility functions were assessed over consequences which had operational significance to the individual decision makers involved.

Spetzler [1968] has quantified the preferences of a number of business executives from one company in an attempt to evaluate a corporate utility function . The objective was to develop a corporate risk policy for capital investment decisions. A major part of this work concerned assessing utility functions for thirty-six managers of this firm, including all the top executives. The initial interview with each individual was to acquaint him with the concept and need for quantifying his preferences and to determine which risk characteristics represented his attitude. To accomplish this, the decision makers were given an investment opportunity yielding a present value of x_s net dollars if successful and x_f net dollars if it failed. The probability of success p was also given, and the decision maker chose whether or not to accept such an investment. The probability p was then varied to find the indifference probability p_o where the decision maker was indifferent between accepting and rejecting the project. For this value p_o,

$$u(o) = p_o\, u(x_s) + (1-p_o)u(x_f),$$

which gives one the relative preferences of three consequences. By repeating this procedure for twenty different investment opportunities at each of two company investment levels, three and fifty million dollars per investment, a number of points on a utility function were empirically determined for each decision maker.

From the questioning, it was found that each of the decison makers was risk averse. It was assumed they should also be decreasingly risk averse. Then, using a least square error approach, decreasingly risk averse utility functions of the form

$$u(x) = a + b \log (x + c), \ b > 0, \qquad (4.48)$$

were fitted to the empirical utility points. Using the resulting "best-fit" utility function, adjusted indifference probabilities were calculated for each investments, and these in turn were discussed with the respective decision makers. Many of the individuals felt these adjustments more adequately expressed their preferences than their original responses. However, some did not, so a more flexible utility function

$$u(x) = a + b \log (x + c - d|x|), \ b > 0, \ 0 \leq d < 1, \qquad (4.49)$$

was tried. This function still satisfied all the original risk characteristics except for a break at the origin. By repeating the procedure just described using (4.49) to calculate adjusted indifference probabilities it was found a few decision makers were still not satisfied. Thus, to partially smooth this break at the origin, another parameter was added while maintaining the decreasing risk aversion property. The newly revised utility function was

$$u(x) = a + b \log \{ x + c - d \, [(x^2 + f^2)^{1/2} - f] \}, \qquad (4.50)$$

where $b > 0$, $0 \leq d < 1$, $f > 0$ and $\{x + c - d \, [(x^2 + f^2)^{1/2} - f]\} > 0$ for all possible amounts of x. The adjusted indifference probabilities calculated using the "best-fit" utility function of form (4.50) were not only acceptable to each decision maker but were preferred to his original probabilities in all cases. For certain values of parameters c,d, and f, one can prove u is decreasingly risk averse, but for some individuals the best-fit utility function violated this condition.

An important result of Spetzler's work was that by using both qualitative risk characteristics and quantitative assessments, he developed utility functions which adequately expressed the preferences of a number of decision makers faced with real-world investments problems. The value of consistency checks, which in this case involved the repeated interviewing of the decision makers concerning the adjusted indifference probabilities, is particularly evident from this work.

4.10.3 Computer Programs to Assess Preferences for Money

Quite a different approach to assess utility functions has been in use at the Harvard Business School since 1966. A number of computer programs (see Schlaifer [1971]) are used to assess utility functions of different forms which are consistent with various input data specifying both qualitative and quantitative characteristics of the utility function. Here, to illustrate the idea, we will briefly discuss the first program which computes a decreasingly risk averse utility function of the form

$$u(x) = -e^{-ax} - be^{-cx}, \quad a > 0, \ bc > 0, \tag{4.51}$$

consistent with a decision maker's certainty equivalents for three 50-50 lotteries. If no function exists which is consistent, this is indicated by the program. By presenting the decision maker with three 50-50 lotteries where the consequences have equal spread, it is easy to check the appropriateness of the decreasingly risk averse assumption.

As an example, suppose we were assessing a decision maker's utility function for change in monetary asset position between - \$1000 and \$3000. We might begin by asking his certainty equivalents for lotteries <\$0, -\$1000> <\$1000, \$0>, and <\$2500, \$1500>. If his certainty equivalent for the second lottery was greater than \$500, we would know he was not risk averse, but risk prone for this region at least, and therefore, not decreasingly risk averse. Another decision maker faced with the same three lotteries might give his certainty equivalents as -\$550, \$400, and \$1850, respectively. Clearly this decision maker is risk averse since his risk premiums, the expected monetary values minus the certainty equivalents, are positive. However, he is increasingly risk averse since the risk premiums increase as the potential payoffs increase. For both of these cases, a utility function of form (4.51) would not be appropriate.

Suppose a third decision maker stated his certainty equivalents were -\$650, \$400, and \$1950, respectively. This decision maker is decreasingly risk averse. By using (4.51) and equating the utilities of the certainty equivalents to the expected utilities of the respective lotteries, we get three equations with three unknowns, a,b, and c. The computer solves for these unknowns and outputs the resulting utility function. Even when the three certainty equivalents are consistent with a decreasingly risk averse utility function, there may not be a utility function of form (4.51) which both fits these data and is decreasingly risk averse for all amounts of x. For instance, if both $b < 0$ and $c < 0$, the resulting utility function becomes risk prone for x greater than some amount. If the decision maker's operational range of x includes part of the risk-prone range, one must either try a different functional form for the utility function or repeat this procedure with a different set of input lotteries.

The research program on the assessment of utilities at the Harvard Business School has the following pragmatic orientation: The researchers assume that a time-sharing, intersective computer terminal can be used during the interrogation procedure. The respondent is first asked a series of qualitative questions which establish the qualitative structure of his utility curve. Next one or two hypothetical numerical questions are posed and the respondent can give either explicit numerical values or ranges of values. At any stage in the protocol the computer program checks for internal consistency of the past responses and for any hypothetical lottery the program is prepared to compute the possible range of certainty equivalents for that lottery that is consistent with the input data. In practice then, one can often resolve one's actual choice problem without fully defining a single utility function. With some familiarity with the programs the respondent can run his own sensitivity tests and, in so doing, build up a sense of confidence in the procedure. And in those cases which the sensitivity analysis undermines one's senses of security it is better that this be overt rather than not realized.

As an intersting sideline, utility functions for money are assessed for M.B.A. students at Harvard Business School using this program. In approximately 70 per cent of more than a thousand assessments, a decreasingly risk averse utility function of the form (4.51) has been found to be satisfactory for the decision maker.

4.10.4 Preferences in a Hospital Blood Bank

A final example of an empirically evaluated utility function in a context quite different from the previous examples concerns the operation of a hospital blood bank. One of the important measures of effectiveness for evaluating hospital blood bank inventory policies is blood shortage. Here, shortage is defined as blood requested by a doctor which could not be assigned from the hospital inventory. As part of a larger effort, which is discussed in detail in Section 5.10, a utility function was assessed for percent of blood shortage in a year, that is, the percent of all blood requested by doctors which could not be assigned from hospital inventory at a particular hospital. In this shortage situation, a special order for the particular type of blood is placed with a central blood bank, professional donors may have to be called, an operation may be postponed, etc., but only in extremely rare circumstances would a death result from shortage as we have defined it.

The person whose preferences were assessed was the nurse in charge of ordering blood at The Cambridge Hospital in Cambridge, Massachusetts.

First it was established that, in this hospital, shortage would never exceed ten percent of the units demanded. The problem was then to evaluate a utility function for shortages between zero and ten percent.

Clearly, preferences decreased as percent shortage increased so the utility function had to be monotonically decreasing. Using the converging technique discussed in subsection 4.9.3, the certainty equivalent for the 50-50 lottery <0,10> yielding zero or ten percent shortage was found to be 6.5 percent shortage. Since preferences were decreasing and the certainty equivalent was greater than the expected consequence, there was reason to believe the decision maker was risk averse. Next, the certainty equivalents for the lotteries <0, 6.5> and <6.5, 10> were found to be 4 and 8.5, respectively. With these responses, it was justifiable to assume the decision maker was risk averse.

If the utility for x percent shortage is represented by $u(x)$, from the certainty equivalents, we know

$$u(6.5) = [u(0) + u(10)]/2, \tag{4.52}$$

$$u(4.0) = [u(0) + u(6.5)]/2, \tag{4.53}$$

and

$$u(8.5) = [u(6.5) + u(10)]/2 \tag{4.54}$$

For simplicity, a constantly risk averse utility function of the form
$b(1-e^{cx})$ was fit to the data using (4.52) after the origin and unit of measure were respectively set by

$$u(0) = 0 \tag{4.55}$$

and

$$u(10) = -1. \tag{4.56}$$

As can be seen from Fig. 4.26, the utility function

$$u(x) = \frac{1}{2.67} (1-e^{0.13x}) \tag{4.57}$$

fit the empirical data quite closely.

A consistency check was used to see if indeed the decision maker was risk averse. She was asked whether she preferred $\langle i+1, i-1 \rangle$ or i percent shortage for i = 1,2,..., 9. In all cases, the sure i percent shortage was preferred. This verified that the decision maker was risk averse.

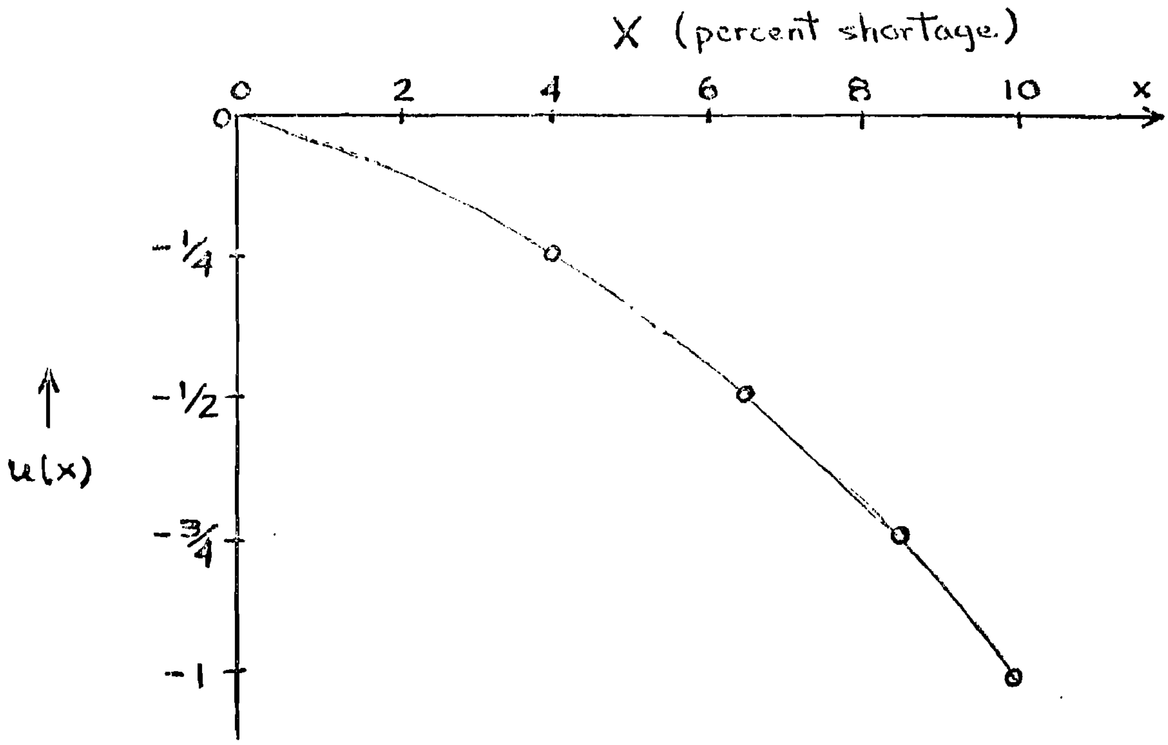


Fig. 4.26 A Utility Function for Blood Shortage

4.10.5 Summary

Actual assessments of unidemenional functions can be categorized into three groups. The first attempts to evaluate utility functions were made in laboratory settings. These experiments indicated that preferences could be quantified and provided some experience with the assessment procedures. Building on this work, utility functions for decision makers faced with operational decisions were next determined by fitting a curve to a number of empirically evaluated utility points. Since the appearance of Pratt's paper on risk aversion in 1964, qualitative characteristics of utility functions have been exploited to complement the quantitative certainty equivalent information. This has led to both a simplification of procedures for assessing utility functions and resulted in utility functions which more accurately express the decision maker's preferences. In Chapter 7 and 8, additional examples illustrating the assessment of unidimensional utility functions and their relevance to multiattribute problems are given.

4.11 EXPLICATING A SINGLE ATTRIBUTE BY MEANS OF MULTIPLE ATTRIBUTES.

In later chapters of this book we shall consider ways of coping with preferences and utilities for consequences that can only be described with multiple attributes. We shall investigate techniques that reduce multidimensional problems down to unidimensional problems, thus enabling us to apply the technique of this chapter. But, as we shall show in this section, there are examples where the reverse procedure may need to be employed. It is sometimes constructive to explode what is seemingly a unidimensional problem into a multidimensional one. Let us explain this by an example. Norman Toy [1971] in his doctoral thesis investigated how individuals, such as we authors or other academics, should manage their retirement pension funds. Take the example of a professor whose sole source of income after retirment will come from social security payments and the retirement funds of his university. He typically has a range of options each year: he can choose to invest the funds set aside for his later retirement years in fixed interest bonds (or comparatively fixed), or else to invest a portion (within bounds) of these funds in equities whose future values depend on the vicissitudes of the stock market. His choice can appreciably affect his post-retirement life style. Not only does the professor have to worry about the uncertainties of the stock market but also about inflation rates, about the longevity of his spouse, and so on. One natural way to approach the problem is to assess a utility function for total wealth at retirement. Toy asked his subjects such questions as: Would you rather have a total retirement fund of \$150,000 for certain or a 50-50 chance at \$100,000 or \$250,000?

This question, if taken really seriously, is terribly difficult to answer. It depends on so many things: What is the inflation rate? ...Well, that's not conceptually so difficult: one can normalize all amounts to today's price index. How certain can one be that one's spouse will be alive to share those retirement years? ... Well, that complication can be handled, as is done in Section 4.12, by assessing utility function for total wealth at retirement conditional on the spouse surviving and on not surviving. But still the problem is not easy to think about--even if one conditions the outcome by the status of one's family obligations. One is forced to think hard about the implications of different monetary amounts in one's post-retirement standard of living. Wealth in itself can be thought of as a surrogate for consumption streams that can be purchased with that level of wealth. It is complicated further by the fact that without the availability of inflation-free annuities, one cannot be certain of which consumption stream one will enjoy (or perhaps not enjoy) starting from a given wealth position.

Toy grappled with this problem in several manners. In one informal approach he had his subjects simulate choices to be made in their post-retirement years. The simulation exercise took place at a time-sharing computer terminal. Take the case of the professor who retires with a retirement fund of $150,000 when he is 67 years and his wife is 66. He must decide in year 1 (first year after retirement) how much to consume, how much to put into stocks, and how much into bounds. Toy's interactive computer model had a built-in simulator of inflation rates, of the equity and bond market, and of longevity rates based on actuarial tables for the male and female partners. The subject is asked to decide year by year what he wants to do (how much he wishes to consume and to invest) and then the computer obligingly handles all the accounting in the probabilistically simulated world. Sooner or later one of the partners dies and the spouse carries on. Since the environment is uncertain, it is important to experience many runs with the same initial conditions before one generates an appreciation of what it means to be left with a retirement fund of $150,000. Since the year-by-year process is slow to simulate, Toy allowed his subjects to choose various strategies over time which obviated the need to make these time consuming simulated year-by-year decisions. By means of this simulated experience, Toy's subjects become better prepared to respond more responsibly to hypothetical questions about wealth at retirement.

In a more formal approach to this problem Toy investigates his subject's utility preferences over consumption streams, a process which involves multidimensional assessments, and he then deduced by this means a <u>derived</u> utility function over the <u>surrogate</u> <u>unidimensional</u> <u>variable</u>:wealth at retirement. Scott F. Richard [1972] in his dissertation addressed the same problem in a more systematic, rigorous, analytical fahsion. Richard's work is based on the path-breaking results of Professor Richard Meyer, which concern utility assessments of consumption streams over time and are discussed in Chapter 9.

We close the subsection by reiterating the point of this discussion: In certain contexts there may be a consequence that can be described quite naturally by a unidimensional attribute but it may not be natural to assess a utility function directly over this attribute. Instead one might have to seemingly complicate the analysis by introducing multiple attributes, over which it may be more natural to assess preferences.

4.12 Conditional Unidimensional Utility Theory

This section is meant to illustrate directly the relevance of unidimensional utility theory to multiattribute utility problems, and as such, to begin a transition to the next chapters.

4.12.1 State Dependent Preferences

As in previous sections, let us assume that the decision maker's choice of an act a determines the probability distribution of an uncertain payoff $\tilde{x}$. But now let us assume that in reacting to simple lotteries with various x payoffs the decision maker is concerned about which state of the world, $w_1,\ldots,w_j,\ldots,w_r$ will prevail. To take a simple example, if x represents the decision maker's wealth at retirement twenty years hence, his certainty equivalent for a 50-50 gamble between x_1 and x_2, say, might depend on the status of the health of his wife and of himself. He can, of course, answer the question posed keeping informally in mind the possible states of health and their probabilities, but instead of answering the question in an <u>un</u>conditional or marginal sense, he may feel more comfortable thinking about the question <u>conditionally</u> on each state, and then somehow combining these conditional evaluations to get an unconditional evaluation.

We simplify by assuming that the choice of act $\underline{a}$ affects the probability distribution of $\tilde{x}$ but not of $\tilde{w}$. Let

$$P(\tilde{w} = w_j) = p_j \text{ for } j = 1,\ldots,r. \tag{4.58}$$

We assume, however, that the decision maker's utility function u depends on both x and w. He wishes to choose the act $\underline{a}$ to

$$\max_{a\varepsilon A} E_a\, u(\tilde{x},\tilde{w}), \tag{4.59}$$

where the expectation operator E_a depends on $\underline{a}$ since the probability distribution of $\tilde{x}$ (not of $\tilde{w}$) depends on $\underline{a}$. How can the decision maker think systematically about constructing his two-dimensional u(.,.) function? That's the issue. We hope to demonstrate the usefulness of uni-dimensional utility theory to this question.

Let's examine our problem in terms of the decision tree in Figure 4.27. At move 1, the decision maker chooses an act $\underline{a}$ from A; at move 2, Chance chooses x from a distribution that depends on $\underline{a}$; at move 3, Chance chooses w_i with probability p_i (for $i = 1,\ldots,r$) independently of the choices at moves 1 and 2. The consequences resulting from the path (a,x,w_i) has a utility $u(x,w_i)$.

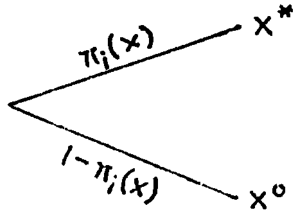
We define the <u>unconditional</u> utility of x to be

$$\bar{u}(x) = \sum_{i=1}^{r} u(x,w_i)p_i \tag{4.60}$$

and for the purpose of making a decision at move 1, the unidimensional <u>un</u>conditional utility function $\bar{u}(\cdot)$ is all that is necessary to know. If the decision maker can directly assess $\bar{u}$, fine; but he still might prefer to get at $\bar{u}$ indirectly through a set of conditional assessments.

4.12.2 <u>Conditional Assessments</u>

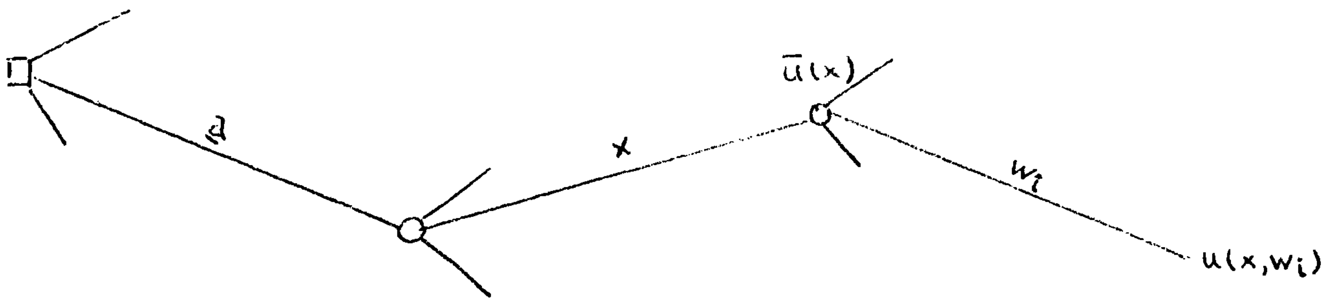
Assume that we are concerned with a range of x values that fall in the interval* from x^o to x^*. If the decision maker knew that w_i were to prevail then let him be indifferent between obtaining x for certain and obtaining the lottery which yields x^* with probability $\pi_i(x)$ and x^o with probability $1 - \pi_i(x)$. Schematically,

$x \sim$ 

, <u>given</u> w_i . (4.61)

---

*This assumption can easily be relaxed but is made for convenience.



Move 1	Move 2	Move 3
Act	Distribution of $\tilde{x}$ depends on a.	Distribution of w does not depend on x or a.

Figure 4.27. A Decision Tree Illustrating State Dependent Preferences

In other words $\pi_i(\cdot)$ is the decision maker's <u>conditional</u> utility function for x-values given the state w_i, normalized by the requirements $\pi_i(x^o) = 0$ and $\pi_1(x^*) = 1$. Clearly π_i is a unidimensional utility function.

In principle, at least, we can think of the utility function in two attributes $u(\cdot,\cdot)$ and it must be such that for any i there are constants c_i and $b_i > 0$ where

$$u(x,w_i) = c_i + b_i\pi_i(x), \text{ for all } x, \qquad (4.62)$$

and for $i = 1,\ldots,r$. Hence in order to assess $u(\cdot,\cdot)$ it is not enough to assess the r conditional utility functions $\pi_1(\cdot),\ldots,\pi_r(\cdot)$--we must somehow also assess the scaling constants $c_1,b_1,c_2,b_2,\ldots,c_r,b_r$. That's our next concern.

From (4.62) and (4.60) we observe that

$$\bar{u}(\mathbf{x}) = \sum_{i=1}^{r} [c_i + b_i\pi_i(x)]p_i$$

$$= \sum_{i=1}^{r} c_ip_i + \sum_{i=1}^{r} b_i\,\pi_i(x)p_i. \qquad (4.63)$$

But for decision purposes we can ignore the constant term on the right-hand side of (4.63) and thus we see that we do <u>not</u> have to determine the c_i's. This is a tremendous help because otherwise we would have to ask such disconcerting questions as: "If you were at postion (x,w_i), how much, in terms of attribute X, would you be willing to give up in assets to modify w_i to w_j?" And fortunately we can avoid such questions.

4.12.3 Conditional Certainty Equivalents

For any act a let the resulting payoff be denoted by the uncertain quantity $\tilde{x}^{(a)}$. The conditional certainty equivalent for $\tilde{x}^{(a)}$ given w_i, denoted by $\hat{x}_i^{(a)}$, satisfies the relation

$$\pi_i(\hat{x}_i^{(a)}) = E_a\ \pi_i(x^{(a)}). \qquad (4.64)$$

Hence any act a can be evaluated by the r-tuple of conditional certainty equivalents $[\hat{x}_1^{(a)},\dots,\hat{x}_r^{(a)}]$. In practice if one has only a few acts to choose amongst one might wish to directly assess $\hat{x}_i^{(a)}$ for all i and a without formalizing the conditional utility functions π_i for $i = 1,\dots,r$. But now the problem boils down to tradeoffs or substitution rates amongst the r component, conditional, certainty-equivalent values.

Let us now consider the lottery, which will yield a certain amount x_i if w_i prevails, for $i = 1,\dots,r$; illustrated in Fig. 4.28. Let's characterise this lottery by the symbol $\langle x_1,\dots,x_i,\dots,x_r\rangle$, and our task is to structure the decision maker's preferences in this evaluation space. If we let

$$\langle \underline{x}' \rangle \equiv \langle x_1',\dots,x_r' \rangle \quad \text{and} \quad \langle \underline{x}'' \rangle \equiv \langle x_1'',\dots, x_r'' \rangle ,$$

then by (4.63) we see that

$$\langle \underline{x}' \rangle \succsim \langle \underline{x}'' \rangle \Longleftrightarrow \sum_{i=1}^{r} b_i p_i \pi_i(x_i') \geq \sum_{i=1}^{r} b_i p_i \pi_i(x_i'') \quad . \qquad (4.65)$$

Recall, however, that we still have to develop a method for

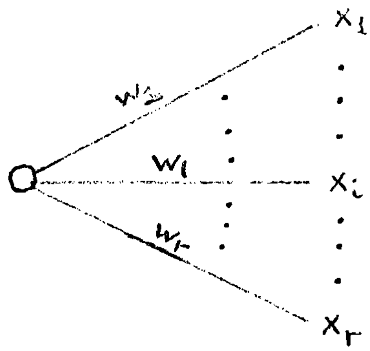


Figure 4.28. A Lottery Where the Consequence Depends on The State w

determining appropriate b_i values.

Let us compare the following two lotteries.

Lottery L': The return is x^o for each state w_1 to w_r.

Lottery L": The return is x^o for each state w_1 to w_r except for states w_i and w_j; the return for w_i is $x^o + \alpha_i$ for w_j is $x^o - \beta_j$.

Now suppose the decision maker adjusts α_i and β_j so that L" is indifferent to L'. Then from (4.65) we have

$$b_i p_i \pi_i(x^o) + b_j p_j \pi_j(x^o) = b_i p_i \pi_i(x^o + \alpha_i) + b_j p_j \pi_j(x^o - \beta_j). \quad (4.66)$$

Since in (4.66) the α_i and β_j values are known, it is a simple matter to solve for the ratio

$$b_i p_i / b_j p_j.$$

If, for example, we repeatedly use this pairwise indifference procedure by letting i = 1 and j = 2,...,r successively then we can determine the ratios

$$b_1 p_1 / b_j p_j \quad \text{for } j = 2,\ldots,r. \quad (4.67)$$

Now, since u in (4.62) can be arbitrarily scaled, there is no loss of generality in letting $b_1 p_1 = 1$. Using this and (4.67), we can determine the appropriate scaling constants $b_1,\ldots,b_r$. Observe also that if one wishes to do so, one can always suppress the formal determination of the p_i's. But, of course, the tradeoff question between the lotteries in Figure 4.27 does implicitly require the decision maker to weigh in his mind the chances of w_i and w_j.

4.13 WHERE WE STAND

Many of the important aspects of utility theory have been introduced in this chapter. The theory necessary to make the concept of utility operationally useful has been discussed in detail, methods for assessing unidimensional utility functions have been described, and examples where utility functions have been assessed in operational situations illustrated. The conditional unidimenional utility theory introduced in the proceeding sections begins to bridge the gap between unidimensional and multiattribute utility theory. Only with a firm understanding of the fundamentals in this chapter do we begin to tackle the main problem of concern in the next two chapters, the structure and assessment of multiattribute utility functions.

DECISION ANALYSIS WITH MULTIPLE CONFLICTING OBJECTIVES PREFERENCES AND VALUE TRADEOFFS (Chapters 5 & 6)

Ralph L. Keeney and Howard Raiffa

May 1975 WP-75-53

CHAPTER 5

MULTIATTRIBUTE PREFERENCES UNDER UNCERTAINTY: THE TWO-ATTRIBUTE CASE

In this and the following chapter, the ideas developed and results presented are useful for assessing multiattribute utility functions. The results are mainly representation theorems specifying the functional form of the utility function provided certain assumptions concerning the decision maker's preferences are appropriate. We shall develop reasonable preference assumptions, determine when such assumptions are appropriate, and assess the resulting utility functions.

Many of the concepts of importance in multiattribute utility theory can be illustrated with the two-attribute problem. Therefore, to avoid unnecessary complications and detail, we have chosen to focus on this case in Chapter 5. Assessments involving three or more attributes are addressed in Chapter 6. However, the material in this first section is relevant to both situations.

5.1. The Basic Problem

We will assume that an objectives hierarchy has been specified and that attributes $X_1, X_2, \ldots, X_n$ have been identified and are appropriate for the problem. If x_i designates a specific level of X_i, then our task is to assess a utility function* $u(\underline{x}) = u(x_1, x_2, \ldots, x_n)$ over the n attributes.

The utility function u has the salient characterizing property that given two probability distributions A and B over multiattribute consequences $\underline{\tilde{x}}$ then: <u>Probability distribution</u> A <u>is at least as desirable as</u> B <u>if and only if</u>

$$E_A[u(\underline{\tilde{x}})] \geq E_B[u(\underline{\tilde{x}})], \tag{5.1}$$

where E_A and E_B are the usual expectation operators taken with respect to distribution measures A and B respectively**. This merely asserts that expected utility is the appropriate criterion to use in choosing among alternatives.

As a special degenerate case of (5.1) we conclude that: <u>Alternative</u> $\underline{x}^A$ <u>is at least as desirable as</u> $\underline{x}^B$ <u>if and only if</u>

$$u(\underline{x}^A) \geq u(\underline{x}^B). \tag{5.2}$$

---

*To be consistent with our past usage we should refer to the utility <u>function</u> as u or $u(\cdot)$ and not $u(\underline{x})$, which is strictly speaking the value of u at $\underline{x}$. But we believe our occasional sloppiness in notational usage will simplify the presentation a little and will not cause any real confusion --perhaps a bit of aesthetic displeasure.

**If probability distribution A is defined in terms of a joint probability density function $f_A(\cdot)$ in R_n, Euclidean n-space, then

$$E_A[u(\underline{\tilde{x}})] \equiv \int_{R_n} u(\underline{x})\, f_A(\underline{x})\, d\underline{x}.$$

In our presentation, we will differentiate between cases when one already has determined a value function over the attributes and when one has not. The value function can be exploited in determining the utility function.

5.1.1. Assessing a Utility Function Over the Attribute "Value"

Recall from Chapter 3 that a value function $v(\underline{x}) = v(x_1, x_2, \ldots, x_n)$ over n attributes assigns a ranking to all possible consequences. It is a function which satisfies the special case (5.2) required of a utility function. And so, by definition, a utility function is a value function, but a value function is not necessarily a utility function*.

---

*Unfortunately, there is no standarized terminology for what we have chosen to call value functions and utility functions. In the literature, our value functions are sometimes referred to as worth functions, ordinal utility functions, preference functions, Marshallian utility functions and even utility functions. Similarly, out utility functions are referred to as preference functions, cardinal utility functions, von Neumann utility functions, probabilistic utility functions, and utility functions. Although clearly we can't be consistent with all the existing literature, we will try to be internally consistent with our own use of value functions and utility functions as we have defined them.

Chapter 3 indicated several methods one might use to acquire the value function $v(\underline{x})$. Because this function assigns a scalar "value" to each consequence $\underline{x}$, one can consider V as the scalar attribute "value" which takes on levels designated by v. Furthermore, since $v(\underline{x}^A) > v(\underline{x}^B)$ if and only if the decision maker finds $\underline{x}^A$ preferable to $\underline{x}^B$, the utility function over V must be monotonically increasing. Hence any of the ideas discussed in Chapter 4 for assessing unidimensional utility functions are appropriate for assessing $u[v(\underline{x})]$.

Operationally, the problem is not quite the same, however, since different levels of V per se do not have a physical interpretation to the decision maker. The techniques of Chapter 4 are useful in assessing $u[v(\underline{x})]$ but usually one must fall back to the interpretations of the original attributes $X_1, X_2, \ldots, X_n$, in order to implement the assignment task. This idea can probably best be illustrated with a simple example.

Example 5.1. Consider Figure 5.1 and suppose a value function $v(x_1, x_2)$ has been specified over the attribute space $X = X_1 \times X_2$ for $x_i^o \leq x_i \leq x_i^*$, $i = 1,2$. And for convenience assume that v is continuous and increasing in both X_1 and X_2. Also, for any consequence (x_1', x_2'), let us assume that there is a consequence of either form (x_1, x_2^o) for

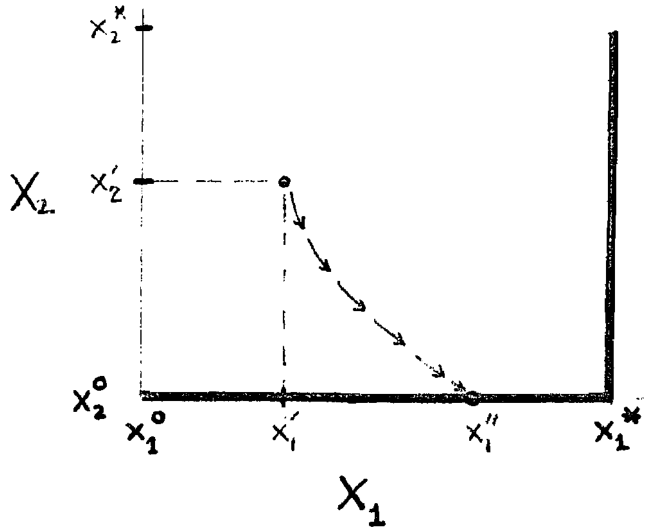


Figure 5.1. Assigning Utility to Consequences When a Value Function is Known

$x_1^o \le x_1 \le x_1^*$ or of the form (x_1^*, x_2) for $x_2^o \le x_2 \le x_2^*$ which is indifferent to (x_1', x_2'). The loci of all points of the form (x_1, x_2^o) or (x_1^*, x_2) are indicated in the figure by heavy lines. Thus if one had a utility function u defined for all points of the form (x_1, x_2^o) or (x_1^*, x_2), it would be easy to extend u to all points (x_1, x_2) in the domain of concern. If $v(x_1', x_2')$ equals $v(x_1'', x_2^o)$, then clearly $u(x_1', x_2')$ must be assigned to equal $u(x_1'', x_2^o)$, which is already known.

The problem then boils down to the assessment of u over the heavy lines in Fig.5.1, but this is a much easier task than assessing u over all X. Furthermore, the techniques of Chapter 4 can be directly applied to assess the two one-attribute (conditional) utility functions $u_1(x_1, x_2^o)$ as a function of x_1 and $u_2(x_1^*, x_2)$ as a function* of x_2. The only additional difficulty is that u_1 and u_2 must be consistently scaled to yield an appropriate u. Procedures for doing this are discussed in Section 5.8 later in this chapter. ∎

The generalization to more than two attributes is conceptually simple. One assesses a number of one-attribute (conditional) utility functions over the X_i attributes and

---

*Once again we apologize for our notational inconsistencies. We could talk about the functions $u_1(\cdot, x_2^o)$ and $u_2(x_1^*, \cdot)$ but at times it is more natural for us in this chapter to use the notation in the text.

consistently scales these utility functions to form one utility function u over a subspace of X. Then for each $\underline{x}^a$ to which u is not directly assigned, one finds an $\underline{x}^b$ with a u assigned, such that $v(\underline{x}^a) = v(\underline{x}^b)$, and then sets $u(\underline{x}^a) = u(\underline{x}^b)$.

5.1.2. Use of x^* and x^o

Now consider the case where the value function has not been specified over X. If there are only a few possible consequences $\underline{x}^1, \underline{x}^2, \ldots, \underline{x}^R$, it may be reasonable to assign a utility to each of these directly. One sets the utility of two of the consequences and evaluates the others in terms of the first two (or other consequences with utilities already assigned). For example, if we define $\underline{x}^o$ to be the least preferable of $\{\underline{x}^1, \underline{x}^2, \ldots, \underline{x}^R\}$ and $\underline{x}^*$ to be the most preferable of this set, then we can arbitrarily set

$$u(\underline{x}^o) = 0 \text{ and } u(\underline{x}^*) = 1 \quad . \tag{5.3}$$

For each $\underline{x}^r$, one empirically assesses a probability π_r such that $\underline{x}^r$ is indifferent to the lottery yielding either $\underline{x}^*$ with probability π_r or $\underline{x}^o$ with probability $(1-\pi_r)$. By equating expected utilities, it follows that

$$u(\underline{x}^r) = \pi_r, \text{all } r. \tag{5.4}$$

This approach is reasonable for perhaps up to fifty consequences, although with this size the procedure would be very tedious, and one would need many consistency checks to develop confidence in the assessments. Note that the basic idea is identical to that used in Chapter 4 for directly assessing utilities of consequences. The only difference is that the stimuli, the $\underline{x}^r$'s, are now vectors rather than scalars.

In situations where there are many possible $\underline{x}$ consequences contained in X, for which utilities are needed, the same approach could be used to assign utilities to a number of consequences in X. A curve-fitting procedure, interpolation, extrapolation, etc. could then be utilized to acquire utilities for all the other consequences. Especially when X represents a continuum in multiattribute space, such a procedure has two major practical shortcomings: (i) it fails to exploit the basic preference structure of the decision maker, and (ii) the requisite information is difficult to assess and the result is difficult to work with in expected utility calculations and sensitivity analysis. The ideas presented in the next subsection are motivated by these inadequacies.

5.1.3. The General Approach

The basic approach utilized in this and the next chapter is (i) to postulate various sets of assumptions about the basic preference attitudes of the decision maker, and (ii) to derive functional forms of the multiattribute utility function consistent with these assumptions. To use the results, one must first verify whether some of the assumptions hold in the particular problem at hand, and then one must assess a utility function consistent with the verified assumptions. The motivation for this procedure is that it addresses the shortcomings of the more direct approach suggested in the last subsection. The basic preference attitudes of the decision maker are exploited in specifying a utility function, and the actual assessment is simplified. Note that this is exactly the same approach used in Chapter 3 to assess value functions and Chapter 4 to assess unidimensional utility functions.

The assumptions investigated are felt to be operationally significant and relevant to many decision problems. Of fundamental importance in identifying simple representations of individual preferences is the verification and exploitation of certain independence properties which may exist among the decision maker's preferences for various amounts of different attributes. Ideally, we would like to obtain a representation of the utility function such that

$$u(x_1, x_2, \ldots, x_n) = f[f_1(x_1), f_2(x_2), \ldots, f_n(x_n)], \tag{5.5}$$

where f_i is a function of attribute X_i only, for $i = 1, 2, \ldots, n$, and where f has a simple form--an additive or multiplicative form, for example. When this is possible, the assessment of u can be greatly simplified. The fruitfulness of this approach, both in theoretical terms and in applications, is illustrated in this and the remaining chapters.

5.1.4. Outline of the Chapter

Chapter 5 develops two-attribute utility functions. First, the concepts of independence and their theoretical implications are presented. Then a procedure for assessing such utility functions is suggested. Finally, the detailed assessment of a utility function in a real-world setting is presented.

For notational convenience we shall denote the generic point in two-space as (y,z) rather than the more cumbersome (x_1,x_2). The utility function u(y,z), which is a two-attribute utility function when written in this form, may have more than two dimensions. For instance, if Y is a two-dimensional vector attribute and Z is a three-dimensional vector attribute, then u(y,z) can be interpreted as five-dimensional utility function. All the results of this chapter are appropriate for all two-attribute utility functions, regardless of the dimensionability of each of the arguments. However, often for convenience, we will treat Y and Z as unidimensional, scalar attributes.*

5.2. Utility Independence

One of the fundamental concepts of multiattribute utility theory is that of <u>utility independence</u>. Its role in multiattribute utility theory is similar to that of probabilistic independence in multivariate probability

---

*Throughout, we will use y and z, rather than the more conventional $\underline{y}$ and $\underline{z}$, to represent what may be either scalar or vector consequences.

theory. Here and in Chapter 6, much attention will be concentrated on utility independence and its implications for the following reasons:

1. Various utility independence conditions imply that the multiattribute utility function must be of a specified form. These forms include many possibilities for the final shape of the utility function including situations involving an interaction of preference among the attributes, and yet these independence assumptions simplify greatly the assessment of the original utility function.
2. The utility independence assumptions are appropriate in many realistic problems and they are operationally verifiable in practice.
3. Utility functions exploiting utility independence have been used in a number of important problems. Chapter 8 presents the details of one such problem concerning the development of the airport facilities of the Mexico City Metropolitan Area. Other problems in which utility independence has been used are covered in less detail in Chapters 7 and 9.

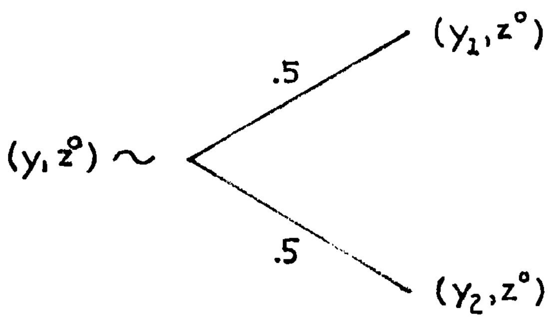
The concept of utility independence can also be viewed as a specialization of concept of preferential independence, which was exploited in Chapter 3.

5.2.1. Definition of Utility Independence

In this section we begin with a definition of utility independence in the two-attribute case. Let the attribute space X be partitioned into Y and Z such that $X = Y \times Z$ and denote a typical point in the attribute evaluation space by (y,z). Let us assume

$$y^{o} \leq y \leq y^{*} \quad \text{and} \quad z^{o} \leq z \leq z^{*}. \tag{5.6}$$

In analyzing a problem of this kind it is natural first to look at various unidimensional conditional utility functions. For example, we might investigate the conditional utility function for various y values given z^{o}; that is, the utility function along the heavy line in Fig. 5.2. We may then inquire if the decision maker's utility function shifts strategically if the given z-level changes from z^{o}. We are led to such questions as: "If z is held fixed throughout at z^{o}, what is your certainty equivalent for a 50-50 gamble yielding values y_1 and y_2 say?" Let us suppose the answer is $\hat{y}$, so that



Now we ask: "If z were held fixed at some other fixed value, say z', would your certainty equivalent $\hat{y}$ shift?"

In a surprisingly large number of contexts the certainty equivalent $\hat{y}$ does not shift. And this would be the case

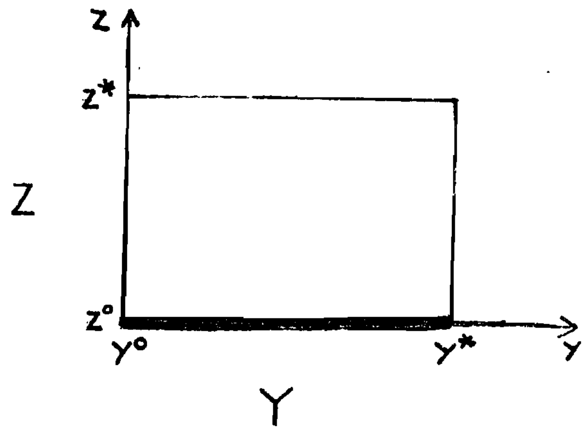


Figure 5.2. Preferences Over the Heavy Line May Be Interpreted as Conditional Preferences for Y Levels Given z^{O}

for any fixed y_1 and y_2. The certainty equivalent $\hat{y}$ would depend solely on the y_1 and y_2 values and not on the fixed z value. In this case the conditional utility functions $u(\cdot,z^o)$ and $u(\cdot,z)$ would be strategically equivalent. Thus, from Theorem 4.1, we know that all the conditional utility functions along horizontal cuts in Figure 5.2 would be positive linear transformations of each other. In particular we would have

$$u(y,z) = g(z) + h(z)\, u(y,z') \tag{5.7}$$

for all y and z, where $g(\cdot)$ and $h(\cdot) > 0$ only depend on z and not on y. Of course the functional form of $g(\cdot)$ and $h(\cdot)$ will depend on the particular z' chosen. Note that if (5.7) holds for one z', then it must be valid for any other level.

Definition. We shall say that Y is utility independent of Z when (5.7) holds*.

---

*An alternate interpretation of utility independence is as follows. Given that Y is utility independent of Z, we know all utility functions of the form $u(\cdot,z)$ are strategically equivalent. If y is scalar and if the second derivative of $u(\cdot,z)$ with respect to y is continuous, one can define a conditional local risk aversion function over Y, for each z, analogous to that in Section 4.5. When Y is utility independent of Z, the local risk aversion function defined on Y for fixed z will not depend on z. The converse is also true. See Keeney [1973d] and Pollak [1973].

When Y is utility independent of Z the conditional utility function over Y given z does not strategically depend on z. Whenever this condition prevails we can elliptically talk about the utility function for Y without referring to any particular z. Already we have a great deal of structure to exploit!

Similarly it is natural to investigate whether Z is utility independent of Y. If we hold the Y-level fixed at y' say, and consider preferences for lotteries over Z, do these preferences depend on y'? If not, then Z is utility independent of Y, and we can talk about the utility function for Z without worrying about a dependence on y'.

In practice it is natural to investigate at an early stage whether Y is utility independent of Z and whether Z is utility independent of Y. Notice that all cases are possible: neither holds, one holds without the other, or both hold. To show that this is possible mathematically let us consider the following utility functions:

a. $u(y,z) = \dfrac{y^{\alpha}\ z^{\beta}}{y+z}$

b. $u(y,z) = g(z) + h(z)\ u_Y(y)$

c. $u(y,z) = k(y) + m(y)\ u_Z(z)$

d. $u(y,z) = k_1 u_Y(y) + k_2 y_Z(z) + k_3\ u_Y(y)\ u_Z(z)$

e. $u(y,z) = [\alpha + \beta\ u_Y(y)]\ [\gamma + \delta\ u_Z(z)]$

f. $u(y,z) = k_Y u_Y(y) + k_Z u_Z(z)$.

In case (a), neither attribute is utility independent of the other. In case (b), Y is utility independent of Z, but not in general vice versa. In (c), Z is utility independent of Y but not vice versa, and in cases (d), (e), and (f), each is utility independent of the other. We will investigate representation theorems in the sequel so that we shall be able to recognize from purely qualitative considerations whether a particular form is appropriate. Naturally, these representation results will materially affect the assessment protocol.

Utility independence is important because it is a necessary and sufficient condition for one to speak about a single utility function over one of the attributes. When Y is utility independent of Z, there is "a" utility function over Y. In this case, preferences for varying amounts of Y can be assessed after fixing Z at any convenient level. When Y is not utility independent of Z, then it is not meaningful to speak of a utility function over Y, and assessment of $u(\cdot,\cdot)$ becomes much more difficult. In this case, the conditional utility function for Y given $z = z'$ and the conditional utility function for Y given $z = z''$, that is $u(\cdot,z')$ and $u(\cdot,z'')$, respectively, are not strategically equivalent. Each must be assessed separately, and completely, since knowing one may imply little about the other.

5.2.2. Getting a Feeling for Utility Independence

Before proceeding, let us try to get a flavor for the manner in which utility independence helps us out considerably in the assessment of utility functions. If we are interested in preferences over (y,z) such that $y^o \leq y \leq y^*$ and $z^o \leq z \leq z^*$, then in the absence of any simplifying assumptions, one must directly assess the utility function u over the entire shaded region of Figure 5.3A.

However, suppose that Y is utility independent of Z. Then the general shape of the conditional utility functions $u(\cdot,z)$ cutting across Y for various levels of z must be positive linear transformations of each other. Hence, as we shall see later, we can get enough information to completely specify u by knowing the utilities of the darkened consequences in Figure 5.3B. This means we would have to assess and consistently scale three one-attribute conditional utility functions.

To take another case, if Z is utility independent of Y, and Y is not utility independent of Z, one can, for example, completely specify u by consistent assessment of the three one-attribute conditional utility functions in Figure 5.3C. In this case, the conditional utility functions $u(y,\cdot)$ cutting across Z for various levels of y are all positive linear transformations of each other. To illustrate this using the notation of the figure, we know

$$u(y^2,z) = k_1 + k_2\, u(y^1,z), \text{ all } z. \tag{5.8}$$

The conditional utility function $u(y^1,\cdot)$ is known because it is assessed, and the k_1 and k_2 are found by evaluating (5.8) at (y^2,z^1) and (y^2,z^2), two consequences whose utilities are known. The resulting simple equation is then easily solved. More about this later.

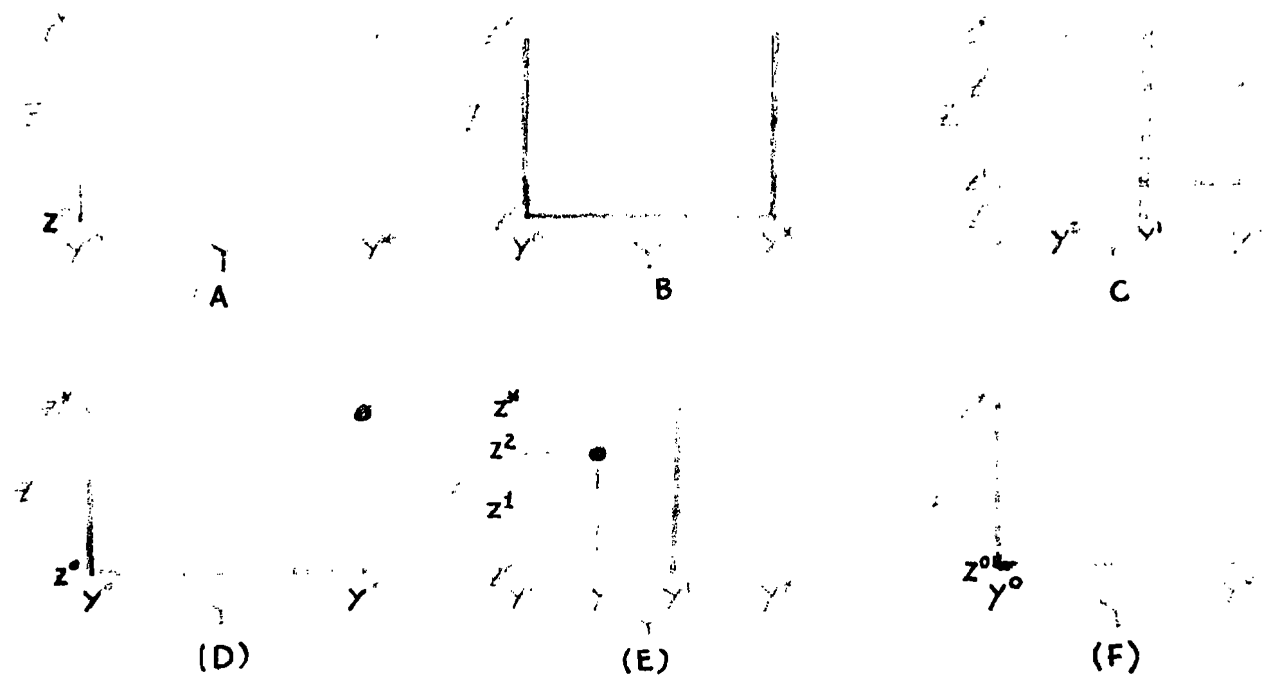


Figure 5.3. Exploiting Utility Independence to Simplify the Assessment of the Utility Function

Now suppose that Y and Z are utility independent of each other, a condition which we will define as <u>mutual utility independence</u>. Then taking Figure 5.3B as a starting point, one can see that the two conditional utility functions $u(y^o,\cdot)$ and $u(y^*,\cdot)$, as functions of z, must be positive linear transformations of each other. Thus, rather than assess $u(y^*,\cdot)$ for all z, we just need, for instance, the utilities of two points on the curve to fix the correct transformation. The implication is that, if Y and Z are mutually utility independent, one need only consistently assess two conditional utility functions and the utility of (y^*,z^*) to completely specify u. The consequences whose utilities are needed are blackened in Figure 5.3D.

Actually, when mutual utility independence holds, one has the freedom to choose any arbitrary conditional utility functions $u(\cdot,z^1)$ and $u(y^1,\cdot)$ and the utility of any arbitrary consequence (y^2,z^2) to specify $u(y,z)$ for all y,z. This freedom can be used to select y^1,z^1,y^2, and z^2 to simplify the decision maker's assessment problem. That is, he may feel more comfortable assessing $u(\cdot,z^1)$ than $u(\cdot,z^o)$ because his accumulated experience with consequences of the form (y,z^1) may be considerably greater. Figure 5.3E indicates what needs to be assessed in this case.

If mutual utility independence holds, and if also an additivity assumption holds, which we will describe later, we can completely assess $u(y,z)$ for all (y,z) using only the two conditional utility functions darkened in Figure 5.3F. This is the simplest two-attribute utility function one can have without simplifying the form of the one-attribute conditional utility functions or without making various tradeoff assumptions, such as a constant rate of substitution, discussed in Chapter 3. Thus, in some sense, the darkened information in Figure 5.3F represents the minimum actual information that needs to be assessed to specify $u(y,z)$ for all (y,z).

In the following sections, we begin discussing different forms of the utility function implied by the various sets of assumptions beginning with the simplest case (Fig. 5.3F) first. After presenting the results, we suggest procedures for verifying the requisite assumptions and assessing such utility functions and illustrate the techniques with a real-world example.

5.3. Additive Independence and the Additive Utility Function

The additive utility function which has the form

$$u(y,z) = k_Y u_Y(y) + k_Z u_Z(z),$$

where k_Y and k_Z are positive scaling constants, allows one to add the separate contributions of the two attributes to obtain the total utility. It is the best known of the multiattribute utility functions and important both because of its relevance to some real problems and its relative simplicity.

As one can easily verify, and as indicated in the previous section, the additive utility function implies that Y and Z are mutually utility independent. However, the converse is not true. Mutual utility independence does not imply that the utility function is additive.* The assumptions, in addition to mutual utility independence, which imply that the two-attribute utility function is additive are presented in Section 5.4. Here, an alternate set of assumptions about the decision maker's preferences which allow one to use the additive utility function is discussed.

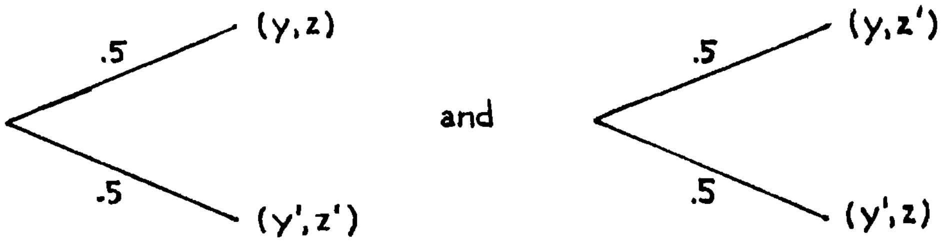
Necessary and sufficient conditions for the existence of an additive utility function can be stated using the concept of additive independence. Unfortunately, this terminology is not universal, and what we refer to as the "additive independence" condition has been referred to elsewhere as "independence." However, the adjective "additive" is needed to differentiate it from other independence conditions which we have introduced.

---

*For example, if $u(y,z) = y^{\alpha}z^{\beta}$, $1 \leq y \leq 10$, $1 \leq z \leq 10$, then Y and Z are mutually utility independent, but u is not additive. Taking logarithms, one gets $\log u(y,z) = \alpha \log y + \beta \log z$, which is clearly additive. However, this log u is not a utility function since it is not a positive linear transformation of u. On the other hand, log u is an appropriate value function since it preserves the ordering of the consequences (y,z).

<u>Definition</u>. Attributes Y and Z are <u>additive</u> <u>independent</u> if the paired preference comparison of any two lotteries, defined by two joint probability distributions on $Y \times Z$, depends only on their marginal probability distributions.

The condition above is written in the form stated because it is easy to generalize. In two dimensions, as we shall soon verify, an equivalent condition for Y and Z to be additive independent is that the lotteries



must be equally preferable (i.e. indifferent) for all (y,z) given an arbitrarily chosen y' and z'. Note that in each of these two lotteries, there is a one-half probability of getting either y or y' and a one-half probability of getting either z or z'. The only difference is how the levels of Y and Z are combined. From this it should be clear that it is not meaningful to have Y additive independent of Z, but Z not additive independent of Y. The property is reflexive, which is not the case with the other independence conditions we shall discuss.

<u>5.3.1. A Fundamental Result of Additive Utility Theory</u>

The following result is due to Fishburn [1965a], but exposited slightly differently here.

Theorem 5.1. Attributes Y and Z are additive independent, if and only if the two attribute utility function is additive. The additive form may be written either as

$$u(y,z) = u(y,z^o) + u(y^o,z), \tag{5.9}$$

or as

$$u(y,z) = k_Y u_Y(y) + k_Z u_Z(z) \tag{5.10}$$

where

(1) $u(y,z)$ is normalized by $u(y^o,z^o) = 0$ and $u(y^1,z^1) = 1$ for arbitrary y^1 and z^1 such that $(y^1,z^o) \succ (y^o,z^o)$ and $(y^o,z^1) \succ (y^o,z^o)$,

(2) $u_Y(y)$ is a conditional utility function on Y normalized by $u_Y(y^o) = 0$ and $u_Y(y^1) = 1$,

(3) $u_Z(z)$ is a conditional utility function on Z normalized by $u_Z(z^o) = 0$ and $u_Z(z^1) = 1$,

(4) $k_Y = u(y^1,z^o)$,

and

(5) $k_Z = u(y^o,z^1)$.

Proof. Clearly additive independence implies indifference between the two lotteries* $\langle (y,z),\ (y^o,z^o)\rangle$ and $\langle (y,z^o),\ (y^o,z)\rangle$, since they have the same marginal probability

---

*We remind the reader that the lottery denoted by <A,B> has consequences A and B each with probability one-half.

distributions on the attributes. Equating the expected utilities of these two lotteries gives us

$$\tfrac{1}{2}\, u(y,z) + \tfrac{1}{2}\, u(y^o,z^o) = \tfrac{1}{2}\, u(y,z^o) + \tfrac{1}{2}\, u(y^o,z). \tag{5.11}$$

If we arbitrarily set $u(y^o,z^o) = 0$, equation (5.9) follows directly from (5.11). Defining

$$u(y,z^o) = k_Y u_Y(y) \tag{5.12}$$

and

$$u(y^o,z) = k_Z u_Z(z), \tag{5.13}$$

to allow for free scaling of the one-attribute utility functions, and substituting these into (5.11) yields the result (5.10),

$$u(y,z) = k_Y u_Y(y) + k_Z u_Z(z).$$

To prove the other half of the theorem, that an additive utility function implies additive independence, note that the expected utility of any lottery using (5.9) or (5.10) depends only on the marginal probability distributions for Y and Z. Hence, preferences among such lotteries cannot depend on the joint probability distribution of Y and Z so the two attributes are additive independent. ◀

The assumptions required for the justification of an additive utility function are rather restrictive. They allow for no interaction of the decision maker's preferences for various amounts of the two attributes. Often, one might expect the desirability of various amounts of one attribute to depend on the specified level of the other attribute. For instance, consider a farmer with preferences for various amounts of sunshine and rain because of the impact this will have on the season's crops. Here, one might expect that the farmer's preferences for various amounts of sunshine to be different depending on whether there had been only a little rain or much rain. Such an interaction of preferences cannot be expressed with the additive utility function. In the subsequent sections, we will present some more general two-attribute utility functions, which do allow for certain types of interaction.

In Section 5.8 we discuss procedures and techniques that can be employed to (1) verify additive independence and (2) assess the appropriate unidimensional utility functions and scaling constants.

5.4. The Implications of Mutual Utility Independence

In this section we derive functional forms for evaluating the utility function $u(y,z)$ when attributes Y and Z are mutually utility independent. First it is illustrated how this assumption restricts the form of $u(y,z)$. Then the manner in which the resulting utility function accounts for possible interactions in the decision maker's preferences for the two attributes is discussed.

The theorems and proofs in this and the following sections are presented in terms that indicate exactly what must be empirically assessed to specify the utility function.

The results stated are consequently a bit more "bulky" than would be the case if we just wanted to prove the mathematical result and to ignore the assessment aspect.

Throughout Chapters 5 and 6, algebraic proofs will be given for the theorems. While this demonstrates the result for the general case we had in mind, it does not communicate as much of an intuitive feeling for the result as is possible with alternate less formal proofs. With a loss of some generality more natural proofs can be given for the results. So in some cases, especially here where we begin to discuss utility independence, we will offer a more intuitive, less formal proof in addition to the main algebraic one.

From (5.7), one sees that the assumption of mutual utility independence can be mathematically represented by

$$u(y,z) = c_1(z) + c_2(z)\, u(y,z_o), \text{ for all } y,z, \tag{5.14}$$

for an arbitrarily chosen z_o, and

$$u(y,z) = d_1(y) + d_2(y)\, u(y_o,z), \text{ for all } y,z, \tag{5.15}$$

for an arbitrarily chosen y_o. Equation (5.14) says Y is utility independent of Z and (5.15) says that Z is utility independent of Y.

5.4.1. The Multilinear Utility Function

When Y and Z are mutually utility independent, then

$u(y,z)$ can be expressed by the multilinear representation*

$$u(y,z) = k_Y u_Y(y) + k_Z u_Z(z) + k_{YZ} u_Y(y)\ u_Z(z),$$

where u, y_Y, and u_Z have a common origin and are consistently scaled by the scaling constants $k_Y > 0$, $k_Z > 0$, and k_{YZ}. Since the dimensionality of the utility functions u_Y and u_Z is less than the dimensionality of the original utility function u, its assessment is simplified when the stated assumptions hold.

A geometrical interpretation of the result for the case where Y and Z are scalar attributes is shown in Figure 5.4. Our result says that subject to the requisite assumptions, the utility of any consequence in the specified consequence space is uniquely determined by the relative utilities of the consequences along the heavy lines and at the heavy point in the figure.

To see why this is true, refer to Figure 5.4 and follow these steps:

(1) Consistently assess $u(\cdot,z_o)$, $u(y_o,\cdot)$ and $u(y_1,z_1)$.

(2) For any point Q (where Q can assume values A,B,...,H) denote the value of u at Q by u_Q. Let

---

*Because there are just two attributes, we could have referred to this utility function as the bilinear utility function. Since the representation is generalized to n attributes in Chapter 6, we have chosen to use the general term "multilinear" in this chapter also.

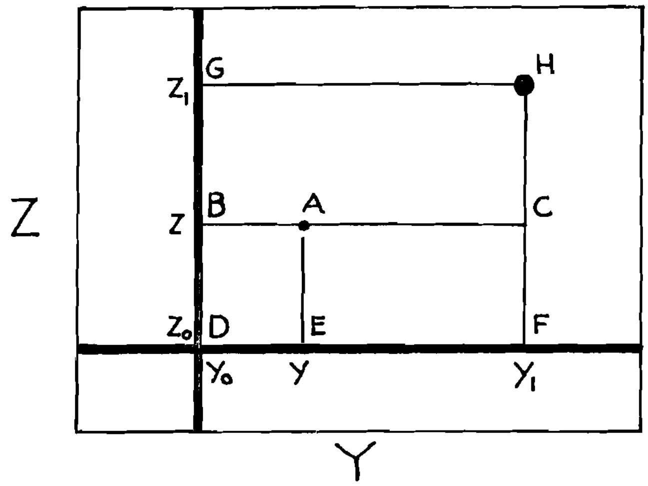


Figure 5.4. Mutual Utility Independence Implies u(y,z) is Completely Specified from the Utilities of the Heavy Shaded Consequences

A represent the generic point (y,z) and denote $u(y,z)$ as u_A.

(3) Express u_A in terms of u_B and u_C. This follows from the relationship of u_D, u_E, u_F, since Y is utility independent of Z.

(4) We know u_B but not u_C. Therefore, express u_C in terms of u_F and u_H, using the fact that Z is utility independent of Y and using the utilities u_D, u_B, and u_G.

(5) The utility u_A is now expressed in terms of the known utilities u_B, u_F, and u_H.

If A were not originally chosen to fall within the region 'cornered' by D, F, G, and H, slightly different steps--using identical reasoning--would be required. With this motivation, we prove the following.

Theorem 5.2. If Y and Z are mutually utility independent, then the two-attribute utility function is multilinear. In particular u can be written in the form

$$u(y,z) = u(y,z_o) + u(y_o,z) + ku(y,z_o)u(y_o,z), \tag{5.16}$$

or

$$u(y,z) = k_Y u_Y(y) + k_Z u_Z(z) + k_{YZ} u_Y(y)\, u_Z(z), \tag{5.17}$$

where

(1) $u(y,z)$ is normalized by $u(y_o,z_o) = 0$ and $u(y_1,z_1) = 1$ for arbitrary y_1 and z_1 such that $(y_1,z_o) \succ (y_o,z_o)$ and $(y_o,z_1) \succ (y_o,z_o)$,

(2) $u_Y(y)$ is a conditional utility function on Y normalized by $u_Y(y_o) = 0$ and $u_Y(y_1) = 1$,

(3) $u_Z(z)$ is a conditional utility function on Z normalized by $u_Z(z_o) = 0$ and $u_Z(z_1) = 1$,

(4) $k_Y = u(y_1, z_o)$,

(5) $k_Z = u(y_o, z_1)$,

and

(6) $k_{YZ}=1-k_Y-k_Z$ and $k = k_{YZ}/k_Y k_Z$.

<u>Proof.</u> Let us set the origin of $u(y,z)$ by

$$u(y_o, z_o) = 0. \tag{5.18}$$

Evaluating (5.14) at $y = y_o$,

$$u(y_o, z) = c_1(z) + c_2(z)\ u(y_o, z_o) = c_1(z). \tag{5.19}$$

Substituting (5.19) into (5.14) and evaluating at an arbitrary $y_1 \neq y_o$,

$$u(y_1, z) = u(y_o, z) + c_2(z)\ u(y_1, z_o)$$

or

$$c_2(z) = \frac{u(y_1, z) - u(y_o, z)}{u(y_1, z_o)}. \tag{5.20}$$

Using (5.19) and (5.20) in (5.14), we now have

$$u(y, z) = u(y_o, z) + \frac{u(y_1, z) - u(y_o, z)}{u(y_1, z_o)}\ u(y, z_o), \text{ all } z. \tag{5.21}$$

Similarly, by evaluating (5.15) successively at $z = z_o$ and at an arbitrary $z_1 \neq z_o$, it becomes

$$u(y,z) = u(y,z_o) + \frac{u(y,z_1) - u(y,z_o)}{u(y_o,z_1)} u(y_o,z), \text{ all } y. \tag{5.22}$$

Evaluating (5.22) at $y = y_1$ and substituting this into (5.21), we conclude

$$u(y,z) = u(y_o,z) + \left[\frac{u(y_1,z_o) + \dfrac{u(y_1,z_1) - u(y_1,z_o)}{u(y_o,z_1)} u(y_o,z) - u(y_o,z)}{u(y_1,z_o)} \right] \cdot u(y,z_o)$$

$$= u(y_o,z) + u(y,z_o) + \left[\frac{u(y_1,z_1) - u(y_1,z_o) - u(y_o,z_1)}{u(y_1,z_o)\; u(y_o,z_1)} \right] \cdot u(y_o,z)\; u(y,z_o). \tag{5.23}$$

Equation (5.23) can be written as (5.16),

$$u(y,z) = u(y_o,z) + u(y,z_o) + k\, u(y_o,z)\, u(y,z_o),$$

where k is an empirically evaluated constant defined by

$$k = \frac{u(y_1,z_1) - u(y_1,z_o) - u(y_o,z_1)}{u(y_1,z_o)\; u(y_o,z_1)} . \tag{5.24}$$

To provide for arbitrary scaling of the conditional utility functions, we can define u_Y and u_Z such that

$$k_Y u_Y(y) = u(y,z_o) \text{ and } k_Z u_Z(z) = u(y_o,z), \tag{5.25}$$

??

where k_Y and k_Z are positive scaling constants and where u_Y and u_Z are scaled as stated in the theorem. Then, substituting (5.25) into (5.16) and defining $k_{YZ} = k\, k_Y\, k_Z$ gives us (5.17). From (5.18) and (5.25), it follows that the origins of u_Y and u_Z must be

$$u_Y(y_o) = 0 \text{ and } u_Z(z_o) = 0$$

respectively. It is important to realize that there are no other restrictions on the functional forms of the conditional utility functions u_Y and u_Z. ◀

5.4.2. Use of Iso-Preference Curves*

Because the decision maker may be unaccustomed to thinking in terms of a particular attribute, it may be difficult to assess one of the conditional utility functions required to use (5.16). However, one might be able to obtain an iso-preference curve, that is, a set of all consequences which are equally desirable to the decision maker. In this section, we show that an iso-preference curve may be substituted for one of the conditional utility functions required by Theorem 5.2. provided it covers the same range.

---

*This section describes another way of assessing a utility function when each attribute is utility independent of the other. It exploits the existence of an assessed iso-preference curve. The section can be skipped without interfering with the reading of the main results of the chapter. However, other sections using iso-preference curves should then also be skipped. These sections will be appropriately designated.

A geometrical interpretation of the result is shown in Figure 5.5 for the case where Y and Z are scalar attributes. We prove that if Y and Z are mutually utility independent, then u(y,z) is uniquely determined in the specified consequence space by assessing a conditional utility function along the vertical heavy line, a utility for the heavy point in the figure and the iso-preference curve.

The reasoning goes as follows:

(1) Determine u on L in Figure 5.5 setting $u_C = 0$ and assess u_P for consequence P.

(2) Then u along the indifference curve N must be a zero.

(3) Select A with arbitrary coordinates (y,z).

(4) Express u_F in terms of u_H and u_P, using the fact that Z is utility independent of Y and using u_D, u_G, and u_M.

(5) Similarly, express u_K in terms of u_H and u_P using u_J, u_D, and u_M.

(6) Express u_A in terms of u_G and u_F, using the fact that Y is utility independent of Z and the relationship of u_J, u_B, and u_K.

Since u_G and u_F are known, the reasoning is complete. If A had not been in the region cornered by C, H, P, and M, a slightly altered proof using the same reasoning would be required. This provides the motivation for

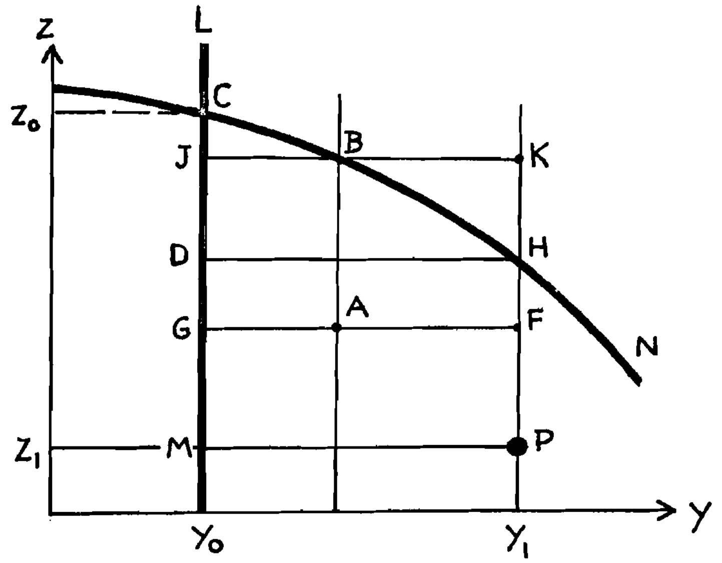


Figure 5.5. Mutual Utility Independence Implies u(y,z) is Completely Specified from the Utilities of the Heavy Shaded Consequences

<u>Theorem 5.3.</u> If Y and Z are mutually utility independent, then

$$u(y,z) = \frac{u(y_o,z) - u(y_o,z_n(y))}{1 + k\, u(y_o,z_n(y))} . \tag{5.26}$$

where

(1) $u(y_o,z_o) = 0$,

(2) $z_n(y)$ is defined such that $(y,z_n(y)) \sim (y_o,z_o)$,

and

(3) $$k = \frac{u(y_o,z_1) - u(y_1,z_1) - u(y_o,z_n(y_1))}{u(y_1,z_1)\, u(y_o,z_n(y_1))} , \tag{5.27}$$

where (y_1,z_1) is arbitrarily chosen such that (y_o,z_o) and (y_1,z_1) are not indifferent.

<u>Proof.</u> Let us define $z_n(y)$ to be such that the set $\{(y,z_n(y));\ \text{all } y\}$ is an iso-preference curve over all Y. We can set the utility level of the iso-preference curve* and the origin of $u(y,z)$ by

$$u(y,z_n(y)) = 0. \tag{5.28}$$

Let us designate $z_n(y_o)$ as z_o. Then of course,

$$u(y_o,z_o) = 0,$$

which is consistent with our origin in Theorem 5.2. Thus

---

*For any y, we only need to determine one $z_n(y)$ such that (5.28) holds in order to specify $u(y,z)$.

we can evaluate (5.16) at $(y, z_n(y))$ and solve for $u(y, z_o)$ to find

$$u(y,z_o) = \frac{-u(y_o,z_n(y))}{1 + k\, u(y_o,z_n(y))} \,. \tag{5.29}$$

Now substituting (5.29) into (5.16) and rearranging, we get

$$u(y,z) = u(y_o,z) + \frac{-\,u(y_o,z_n(y))}{1+k\; u(y_o,z_n(y))} + k\; u(y_o,z)\left[\frac{-\,u(y_o,z_n(y))}{1+k\; u(y_o,z_n(y))}\right]$$

$$= \frac{u(y_o,z) - u(y_o,z_n(y))}{1 + k\, u(y_o,z_n(y))} \,. \tag{5.30}$$

To determine k from (5.24), we need to know $u(y_1, z_o)$. We can assess $u(y_1, z_1)$ for an arbitrary (y_1, z_1) such that it is not indifferent to (y_o, z_o). Substituting this into (5.30) yields

$$u(y_1,z_1) = \frac{u(y_o,z_1) - u(y_o,z_n(y_1))}{1 + k\, u(y_o,z_n(y_1))}$$

which, after rearranging, gives us the desired result

$$k = \frac{u(y_o,z_1) - u(y_1,z_1) - u(y_o,z_n(y_1))}{u(y_1,z_1)\; u(y_o,z_n(y_1))} \,. \blacktriangleleft$$

5.4.3. The Product Representation

The multilinear form

$$u(y,z) = u(y,z_o) + u(y_o,z) + ku(y,z_o)\; u(y_o,z) \tag{5.16}$$

of Theorem 5.2 has a strategically equivalent product representation provided that $k \neq 0$. To this end, let

$$
\begin{aligned}
u'(y,z) &= k\,u(y,z) + 1 \\
&= k\,u(y_o,z) + k\,u(y,z_o) + k^2 u(y_o,z)\,u(y,z_o) + 1 \\
&= [k\,u(y,z_o) + 1][k\,u(y_o,z) + 1] \\
&= u'(y,z_o)\,u'(y_o,z). \qquad (5.31)
\end{aligned}
$$

When $k > 0$, then $u'(y,z_o)$ and $u'(y_o,z)$ are conditional utility functions for Y and Z, respectively. When $k < 0$, they are just the negative of the respective utility functions. Thus, <u>if two attributes are mutually utility independent</u>, <u>their utility function can be represented by either a product form</u>, <u>when</u> $k \neq 0$, <u>or an additive form</u>, <u>when</u> $k = 0$.

5.4.4. Additive Representation

It would be interesting to know when k in (5.16) is zero. In this case the multilinear representation reduces to the additive representation discussed in Section 5.3. We can state the following

<u>Theorem 5.4.</u> If Y and Z are mutually utility independent and if

$$\langle (y_3,z_3),\ (y_4,z_4)\rangle \sim \langle (y_3,z_4),\ (y_4,z_3)\rangle$$

for some y_3,y_4,z_3,z_4, such that (y_3,z_3) is not indifferent to either (y_3,z_4) or (y_4,z_3) then

$$u(y,z) = u(y,z_o) + u(y_o,z)$$

where $u(y,z)$ is normalized by

(1) $u(y_o,z_o) = 0$

and

(2) $u(y_1,z_1) = 1$ for arbitrary y_1 and z_1 such that $(y_1,z_o) \succ (y_o,z_o)$ and $(y_o,z_1) \succ (y_o,z_o)$.

Remark A. Given the above hypotheses, an alternate form of the utility function is given by (5.17) with its usual normalizations and with $k_{YZ} = 0$.

Remark B. The difference between Theorems 5.1 and 5.4 should be clarified. In Theorem 5.1 we require that $\langle (y,z),(y',z')\rangle \sim \langle (y,z'),(y',z)\rangle$ for all (y,z). Theorem 5.4 requires this indifference condition to hold for only one set of four points. However, of course, Theorem 5.4 also requires mutual utility independence.

Proof. Equating the expected utilities of the lotteries, we have

$$\tfrac{1}{2}\, u(y_3,z_3) + \tfrac{1}{2}\, u(y_4,z_4) = \tfrac{1}{2}\, u(y_3,z_4) + \tfrac{1}{2}\, u(y_4,z_3).$$

Evaluating these terms using (5.16), canceling and transposing, we find

$$k\, [u(y_o,z_3) - u(y_o,z_4)][u(y_3,z_o) - u(y_4,z_o)] = 0.$$

Since $u(y_3,z_3) \neq u(y_3,z_4)$, because of utility independence, $u(y_o,z_3) \neq u(y_o,z_4)$, and similarly, $u(y_3,z_o) \neq u(y_4,z_o)$. Therefore k must be zero and (5.16) reduces to the additive representation. ◀

From Theorem 5.4, it should be clear that additive independence implies mutual utility independence, but the converse is not true. Additive independence is obviously the stronger condition.

COROLLARY.* Given the same conditions as the Theorem 5.4, $u(y,z)$ is completely specified by

(1) $u(y_o,z)$, a conditional utility function for Z, for arbitrary y_o;

(2) an iso-preference curve over all Y.

Proof. In this case, $k = 0$, and (5.26) becomes

$$u(y,z) = u(y_o,z) - u(y_o,z_n(y)). \blacktriangleleft$$

5.4.5. Interpretation and Implications of Parameter k

There is an interesting manner in which to interpret the parameter k. Consider the two fifty-fifty lotteries $\langle A,C\rangle$ and $\langle B,D\rangle$ illustrated in Figure 5.6. We will assume that preferences are increasing in both Y and Z in the figure. If this were not originally the case, simple transformations as indicated in Chapter 4 could be used to meet this requirement. Using the multilinear utility function (5.16) to calculate expected utilities, it is easy to show that

$$\langle A,C\rangle \begin{Bmatrix} \succ \\ \sim \\ \prec \end{Bmatrix} \langle B,D\rangle \Longleftrightarrow k \begin{Bmatrix} > \\ = \\ < \end{Bmatrix} 0 \quad .$$

In some sense, consequences A and C are such that one either gets a high level of both Y and Z or a low level of each. On the contrary, with B and D, one either gets a high

*This corollary should also be skipped if the reader did not read Section 5.4.2.

$$\langle A,C\rangle \left\{\begin{matrix}\succ\\ \sim\\ \prec\end{matrix}\right\} \langle B,D\rangle \Longleftrightarrow k \left\{\begin{matrix}>\\ =\\ <\end{matrix}\right\} 0 \Longleftrightarrow \left\{\begin{matrix}\text{Y and Z are complements}\\ \text{no interaction of preference between Y and Z}\\ \text{Y and Z are substitutes}\end{matrix}\right.$$

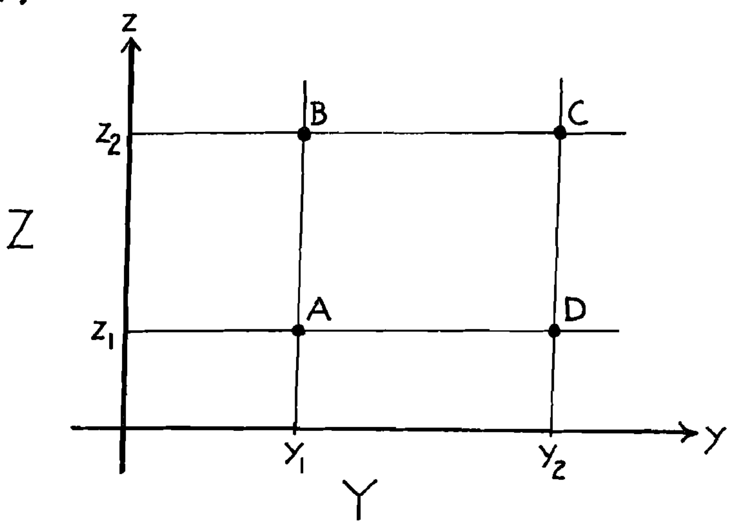


Figure 5.6. Using Lotteries <A,C> and <B,D> to Interpret the Interaction Term in the Multilinear Utility Function

level of Y or Z, but not a lot (or a little) of both. Thinking about it this way, if $\langle A,C \rangle$ is preferred, it is as if one needs an increase of Y to complement an increase in Z in going from A to C. Otherwise the full worth of the increase in Z could not be exploited. On the otherhand, to prefer $\langle B,D \rangle$ implies that it is important to do well in terms of at least one attribute, and given a high level of Y, the increased preference due to an increase in Z is not so much. Thus Y and Z can be thought of as substitutes for each other.

Two simple illustrations may help clarify the idea. First suppose the president of a corporation has two large divisions operating in entirely different markets. She may be interested in profits of division 1, represented by Y, and profits of division 2, represented by Z. Achievement on these attributes would likely be substitutes for each other. If division 1 was doing well financially, the president would likely not be as concerned about division 2, as in the case when division 1 is doing poorly. If either division was quite successful, the corporation as a whole would probably live comfortably.

To illustrate a complementary case, consider the general who is fighting a battle on two fronts. Attribute Y and Z represents the performance on the respective fronts. Here, if either of the fronts break, the consequence is probably almost as bad as if both break. 'Average' achievement on both fronts would likely be preferred by the general to 'fantastic success' on one and 'miserable failure' on the other. Hence, these attributes have a complementary effect. Complementarity, as we have used it here, is just a formalization--though somewhat weakened--of the saying "a chain is only as strong as its weakest link".

Further insight into the implications of parameter k can be seen if we rewrite (5.16) as

$$u(y,z) = u(y,z_o) + u(y_o,z)\ [1 + k\ u(y,z_o)]. \qquad (5.32)$$

Now from (5.32) it is clear that if $u(y_o,z)$ is increasing in z,

$$k \begin{Bmatrix} < \\ = \\ > \end{Bmatrix} 0 \Longrightarrow \left.\frac{\partial u(y,z)}{\partial z}\right|_{y=y_1} \begin{Bmatrix} > \\ = \\ < \end{Bmatrix} \left.\frac{\partial u(y,z)}{\partial z}\right|_{y=y_2}$$

for $u(y_2,z_o) > u(y_1,z_o)$.

Thus, if k is negative [positive, zero], and $u(y_o,z)$ is increasing, the increase in utility due to an incremental increase in Z is smaller [greater, the same] for more preferred amounts of Y. In the case where $u(y_o,z)$ is decreasing in z,

$$k \begin{Bmatrix} < \\ = \\ > \end{Bmatrix} 0 \Longrightarrow \left.\frac{\partial u(y,z)}{\partial z}\right|_{y=y_1} \begin{Bmatrix} < \\ = \\ > \end{Bmatrix} \left.\frac{\partial u(y,z)}{\partial z}\right|_{y=y_2}$$

for $u(y_1,z_o) > u(y_2,z_o)$.

In this sense, again k may be interpreted as a parameter that indicates the manner in which the amount of one attribute affects the value of the other attribute. If k is positive, more preferred amounts of Y complement more preferred amounts of Z. Just the reverse is true where k is negative. Here, one can consider more preferred amounts of Y and Z as being substitutes for each other. And in the additive case when $k = 0$, there is no interaction of preference between Y and Z.

5.5. Use of Certainty Equivalents

Recall that if we have a lottery* $(\tilde{y},z)$, the certainty equivalent for $\tilde{y}$ given z is the amount $\hat{y}_z$ such that

$$u(\hat{y}_z,z) = E[u(\tilde{y},z)] \tag{5.33}$$

where $\hat{y}_z$ in general will depend on the level z. Because the expected utility $E[u(\tilde{y},z)]$ of the lottery in (5.33) is difficult to interpret physically, it is often easier for the decision maker to visualize the situation by considering the equivalent certain consequence $(\hat{y}_z,z)$, a consequence with the same utility as the lottery. It would be especially convenient if the certainty equivalent for lotteries on Y did not depend on the level of Z, and similarly if the certainty equivalent for lotteries on Z did not depend on Y. Provided certain conditions hold, this is true, so one may use the respective certainty equivalents in calculating

---

*A lottery over $Y \times Z$ with an uncertain outcome $\tilde{y}$ coupled with a certain outcome z will be denoted by $(\tilde{y},z)$. We assume that a probability measure is known for the uncertain quantity (random variable) $\tilde{y}$.

expected utility and assessing implications of alternative decisions.

Consider the lottery represented by $(\tilde{y},\tilde{z})$ where Y and Z are mutually <u>utility</u> independent. We do <u>not</u> assume, however, that random variables $\tilde{y}$ and $\tilde{z}$ are <u>probabilistically</u> independent. Then using (5.16), expected utility can be calculated as follows

$$E[u(\tilde{y},\tilde{z})] = E[u(\tilde{y},z_o)] + E[u(y_o,\tilde{z})] + k\; E[u(\tilde{y},z_o)\; u(y_o,\tilde{z})], \tag{5.34}$$

since the expected value of a sum is the sum of the expected values. In the cases where Y and Z are also probabilistically independent, (5.34) becomes

$$E[u(\tilde{y},\tilde{z})] = E[u(\tilde{y},z_o)] + E[u(y_o,\tilde{z})] + k\; E[u(\tilde{y},z_o)]\; E[u(y_o,\tilde{z})]. \tag{5.35}$$

Now (5.35) can be reduced using (5.33) to

$$E[u(\tilde{y},\tilde{z})] = u(\hat{y},z_o) + u(y_o,\hat{z}) + k\; u(\hat{y},z_o)\; u(y_o,\hat{z}) = u(\hat{y},\hat{z}). \tag{5.36}$$

This illustrates

<u>Theorem 5.5.</u> Given a lottery of the form $(\tilde{y},\tilde{z})$, separate certainty equivalents $\hat{y}$ and $\hat{z}$ for $\tilde{y}$ and $\tilde{z}$ respectively may be calculated using the marginal probability distributions on $\tilde{y}$ and $\tilde{z}$ to form a joint certainty equivalent $(\hat{y},\hat{z})$ for $(\tilde{y},\tilde{z})$ provided either

(i) the attributes are mutually utility independent and probabilistically independent, or

(ii) the attributes are additive independent.

That condition (i) is sufficient is proven by (5.36). When additive independence holds, $k = 0$ in (5.34) from which the desired result immediately follows.

5.6. Utility Functions With One Utiltiy Independent Attribute*

In the previous sections, we have been concerned with representing and assessing two-attribute utility functions when the attributes are at least mutually utility independent. That is, all our work dealt with assumptions which were at least that strong. In this section, we look at the implications of the weaker assumption, where only one attribute is utility independent of the other. It is shown that the two-attribute utility function can be specified by either three conditional utility functions, or two conditional utility functions and an iso-preference curve, or one conditional utility function and two iso-preference curves. Special cases of these results, including the additive and multilinear utility functions, are indicated.

For all the work in this section, we will denote the attributes as Y and Z and assume Z is utility independent of Y. That is, for any arbitrary y_o,

$$u(y,z) = c_1(y) + c_2(y)\, u(y_o,z), \text{ where } c_2(y) > 0, \text{ all } y. \tag{5.37}$$

---

* Material in this section is adapted from Keeney [1971].

5.6.1. Assessments In Terms of Three Conditional Utility Functions

Let us begin with an illustration of what we will prove. If Z is utility independent of Y, then $u(y,z)$ is completely specified by two arbitrary conditional utility functions for Y and one conditional utility function for Z, subject to consistent scaling. To see this in the case where Y and Z represent scalar attributes, consider Figure 5.7. If we consistently assess the utilities along the heavy lines in the figure, we will have enough information to assign the utility to every consequence. For example, consider an arbitrary point A with coordinates (y,z). The utility of A can be expressed as a linear combination of the utilities u_B and u_C where the weights are determined (since Z is utility independent of Y) by the values of u_D, u_E, and u_F.

As an alternate way of looking at the same proof, consider any vertical line at arbitrary point y. The utility function $u(y,\cdot)$ must be strategically equivalent to the function $u(y_o,\cdot)$--which is given. The utilities at B and C serve to normalize $u(y,\cdot)$.

To formalize this argument, we have

<u>Theorem 5.6.</u> If Z is utility independent of Y, then

$$u(y,z) = u(y,z_o)\,[1-u(y_o,z)] + u(y,z_1)\,u(y_o,z), \tag{5.38}$$

where $u(y,z)$ is normalized by $u(y_o,z_o) = 0$ and $u(y_o,z_1) = 1$.

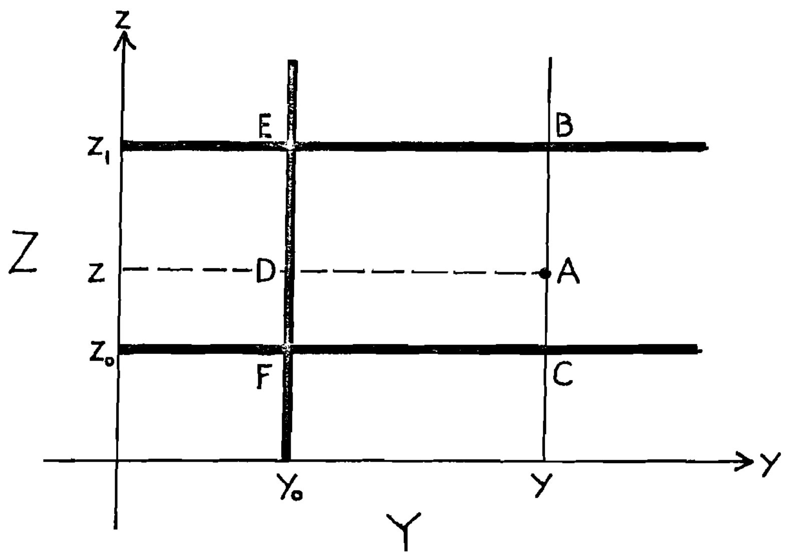


Figure 5.7. Z Utility Independent of Y Implies u(y,z) is Completely Specified from the Utilities of the Heavy Shaded Consequences

Proof. We can define z_o and z_1 to insure $u(y_o,z_1) > u(y_o,z_o)$ and then arbitrarily set the origin and unit of measure of $u(y,z)$ by

$$u(y_o,z_o) = 0 \tag{5.39}$$

and

$$u(y_o,z_1) = 1 \quad . \tag{5.40}$$

Since Z is utility independent of Y, (5.37) holds. Evaluating (5.37) at $z = z_o$ and using (5.39), we find

$$u(y,z_o) = c_1(y) + c_2(y)\, u(y_o,z_o) = c_1(y). \tag{5.41}$$

Combining (5.41) and (5.37) and evaluating at $z = z_1$,

$$u(y,z_1) = u(y,z_o) + c_2(y)\, u(y_o,z_1),$$

and using (5.40), we conclude

$$c_2(y) = u(y,z_1) - u(y,z_o). \tag{5.42}$$

Now, (5.41) and (5.42) can be substituted into (5.37) to give

$$\begin{aligned} u(y,z) &= u(y,z_o) + [u(y,z_1) - u(y,z_o)]\, u(y_o,z) \\ &= u(y,z_o)\, [1 - u(y_o,z)] + u(y,z_1)\, u(y_o,z), \end{aligned}$$

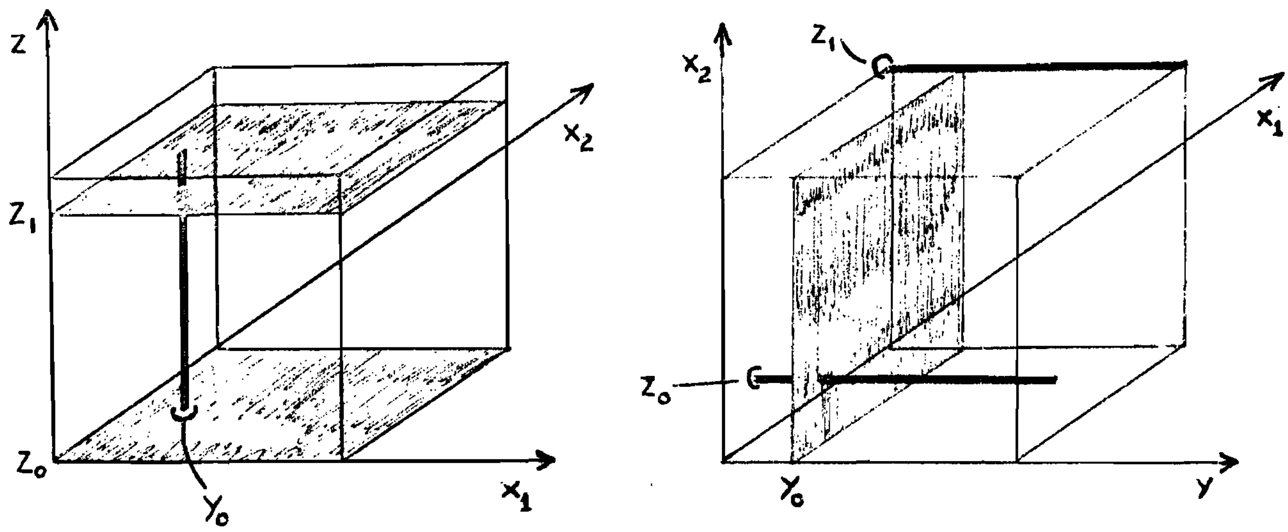
which is the desired result. ◀

It should be noted that $u(y_o,\cdot)$, $u(\cdot,z_o)$, and $u(\cdot,z_1)$ are conditional utility functions. Equations (5.39) and (5.40) specify the origin and unit of measure of $u(y_o,\cdot)$

and fix one point on the $u(\cdot,z_o)$ and $u(\cdot,z_1)$ curves. One other point on each of the latter two conditional utility functions must be evaluated empirically in order to set their units of measure equal to that of $u(y_o,\cdot)$ and thus insure consistency of the unit of measure of $u(\cdot,\cdot)$. This can be done by finding a consequence (y_o,z_2) which is indifferent to a consequence (y_2,z_o). Thus, $u(y_o,z_2)$ equals $u(y_2,z_o)$ which specifies a second point on $u(\cdot,z_o)$, thereby fixing its unit of measure. Similarly, one can find a (y_o,z_3) which is indifferent to (y_3,z_1) to consistently fix the unit of measure of $u(\cdot,z_1)$.

To provide a better understanding of (5.38), we offer graphical illustrations of two special cases. First, let us assume Y is two-dimensional, that is $y \equiv (x_1,x_2)$, and Z is one-dimensional. For this case, Theorem 5.6 states that provided Z is utility independent of Y, $u(y,z)$ can be specified by assessing two two-dimensional conditional utility functions, $u(\cdot,z_o)$ and $u(\cdot,z_1)$, and the one-dimensional conditional utility function $u(y_o,\cdot)$. Referring to Figure 5.8A, this means we must assess the relative utilities of the shaded consequences to specify $u(\cdot,\cdot)$.

As a second illustration, suppose Y is one-dimensional and $z \equiv (x_1,x_2)$. In this case, the theorem states $u(\cdot,\cdot)$ is specified by two one-dimensional conditional utility functions, $u(\cdot,z_o)$ and $u(\cdot,z_1)$, and the two-dimensional conditional utility function $u(y_o,\cdot)$, provided Z is utility independent of Y. Thus, to determine $u(\cdot,\cdot)$ in this case,



(A) Z is utility independent of $Y \equiv \{X_1, X_2\}$

(B) $Z \equiv \{X_1, X_2\}$ is utility independent of Y

Figure 5.8. Assessing Utilities for the Shaded Consequences Completely Specifies the Utility Function

one must assess the relative utilities of the consequences shaded in Figure 5.8B.

5.6.2. Substitution of One Iso-Preference Curve for One Conditional Utility Function*

In certain problems, it may be more convenient to assess an iso-preference curve than a conditional utility function. We prove that in evaluating $u(y,z)$ an iso-preference curve may be substituted for either a conditional utility function for Y or Z provided it covers the same range. Let us begin with

Theorem 5.7. If Z is utility independent of Y, then

$$u(y,z) = u(y,z_o) + \left[\frac{u(y_o,z_1) - u(y,z_o)}{u(y_o,z_n(y))}\right] u(y_o,z) \tag{5.43}$$

where

(1) $u(y_o,z_o) = o$

(2) $z_n(y)$ is defined such that $(y,z_n(y)) \sim (y_o,z_1)$ for an arbitrary z_1.

[Remark before proof: Thus to implement the results of this theorem one must ascertain that Z is utility independent of Y, assess the functions $u(\cdot,z_o)$ and $u(y_o,\cdot)$, and determine one iso-preference curve with a full range of y's.]

---

*This subsection can be omitted without interfering with the continuity of the presentation. It should be omitted if the reader skipped subsection 5.4.2.

Proof. We will set the origin of $u(y,z)$ by

$$u(y_o,z_o) = 0, \tag{5.44}$$

and define $z_n(y)$ to be such that the set $\{(y,z_n(y)):$ all $y\}$ is an iso-preference curve over all Y. Since the curve $\{(y,z_n(y)):$ all $y\}$ must intersect the line $\{(y_o,z):$ all $z\}$, we can denote the intersection point as (y_o,z_1) and set the utility level of the iso-preference curve by

$$u(y,z_n(y)) = u(y_o,z_1). \tag{5.45}$$

Evaluating (5.37) at $z = z_o$ and at $z = z_n(y)$, we respectively find

$$u(y,z_o) = c_1(y) + c_2(y)\ u(y_o,z_o) = c_1(y) \tag{5.46}$$

and

$$\begin{aligned} u(y,z_n(y)) &= u(y_o,z_1) = c_1(y) + c_2(y)\ u(y_o,z_n(y)) \\ &= u(y,z_o) + c_2(y)\ u(y_o,z_n(y)), \end{aligned}$$

or

$$c_2(y) = \frac{u(y_o,z) - u(y,z_o)}{u(y_o,z_n(y))} . \tag{5.47}$$

Substituting (5.45) and (5.46) into (5.37), we conclude

$$u(y,z) = u(y,z_o) + \left[\frac{u(y_o,z_1) - u(y,z_o)}{u(y_o,z_n(y))}\right] u(y_o,z). \blacktriangleleft$$

In the special case when the iso-preference curve goes through (y_o, z_o) [i.e., the case where $z_1 = z_o$], $u(y_o, z_1) = 0$ and (5.43) simplifies to

$$u(y,z) = u(y,z_o)\left[1 - \frac{u(y_o,z)}{u(y_o,z_n(y))}\right]. \tag{5.48}$$

The geometrical interpretation of Theorem 5.7 in the case where Y and Z are scalar attributes is given in Fig. 5.9. To specify $u(\cdot,\cdot)$ one must consistently assess the utilities of the consequences covered by heavy lines in the figure.

It is also of interest to use an iso-preference curve in place of the conditional utility function for Z in the assessment of $u(\cdot,\cdot)$. Let us formalize this with

Theorem 5.8. If Z is utility independent of Y, then

$$u(y,z) = \frac{u(y,z_o)u(y_n(z),\ z_1) - u(y,z_1)u(y_n(z),z_o)}{u(y_n(z),z_1) - u(y_n(z),z_o)} \tag{5.49}$$

where

(1) $u(y_o,z_o) = 0,\ z_1 \neq z_o$,

and

(2) $y_n(z)$ is defined such that $(y_n(z),z) \sim (y_o,z_o)$.

[Remark before proof: Thus to implement the results of this theorem one must ascertain that Z is utility independent of Y, assess the functions $u(\cdot,z_o)$, $u(\cdot,z_1)$, and determine one

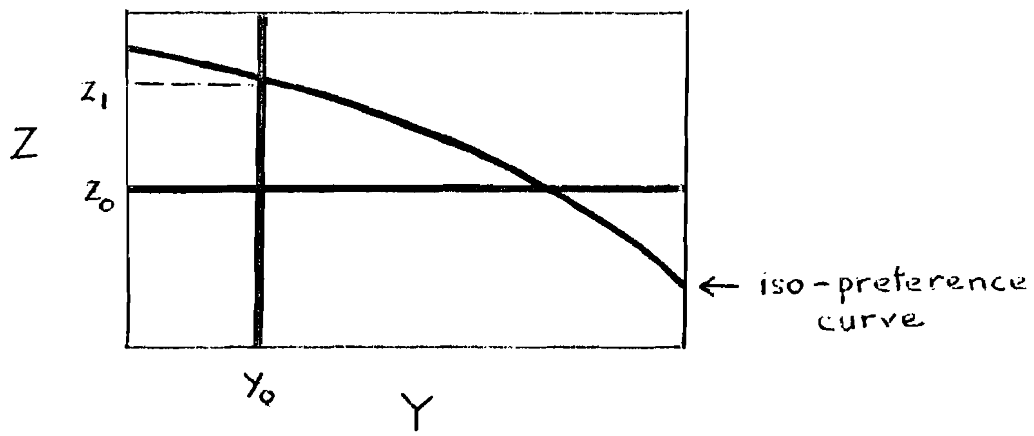


Figure 5.9. Z Utility Independent of Y Implies u(y,z) is Completely Specified from the Utilities of the Heavy Shaded Consequences

iso-preference curve with a full range z's.]

Proof. Let us define the origin of $u(y,z)$ as the point where the iso-preference curve, $\{(y_n(z),z): \text{ all } z\}$, crosses the line $\{(y,z_o): \text{ all } y\}$. This must occur at some y, call it y_o, and thus

$$u(y_n(z),z) = u(y_o,z_o) = 0 \ .$$

Furthermore, we can set the unit of measure by $u(y_o,z_1) = 1$. Thus, since Z is utility independent of Y, we can use (5.38) to evaluate $u(y_n(z),z)$ to yield

$$u(y_n(z),z) = 0 = u(y_n(z),z_o)[1-u(y_o,z)] + u(y_n(z),z_1)\ u(y_o,z),$$

which, after rearranging, becomes

$$u(y_o,z) = \frac{-u(y_n(z),z_o)}{u(y_n(z),z_1) - u(y_n(z),z_o)} \ . \tag{5.50}$$

Substituting (5.50) into (5.42) we get the result (5.49). ◀

A geometrical illustration of Theorem 5.9 is given in Figure 5.10 for the case where Y and Z are scalar attributes. Expression (5.49) gives one a method of evaluating $u(\cdot,\cdot)$ from the relative utilities of the consequences along the heavy lines in the figure. From the orientation of the iso-preference curve in Figure 5.10, it should be clear that preferences must be increasing in one attribute and decreasing in the other.

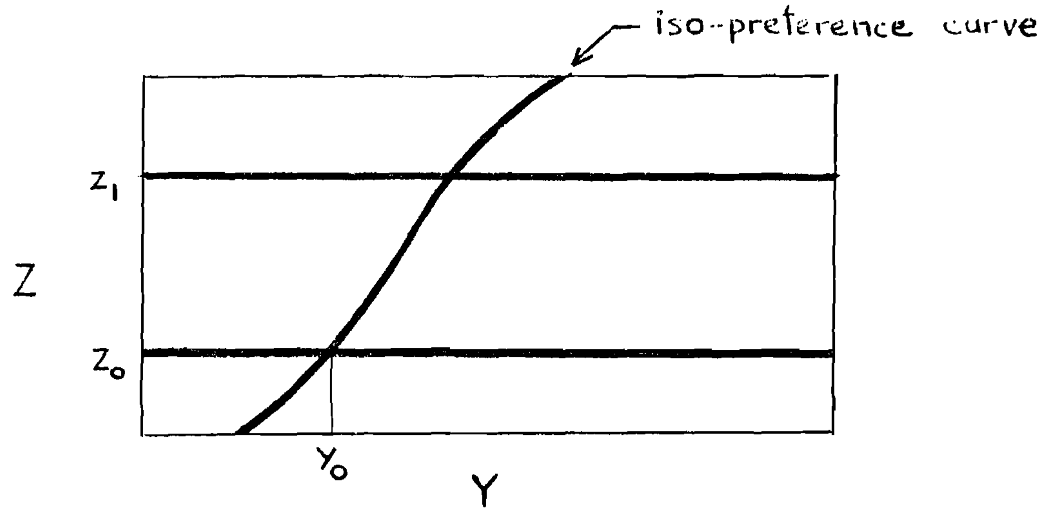


Figure 5.10. Z Utility Independent of Y Implies u(y,z) is Completely Specified from the Utilities of the Heavy Shaded Consequences

5.6.3. Use of Two Iso-Preference Curves*

It is possible to substitute an iso-preference curve for each of the conditional utility functions for Y necessary to implement (5.38). To this end, we prove

Theorem 5.9. If Z is utility independent of Y, then

$$u(y,z) = \frac{u(y_o,z) - u(y_o,z_m(y))}{u(y_o,z_n(y)) - u(y_o,z_m(y))} \quad , \tag{5.51}$$

where

(1) $u(y,z)$ is normalized by $u(y_o,z_o) = 0$ and $u(y_o,z_1) = 1$,

(2) $z_m(y)$ is defined such that $(y, z_m(y)) \sim (y_o,z_o)$,

and

(3) $z_n(y)$ is defined such that $(y, z_n(y)) \sim (y_o,z_1)$.

[Remark before proof: To implement this theorem, one must ascertain that Z is utility independent of Y, assess the function $u(y_o,\cdot)$, and determine two iso-preference curves with the full range of z's.]

Proof. Let us define $z_m(y)$ and $z_n(y)$ such that the sets $\{(y,z_m(y)):$ all $y\}$ and $\{(y,z_n(y)):$ all $y\}$ represent two iso-preference curves over all Y. Both iso-preference curves must intersect the line $\{(y_o,z):$ all $z\}$, so we can set the origin and unit of measure of $u(\cdot,\cdot)$ and define z_o and z_1 by

---

*Skip this subsection if the previous subsection was skipped.

$$u(y,z_m(y)) = u(y_o,z_o) = 0 \tag{5.52}$$

and

$$u(y,z_n(y)) = u(y_o,z_1) = 1. \tag{5.53}$$

By evaluating $u(y,z_m(y))$ and $u(y,z_n(y))$ using (5.38) we find, respectively

$$u(y,z_m(y)) = 0 = u(y,z_o)[1 - u(y_o,z_m(y))] + u(y,z_1)\ u(y_o,z_m(y)) \tag{5.54}$$

and

$$u(y,z_n(y)) = 1 = u(y,z_o)[1 - u(y_o,z_n(y))] + u(y,z_1)\ u(y_o,z_n(y)). \tag{5.55}$$

Equations (5.54) and (5.55) are two equations with two unknowns, which can be solved to yield

$$u(y,z_o) = \frac{-\ u(y_o,z_m(y))}{u(y_o,z_n(y)) - u(y_o,z_m(y))} \tag{5.56}$$

and

$$u(y,z_1) = \frac{1 - u(y_o,z_m(y))}{u(y_o,z_n(y)) - u(y_o,z_m(y))}\ . \tag{5.57}$$

Substituting (5.56) and (5.57) into (5.38), we conclude

$$u(y,z) = \frac{-u(y_o,z_m(y))[1 - u(y_o,z)] + [1 - u(y_o,z_m(y))]\ u(y_o,z)}{u(y_o,z_n(y)) - u(y_o,z_m(y))} \tag{5.58}$$

$$= \frac{u(y_o,z) - u(y_o,z_m(y))}{u(y_o,z_n(y)) - u(y_o,z_m(y))}\ . \quad \blacktriangleleft$$

When Y and Z both represent scalar attributes, Theorem 5.9 can be illustrated geometrically as shown in Figure 5.11. We have proven that provided Z is utility independent of Y, $u(\cdot,\cdot)$ is specified by assessing the relative utilities of the consequences along the heavy lines.

A utility function gives us a measure of the decision maker's attitude toward risky or uncertain situations. To assess the utility function, the decision maker must specify his preferences for lotteries. An iso-preference curve, on the other hand, yields no information about the decision maker's attitudes towards risk and can be assessed by comparing only certain consequences. Thus, since only one conditional utility function is necessary to implement (5.58), the decision maker's attitudes toward risks involving both uncertain Y and Z can be specified by considering risky situations involving only uncertain Z.

5.6.4. Special Cases*

As proven in Section 5.4, if Y and Z are mutually utility independent,

$$u(y,z) = u(y,z_o) + u(y_o,z) + k\, u(y,z_o)\, u(y_o,z), \tag{5.59}$$

where k is an empirically evaluated constant. It would be interesting to know what additional conditions must hold

---

*In a cursory reading of the book, the remainder of Section 5.6 can be skipped.

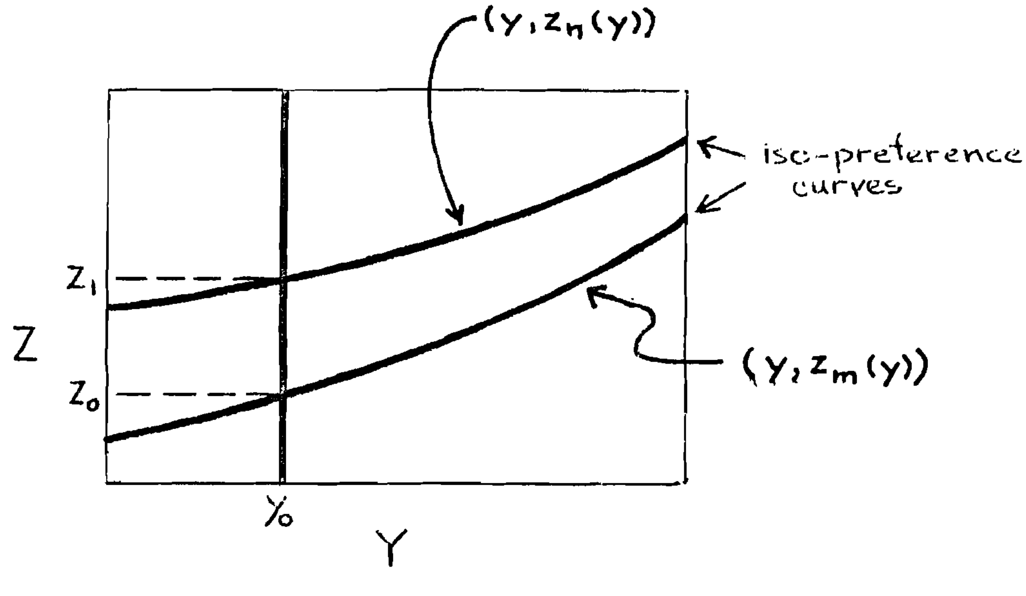


Figure 5.11. Z Utility Independent of Y Implies u(y,z) is Completely Specified from the Utilities of the Heavy Shaded Consequences

for the results of this section to reduce to the form (5.59) or to the additive utility function. To this end, we prove two results which can be thought of as corollaries to Theorem 5.6.

COROLLARY 1. Given Z is utility independent of Y, it is a necessary and sufficient condition for $u(y,z)$ to be of form (5.59) that

$$u(y,z_1) = a + b\,u(y,z_o), \tag{5.60}$$

for arbitrary $z_1 \neq z_o$, where a and $b > 0$ are constants. In other words, this corollary states that if Z is utility independent of Y, in order to get the multilinear utility function (5.59), we do not have to assert that all conditional utility functions $u(\cdot,z)$ be strategically equivalent. It is enough that there be merely a single pair, say $u(\cdot,z_o)$ and $u(\cdot,z_1)$, that are strategically equivalent.

Proof. For sufficiency, let us substitute (5.60) into (5.38) giving

$$\begin{aligned} u(y,z) &= u(y,z_o)[1 - u(y_o,z)] + [a + b\,u(y,z_o)]\,u(y_o,z) \\ &= u(y,z_o) + a\,u(y_o,z) + (b-1)\,u(y,z_o)\,u(y_o,z). \end{aligned} \tag{5.61}$$

Since from (5.39), $u(y_o,z_o) = 0$, evaluating (5.61) at $y = y_o$ yields

$$u(y_o,z) = 0 + a\,u(y_o,z) + 0$$

so

$$a = 1. \tag{5.62}$$

Substituting this result into (5.61) and defining $k \equiv b-1$, we get (5.59).

To prove that (5.60) is a necessary condition for (5.59), we only need to observe that (5.59) implies

$$u(y,z_1) = u(y_o,z_1) + [1 + k\, u(y_o,z_1)]\, u(y,z_o),$$

and that $u(y_o,z_1)$ and $[1 + k\, u(y_o,z_1)]$ are constants. ◀

COROLLARY 2. Given Z is utility independent of Y, $u(y,z)$ is additive if and only if $\langle(y_o,z_o),\ (y,z_1)\rangle$ is indifferent to $\langle(y_o,z_1),\ (y,z_o)\rangle$ for all y.

Proof. Equating expected utilities of the two lotteries,

$$\tfrac{1}{2}\, u(y_o,z_o) + \tfrac{1}{2}\, u(y,z_1) = \tfrac{1}{2}\, u(y_o,z_1) + \tfrac{1}{2}\, u(y,z_o), \quad \text{for all } y. \tag{5.63}$$

Recalling the origin and unit of measure of $u(y,z)$ were set by

$$u(y_o,z_o) = 0$$

and

$$u(y_o,z_1) = 1,$$

we can substitute these into (5.63) to give

$$u(y,z_1) = 1 + u(y,z_o). \tag{5.64}$$

Expression (5.64) is the necessary and sufficient condition for the multilinear form stated in Corollary 1. Noting for this case that a=1 and b=1, the additive utility function follows directly from (5.61). ◀

Recall from Section 5.3 that in general, additivity follows from an assumption that

$$<(y_o,z_o),\ (y,z)> \sim <(y_o,z),\ (y,z_o)> \qquad (5.65)$$

for all y and z given some arbitrarily chosen y_o and z_o. Corollary 2 states that if we can assume Z is utility independent of Y, then additivity follows if we set $z=z_1$ and the above assumption holds for all y given the arbitrarily chosen y_o, z_o, and z_1. Earlier in Theorem 5.4, we proved that if mutual utility independence holds, then the additive utility function follows if $y=y_1$ and $z=z_1$ are both set and assumption (5.65) is valid for the single set of four values y_o, z_o, y_1, and z_1.

5.6.5. Usefulness of Certainty Equivalents

As before, a certainty equivalent $\hat{y}$ for $\tilde{y}$ in the lottery $(\tilde{y},z)$ is defined by the relation

$$u(\hat{y},z) = E[u(\tilde{y},z)].$$

When Z is utility independent of Y and when $\tilde{y}$ and $\tilde{z}$ are probabilistically independent, the expected utility of $(\tilde{y},\tilde{z})$ using (5.38) is

$$E[u(\tilde{y},\tilde{z})] = E[u(\tilde{y},z_o)][1-E[u(y_o,\tilde{z})]] + E[u(\tilde{y},z_1)]\ E[u(y_o,\tilde{z})]$$
$$= u(\hat{y}_o,z_o)\ [1-u(y_o,\hat{z})] + u(\hat{y}_1,z_1)\ u(y_o,\hat{z}),$$

where $\hat{y}_o$ and $\hat{y}_1$ are respectively the certainty equivalents for $\tilde{y}$ when $z = z_o$ and $z = z_1$, and $\hat{z}$ is the certainty equivalent for $\tilde{z}$.

The use of certainty equivalents for evaluating lotteries is discussed in more detail in Section 5.5. The basic explanation for their applicability is as follows. Utility independence allows us to express the expected utility of a lottery with more than one uncertain attribute in terms of the expected utilities of lotteries involving only one uncertain attribute. Probabilistic independence allows us to calculate expected utility over these latter lotteries by evaluating the expected utility over each component of the terms separately. Thus we have an expression for expected utility of the multiattribute lottery in terms of the expected utilities of one-variable lotteries. A certainty equivalent may then be substituted for the uncertain attribute in these simple lotteries, which should greatly facilitate interpretation of the implications of the lottery.

5.6.6. Utility Independence as an Approximation Technique

Even if neither attribute is utility independent of the other, the utility representation (5.38) which was derived using the assumption that just one of the attributes was utility independent of the other may provide a good approximation for the true utility function.

The basis for our argument is that (5.38) gives us five degrees of freedom in assessing $u(y,z)$, whereas the multilinear formulation of (5.16) gives us four degrees of freedom, and the additive formulation of (5.10) offers only three degrees of freedom in assessing $u(y,z)$. Consider the two-dimensional illustrations in Figure 5.12.

The degrees of freedom are shown on the figure as heavy lines or points. The two consequences marked "O" represent the consequences chosen to establish the origin and unit of measure of $u(y,z)$.

Using the additive form, we can then arbitrarily determine

(a) the shape of $u(\cdot,z_o)$, a conditional utility function for Y,

(b) the shape of $u(y_o,\cdot)$, a conditional utility function for Z,

(c) the unit of measure of $u(y_o,\cdot)$ by assessing $u(y_o,z_1)$.

These are the three degrees of the additive representation.

With the multilinear form, we have in addition to (a), (b), and (c), the freedom to fix

(d) the unit of measure of $u(\cdot,z_1)$, a conditional utility function for Y, by assessing $u(y_1,z_1)$.

Using (5.38), we can add to this list the freedom to evaluate

(e) the shape of $u(\cdot,z_1)$.

In Figure 5.13, we illustrate some of the general shapes of $u(y,z)$ which one can obtain using (5.38). The

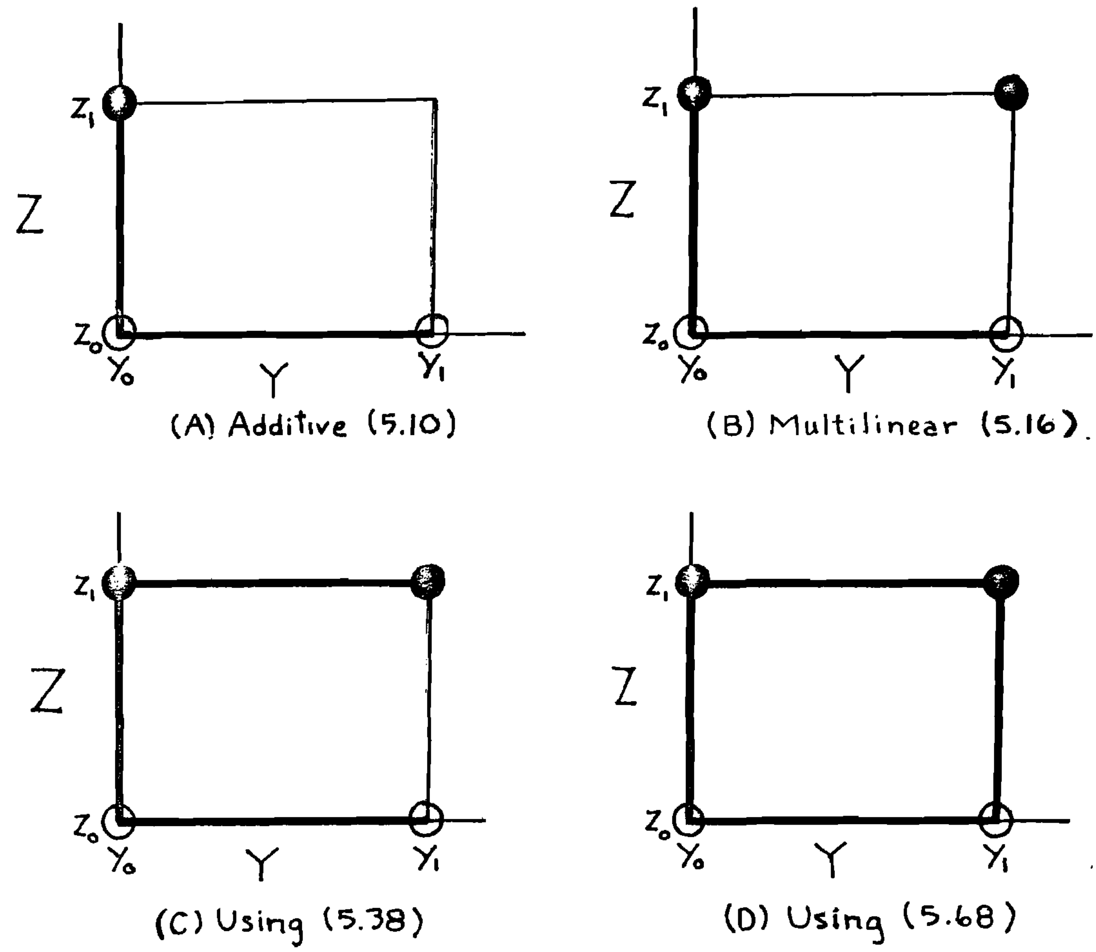


Figure 5.12. Assigning Utilities for Heavy Shaded Consequences Completely Specifies the Utility Function in the Cases Indicated

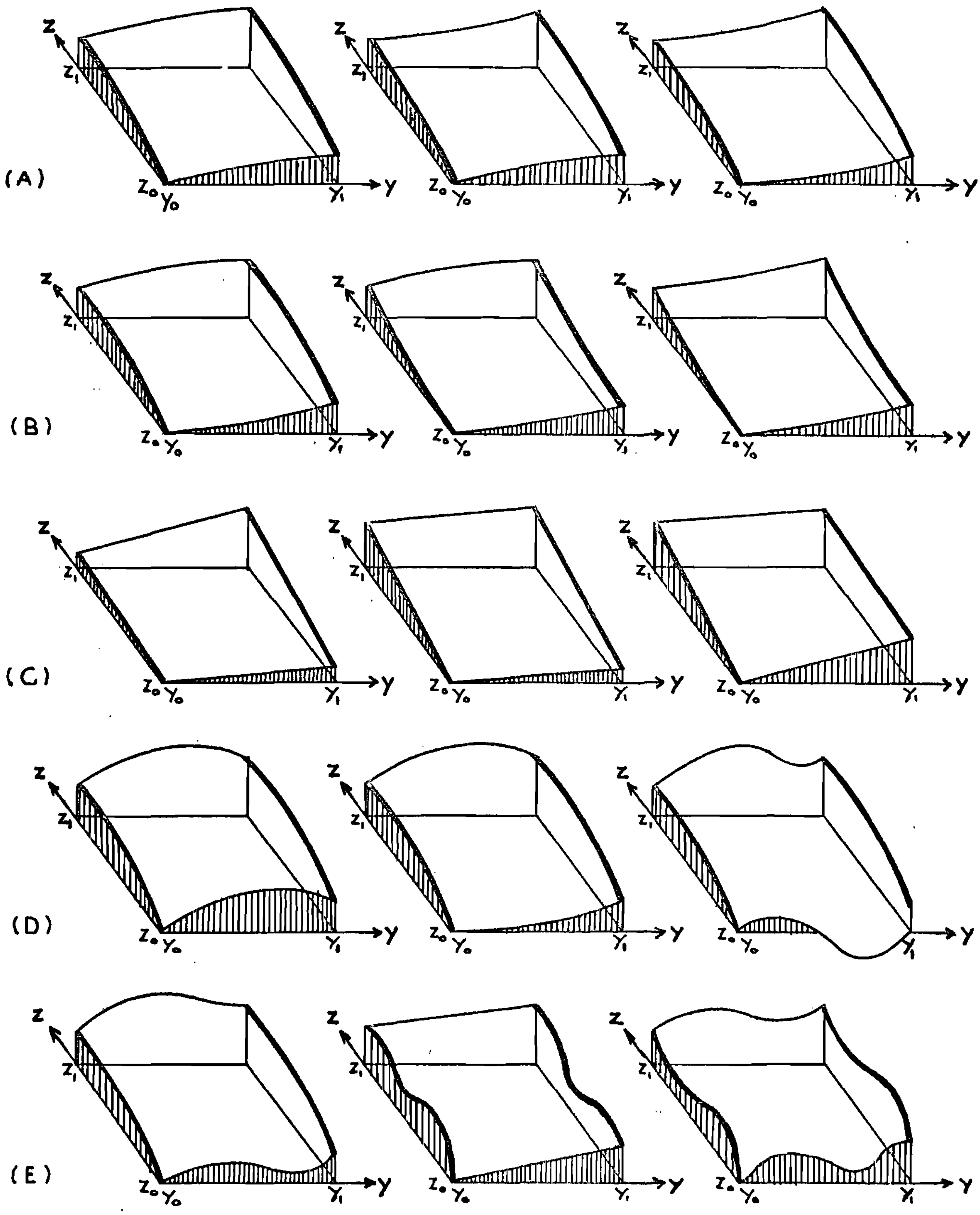


Figure 5.13. Different Shapes of Utility Functions $u(y,z)$ Where Z is Utility Independent of Y

common restriction on each utility function is that all the conditional utility functions over Z must be strategically equivalent. In each of the fifteen drawings, two such functions are darkened. Note however that the $u(y,\cdot)$ can have many shapes. Rows A and B in Figure 5.13 illustrate the effect of varying the shapes of $u(\cdot,z_o)$, $u(\cdot,z_1)$, and $u(y_o,\cdot)$. Various combinations of convex and concave conditional utility functions are shown.

With row C, we intend to illustrate the freedom created by selecting the units of $u(\cdot,z_o)$ and $u(\cdot,z_1)$. Finally, in rows D and E, we wish to point out that there are no restrictions, such as monotonicity or certain risk properties, on the conditional utility functions. To repeat the only restriction on the forms of $u(\cdot,\cdot)$ in Figure 5.13 is that $u(y,\cdot)$ has the same general shape (i.e., is strategically equivalent to) as $u(y_o,\cdot)$ for all values of y.

5.7. What To Do If No Independence Properties Hold*

Suppose we have ascertained, using assessment techniques discussed in the next section, that neither Y or Z is utility independent of the other. Then clearly, since mutual utility independence is a necessary condition for

---

*The reader may wish to omit this entire section, but we suggest that he at least quickly read the introduction of the section before proceeding to Section 5.8.

additive independence, none of the functional forms of two-attribute utility functions discussed in the preceding sections are strictly appropriate. Furthermore, suppose we have tried to implement the techniques discussed in Chapter 3 to reduce the dimensionality of the problem. These did not help either. However, we still want to quantify the decision maker's preferences. The question is, what can one do to obtain a reasonable u over $Y \times Z$ for decision making? Several possibilities exist including:

(A) transformation or adjustment of Y and Z to new attributes which might allow exploitation of utility independence properties,

(B) direct assessment of $u(y,z)$ by acquiring utilities of several consequences in the range of $Y \times Z$, and then using interpolation, extrapolation, and/or curve fitting,

(C) apply various of the results in preceding chapters over subsets of the $Y \times Z$ space, and then consistently scale them,

(D) develop or use existing more complicated assumptions about the decision maker's preference structure which imply more general utility functions.

Let us clarify ourselves on these options. The relative desirability of one approach versus another, of course, is very much a function of the problem at hand.

5.7.1. A Transformation of Attributes

It may be possible to select an alternate set of attributes and proceed to analyze the problem with this new set. Unfortunately, in this case the questions raised in Chapter 2 concerning the appropriateness of the set of attributes, such as completeness and measurability, must be reconsidered. Furthermore, it may make it necessary to repeat much of the analysis, including perhaps probabilistic assessments. To avoid this, perhaps the new attributes can be chosen to have some simple functional relationship to the original ones. Then, very little of the original analysis already completed will be worthless.

As a simple illustration, let Y and Z designate respectively measures of the crime rates in the two sections of a city. It may be that there is a complicated preference structure for (y,z) pairs. The relative ordering of lotteries for criminal activity in one section may depend very much for political reasons on the level of crime in the other section. However, suppose we define $S \equiv (Y + Z)/2$ and $T \equiv |Y - Z|$. Then S may be interpreted as some kind of an average crime index for the city and T is an indicator of the 'balance' of that activity between the two sections. Attributes S and T are functionally related to Y and Z. Given probability distributions over Y and Z, one could derive probability distributions over S and T. In addition, although there may be no simplifying preference assumptions in $Y \times Z$ space, such properties may exist in $S \times T$.

Example 5.2. Suppose that no utility independence properties exist among the original attributes Y and Z. Still it may be possible to define new attributes $S \equiv Y + Z$ and $T \equiv Y - Z$ which do possess independence properties. For instance, S and T might be additive independent with the form

$$u(s,t) = s^2 + t. \tag{5.66}$$

In this case the assessment of (5.66) should not be too difficult.

Notice that

$$\begin{aligned} u(y,z) \equiv u[s(y,z),\ t(y,z)] &= (y+z)^2 + (y-z) \\ &= y^2 + y + z^2 - z + 2yz, \end{aligned}$$

which illustrates that indeed no utility independence properties existed between Y and Z. ■

5.7.2. Direct Assessment of u(y,z)

This procedure is essentially that discussed in the subsection of Section 5.1 entitled "Use of $\underline{x}^*$ and $\underline{x}^o$." One picks as reference, two consequences and assigns utilities to these. Then using reference lotteries and empirical assessments of the decision maker, utilities are successively assigned to a number of consequences throughout $Y \times Z$. Utilizing a curve-fitting technique, a utility can be assigned to all possible consequences.

5.7.3. Employing Utility Independence Over Subsets of $Y \times Z$

The idea is simple--just subdivide the consequence space into parts such that various of the functional forms of preceding sections are appropriate. One needs to be careful to insure consistent scaling on $u(y,z)$.

Example 5.3. Suppose we are interested in assessing $u(y,z)$, $y' \le y \le y''$ and $z' \le z \le z''$ where preferences are increasing in both attributes. For $y \le y_o$, Z is utility independent of Y, so from (5.42), if we set $u_1(y_o,z') = 0$ and $u_1(y_o,z'') = 1$, then

$$u_1(y,z) = u_1(y,z')[1 - u_1(y_o,z)] + u_1(y,z'')\ u_1(y_o,z),$$
$$y \le y_o,\ z' \le z \le z''.$$

For the rest of the original region, suppose Y is utility independent of Z, so if we set $u_2(y_o,z') = 0$ and $u_2(y'',z') = 1$, then

$$u_2(y,z) = u_2(y_o,z)[1 - u_2(y,z')] + u(y'',z)\ u\ (y,z'),$$
$$y \ge y_o,\ z' \le z \le z''.$$

Since both u_1 and u_2 have the same origin, then in order to consistently scale u_1 and u_2 we need only determine a scaling constant λ defined by

$$\lambda = \frac{u_2(y_o,z'')}{u_1(y_o,z'')} \quad .$$

In this case a consistent utility function for all $Y \times Z$ is

$$u(y,z) = \begin{cases} \lambda u_1(y,z), & y \leq y_0,\ z' \leq z \leq z''. \\ u_2(y,z), & y \geq y_0,\ z' \leq z \leq z''. \end{cases} \blacksquare$$

5.7.4. Weaker Assumptions on the Preference Structure*

This subsection is meant to indicate a couple of more general models than those of the previous sections. As is evident and expected, the requisite assumptions for these models are more complex than those used earlier. One could obviously develop even more general models than those in this section. The advantage is clear. Such models are more likely to be appropriate for a specific decision maker's preference structure and, therefore, less likely to misrepresent it. The disadvantage is operational. It is more difficult to verify the assumptions of the more general models and then more difficult to assess $u(y,z)$ once they are verified. This tradeoff must inevitably be considered in selecting a model for one's utility function.

REVERSING PREFERENCES. If Z is utility independent of Y, then

$$u(y,z) = c_1(y) + c_2(y)\ u(y_0,z), \tag{5.67}$$

---

*In this subsection, we will be quite informal. The purpose is (1) to communicate a flavor for some generalizations of material presented earlier in this chapter which have been developed, and (2) to indicate sources for this work.

where $c_2(y)$ must be greater than zero. This implies that the preference order over lotteries on Z will always be the same regardless of the amount y. Suppose one allows $c_2(y)$ to also be negative or zero. Then if $c_2(y') < 0$, the preference order on lotteries over Z given y' is exactly reversed from this order given y_o. If $c_2(y') = 0$, then one is indifferent between all lotteries Z given y'. Fishburn [1974] allowed for these reversals of preference and indifference and derived results analogous to those in Section 5.4.

A GENERALIZATION OF UTILITY INDEPENDENCE. The most general result we have discussed so far is (5.38) which require two one-attribute utility functions over Y and one one-attribute utility function over Z. The question arises as to what type of functional form might be developed using two one-attribute utility functions over each of Y and Z, and what would the associated requisite assumptions on the decision maker's preference structure be? Fishburn [1974] has developed necessary and sufficient conditions for determining u(y,z) by assessing adequately scaled utility functions over the heavy lines of Figure 5.12D. The result is that

$$u(y,z) = u_Y(y) + u_Z(z) + f_Y(y)\ f_Z(z). \tag{5.68}$$

The requisite assumptions and proof for (5.68) along with a discussion of scaling the functions u_Y, u_Z, f_Y, and f_Z is found in Fishburn [1974b].

PARAMETRIC DEPENDENCE. As indicated in Section 5.2, if Z is utility independent of Y, then one's attitude toward risk in terms of lotteries over Z is independent of Y. Kirkwood [1973] developed parametric dependence, which eliminates this restriction, but requires the preferences over Z for different amounts of Y to be representable by members of the same parametric family of utility functions. For instance, if preferences over Y are increasing and constantly risk averse for all z, but the degree of risk aversion varies, we have

$$u(y,z) \sim -e^{-y\theta(z)}, \quad \theta(z) > 0 \quad . \tag{5.69}$$

Equation (5.69) indicates that all conditional utility functions over Y are dependent on z through the parameter $\theta(z)$. In this case we would say that Y is parametrically dependent on Z. More formally, we will say that Y is parametrically dependent on Z if the conditional utility functions over Y given different levels of z depend on z only through a parameter θ. This means that

$$u(y,z) = d_1(z) + d_2(z)u_{Y/z}[y/\theta(z)] \quad , \tag{5.70}$$

where $d_2(z) > 0$ and $u_{Y/z}$ indicates a conditional utility function over Y given z.

To illustrate the use of parametric dependence and to provide an intuitive flavor, consider

Theorem 5.10. If Y is parametrically dependent on Z, then $u(\cdot,\cdot)$ is completely determined by three consistently scaled utility functions on Z given levels of y and one utility function on Y given z.

Rather than a formal proof, refer to Figure 5.14 for the basis of an informal one. Theorem 5.10 says that subject to the stated conditions, the utility of any point can be assigned given the consistently scaled utilities of the darkened lines. From $u(\cdot,z_0)$ we know the functional form of the utility function $u(\cdot,z)$ for all z. To determine the value of the parameter for a particular z, we just use the utilities of (y_0,z), (y',z), and (y_1,z). Then $u(\cdot,z)$ is scaled by $u(y_0,z)$ and $u(y_1,z)$, which allows us to assign a utility to any (y,z).

Obviously the parametric dependence concept could be extended to include families of utility functions involving two parameters rather than one. Then it would not be difficult to derive results analogous to Theorem 5.10. For instance, the only change in Theorem 5.10 would be that four conditional utility functions over Z, one more than before, would need to be assessed. Similarly, results making use of both parametric dependence and utility independence can be derived. Kirkwood [1972] presents some of these.

<u>SUMMARY STATE DESCRIPTORS</u>. Let us terminate this section with one further generalization which will be elaborated on in Chapter 9. Consider the two attributes Y and Z but now assume that Z is multidimensional. In some circumstances the conditional utility function $u(\cdot,z)$ on Y

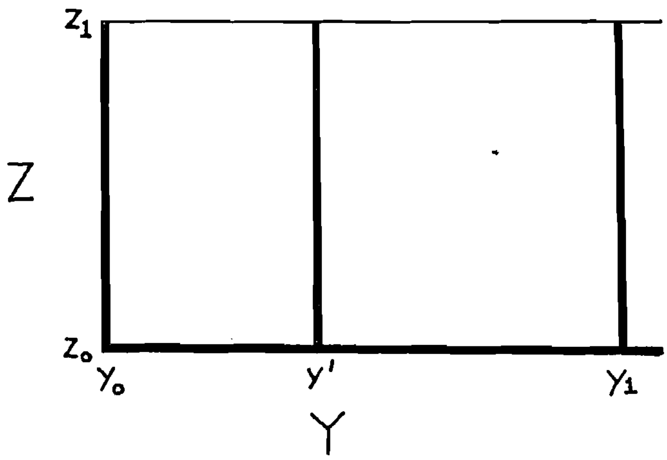


Figure 5.14. When Y is Parametrically Dependent on Z, the Utility Function u(y,z) is Completely Specified by the Utilities Over the Heavy Shaded Consequences

might depend on (the multidimensional) z only through some summary state description, say $\Theta(z)$, of z. In some cases the range of Θ might be unidimensional. For example, suppose we are concerned with time streams of consumption. The utility of future consumption starting from a point in time t_o might depend on past and present consumption. But, as an approximation, we might be able to assume that the utility for future consumption depends only on the past through the present consumption at t_o. Hence the consumption stream up to and including time t_o can be effectively summarized by the state description: consumption at t_o. This example is a natural analogy of Markovian probabilistic dependence, and other weak forms of probabilistic dependence have their analogies in the utility domain. In other words, if we cannot assume as reasonable various utility independence notions, then just as in conventional probabilistic analysis, one can introduce weak forms of utility dependence. As far as we know this research direction has hardly been scratched.

As indicated at the beginning of this subsection, with the greater generality of the models comes the greater complexity of utilizing them. For many problems, the simpler models likely are "good enough" approximations even if they are not precisely valid. However, for those problems where this is not the case, it is important to realize how to add generality to the model and still keep the assessment task within bounds.

5.8. Assessment Procedure for Multiattribute Utility Functions

After reading the unidimensional case in Chapter 4, it should come as no surprise to the reader that we feel that one cannot identify a series of steps which, when followed, will result in a properly assessed multiattribute utility function. Just as before, the process requires a good deal of foresight and improvisation. Before assessing any preferences or utilities, we assume that the analyst (or interrogator) has properly set the stage for the decision maker or his delegated expert. In particular, we assume that the respondent realizes the purpose of the exercise and is sufficiently motivated to think hard about his feelings for the various consequences.

It is at this point that we begin to assess his utility function. As with the one-attribute case, the assessment procedure can be segmented for discussion purposes to highlight various aspects which must be completed. Although our discussion will focus on two-attribute utility functions, the basic ideas are relevant to all multiattribute utility assessments. The sequence one might follow in determining a utility function can be described in five stages:

(1) Introducing the terminology and ideas,

(2) Identifying relevant independence assumptions,

(3) Assessing conditional utility functions or iso-preference curves,

(4) Assessing the scaling constants,

(5) Checking for consistency and reiterating.

5.8.1. Introducing the Terminology and Ideas*

Suppose we have structured the decision problem and specified two attributes Y and Z, which are adequate to describe the consequences. Then we must assess a utility function over all possible (y,z) consequences. A consequence space should be illustrated‡ as in Figure 5.15 as a graphical aid to the decision maker.

Before any assessments are made, it should be clear to the decision maker that the preferences we are interested in are his. It must be understood that there are no objectively correct preferences, that the preferences

---

*In Section 4.9, an assessment procedure for assessing single-attribute utility functions was discussed. The preliminaries to assessment were essentially the same as those discussed in this subsection, since the purpose in both cases is to make sure the decision maker understands the process and its motivation. The basic ideas are included here to render this section a complete unit.

‡The figures and examples in the text in this section concern scalar attributes in order to simplify the presentation. All of the suggestions do generalize for vector attributes, although then clearly the problem becomes more involved.

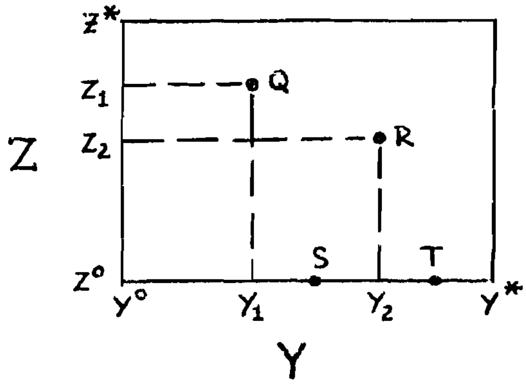


Figure 5.15. A Two-Attribute Consequence Space

of importance represent the subjective feelings of the decision maker. At any time if the decision maker feels uncomfortable with any of the information he has offered about his subjective feelings, it is perfectly all right, in fact, necessary for a correct analysis, for him to change his mind. This is one of the purposes of a utility analysis namely: to help the decision maker think hard about his preferences and hopefully straighten them out in his mind.

Now, the analyst--let's assume this role for ourselves --must make sure that the decision maker understands the consequence space representation of Figure 5.15. We might explain that by consequence Q we mean the consequence where $y = y_1$ and $z = z_1$. Then we might ask him what is meant by consequence R. The answer, of course, is $y = y_2$ and $z = z_2$. The decision maker should realize the directions in which y and z increase in Figure 5.15.

Next, it is helpful to limit the region over which we must assess preferences to as small a region as possible. From our earlier involvement in structuring the problem with the decision maker, we should already know the maximum and minimum amounts which both y and z could assume. Then we would choose a y^o, y^*, z^o, and z^* such that for all possible (y,z),

$$y^o \leq y \leq y^* \text{ and } z^o \leq z \leq z^*.$$

The values should be chosen for convenience and meaningfulness to the decision maker. For instance, if y ranged from

0 to 8.75 in the specific units, we might define $y^o = 0$ and $y^* = 10$. A value of $y^* = 10{,}000$, for example, probably would have little meaning to the decision maker. The preferences which we eventually assess must only be those for consequences (y,z) with $y^o \leq y \leq y^*$ and $z^o \leq z \leq z^*$. This is the region shown in Figure 5.15.

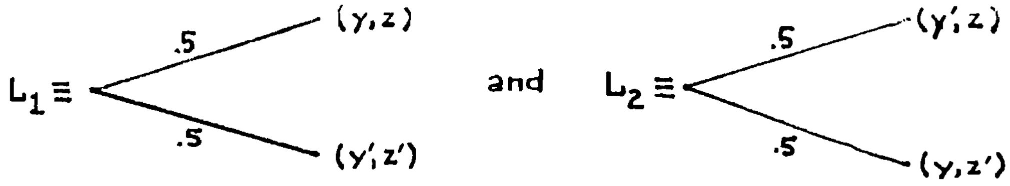
As a final check on the decision maker's understanding of the consequence space representation, we might ask him whether or not he prefers consequence T to consequence S in Figure 5.15. The points S and T should be chosen such that it is clear to us, the analyst, that the decision maker would almost for sure prefer a particular one. If the decision maker's preference in this case agreed with the expected result, we could proceed to more difficult questions. If not, the decision maker's reasoning should be pursued, and perhaps the familiarization process should be repeated, in part or in full.

Enough has been said about the preliminaries. The basic idea is to acquaint the decision maker with the framework which we use in assessing his utility function.

5.8.2. Verification of Independence Assumptions

Here we will discuss procedures to verify if Y and Z are additive independent and if either attribute is utility independent of the other.

ADDITIVE INDEPENDENCE. Suppose we wish to assess preferences over the consequence space $y^o \leq y \leq y^*$ and $z^o \leq z \leq z^*$ as shown in Figure 5.16. As defined in Section 5.3, Y and Z are additive independent if and only if the lotteries

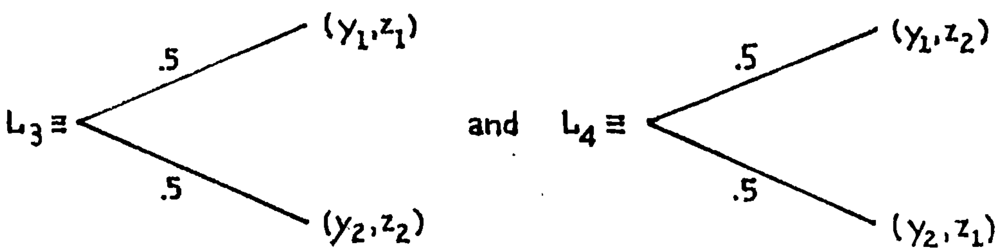


are indifferent for all amounts of y,z given a specific y',z'. So the obvious method to verify additive independence is to select a y' and z' and see if indifference between L_1 and L_2 holds for some (y,z) pairs.

Practically speaking, if Y and Z are divided into four equal subsections by $\{y^o, y^{.25}, y^{.5}, y^{.75}, y^*\}$ and $\{z^o, z^{.25}, z^{.5}, z^{.75}, z^*\}$ respectively as indicated in Figure 5.16 and if L_1 is indifferent to L_2 for each possible (y,z) pair taken from these two groups of five, then it seems justified to assume Y and Z are additive independent.

An alternate procedure to check for additive independence involves first trying to verify that Y and Z are mutually independent.

Recall that mutual utility independence is a necessary, but not a sufficient condition for additive independence. If Y and Z are mutually utility independent, they are additive independent if there exists a y_1, y_2, z_1, and z_2 such that



are equally desirable, where neither (y_1,z_2) or (y_2,z_1) are indifferent to (y_1,z_1). On the other hand, if there

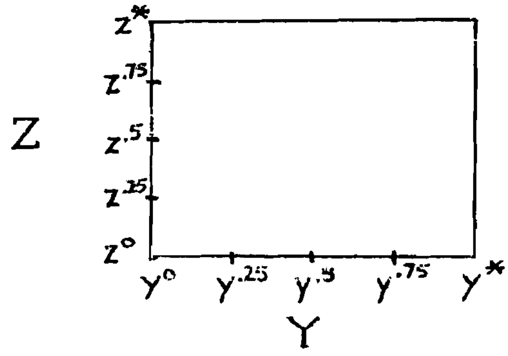


Figure 5.16. A Graphical Aid for Verifying Additive Independence Conditions

exists any L_3 and L_4 such that they are not indifferent, then clearly additive independence cannot hold.

<u>UTILITY INDEPENDENCE.</u> Again, suppose we have two scalar attributes, Y and Z, and wish to assess preferences over $y^o \leq y \leq y^*$ and $z^o \leq z \leq z^*$ as shown in Figure 5.17. The letters P,Q,R,S, etc., designate consequences referred to in the discussion.

To verify whether Y is utility independent of Z, begin by asking the decision maker if he prefers <P,Q>, a lottery yielding either P or Q with equal probability, or S. The consequence S is chosen so a particular answer is expected. Suppose the decision maker prefers <P,Q> to S and this agrees with our expectations. Then we ask the decision maker whether he prefers <P,Q> or T, where T is chosen such that we expect T is preferred to <P,Q>. Next, we inquire about the preferences of <P,Q> relative to W. Since W is "near to" S, we somewhat expect that <P,Q> will be preferred to W, but perhaps not. We continue with this convergence procedure until we reach a consequence R such that <P,Q> and R are equally desirable (or undesirable) to the decision maker.

If the decision maker indicates any preferences which we do not feel are consistent with his "true" preferences, this should be pointed out and discussed again.

Notice from Figure 5.17 that consequences P,Q,R,S,T, and W all had a common amount of Z and only differed in their amount of Y. Thus, R is a certainty equivalent for the lottery <P,Q>.

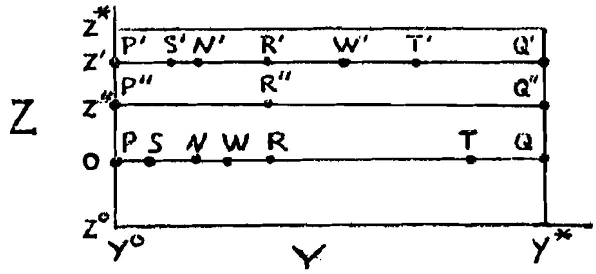


Figure 5.17. A Graphical Aid for Verifying Whether Y is Utility Independent of Z

Now we move our concentration to the set of consequences with a different amount of Z in common, say z', and ask a similar set of questions. First we want to determine whether or not the decision maker prefers T' to $<P',Q'>$. To avoid a repetition of the previous answers without thinking about the current questions, T' should be chosen such that the amount of Y, and not only the amount of Z, in T and T' are different. Suppose the decision maker prefers T' to $<P',Q'>$. Then we ask about his preferences between $<P',Q'>$ and S', between $<P',Q'>$ and W', and eventually determine R' such that $<P',Q'>$ is indifferent to R' for the decision maker. If R' and R have a common amount of Y (i.e., R' lies directly above R in Figure 5.17), then we begin to think Y might be utility independent of Z. What we have so far determined is that the relative preferences of P,Q, and R and the relative preferences of P',Q', and R' are the same.

Again we repeat the procedure for another amount of Z, say z'', and hopefully find that R'', a certainty equivalent for $<P'',Q''>$, has a common amount of Y with R and R'. If this is the case, we can feel reasonably confident that Y is utility independent of Z. The assumption can be further checked following the same procedure to determine a consequence N indifferent to $<P,R>$ and a consequence N' indifferent to $<P', R'>$, for example. If Y is utility independent of Z, we would expect N and N' to have a common amount of Y.

Finally, we ask the decision maker in general if $<(y^*,z'), (y^o,z')>$ is indifferent to (y',z') and if $<(y^*,z''), (y^o,z'')>$ is indifferent to (y',z''), then can we conclude that $<(y^*,z), (y^o,z)>$ is indifferent to (y',z) for all possible values of Z? If Y is utility independent of Z, the answer of course, must be yes. As a last question on this point, we ask if for any arbitrary y_1, y_2, and y_3, if $<(y_1,z), (y_2,z)>$ is indifferent to (y_3,z) for one particular value of z, will the same relation hold for all possible values of Z? A yes answer to this definitely implies that Y is utility independent of Z.

Another way of verifying that Y is utility independent of Z might go as follows.

"Consider a 50-50 lottery between y_1 and y_2 for a fixed level of z, say $z = z_1$", our analyst asks the decision maker. "Look at the figure on the page in front of you. Now think hard about what y amount you would want for certain, always keeping z_1 fixed, so that you are indifferent between the certainty amount and the 50-50 lottery. Reflect on this problem for a while"..."O.K." the analyst continues, "Now when you were thinking about your breakeven y, was it important to you to keep in mind the level of z. Suppose we let $z = z_2$ instead of z_1, would it have made any difference?".

Now if the answer were "No, it would not", then the analyst should check with the respondent whether this could be assumed to be generally the case if y_1 and y_2 were changed and if z_1 and z_2 were changed. If this is verified, then we could assume Y is utility independent of Z.

If in our original set of questions, we found R and R' did not have a common amount of Y, then assuming R and R' correctly represented the decision maker's preferences, Y could not be utility independent of Z. However, since utility independence is not reflexive, Z may yet be utility independent of Y. Even if this is not so, if the amounts of Y in R and R' are "reasonably" close to each other, we might approximate the true utility function by assuming Y to be utility independent of Z and assessing a utility function accordingly.

Suppose we label the certainty equivalent of the lottery $\langle (y',z),\ (y'',z) \rangle$ by $(\hat{y}_z, z)$. Often, in practice, the decision maker might feel there is a slight dependence of $\hat{y}_z$ on z but it might be a convenient "lie" to set $\hat{y}_z$ equal to a fixed value for all z <u>provided</u> that the <u>relevant</u> range of z is small. And therefore in practice it is often crucially important to be able to restrict the range set of an attribute such as Z. One way of achieving this restriction is by eliminating acts that are dominated or "practically" dominated by others. By restricting the domain of Z, one can make an idealized abstraction--or lie--such as, <u>Y is utility independent of Z</u> more palatable. This issue was discussed in subsection 5.6.6.

5.8.3. Assessing Conditional Utility Functions

The conditional utility functions $u_Y(\cdot)$ over Y and $u_Z(\cdot)$ over Z may be either multidimensional or unidimensional. That is, the arguments y and z respectively may be vectors or scalars. If they are vectors, hopefully we can further decompose the utility function using the independence properties discussed in this and the next chapter in order to decrease the dimensionality of utility functions which must be directly assessed. If this is not possible, then some of the ideas of Sections 5.1 or 5.7 must be utilized.

On the other hand if the conditional utility functions are unidimensional, then the procedures discussed in Chapter 4 are appropriate. If this is the case and if the previously suggested procedure to verify utility independence was used, then one already has a number of certainty equivalents, which are appropriate in assessing the conditional utility functions. Obviously, this information, and any other obtained in verifying independence assumptions, should be utilized wherever possible.

5.8.4. Assessing the Scaling Constants

In all the models in this chapter, the form of the utility function u(y,z) has been specified in terms of a number of conditional utility functions over either Y or Z and scaling constants. For example, with the multilinear utility function discussed in Section 5.4,

$$u(y,z) = k_Y u_Y(y) + k_Z u_Z(z) + k_{YZ} u_Y(y) u_Z(z) \quad , \qquad (5.71)$$

there is one conditional utility function for each of Y and Z and three scaling constants: k_Y, k_Z, and k_{YZ}. Both of the utility functions in (5.71) can be scaled from zero to one since the scaling constants are used to insure internal consistency.

The basic idea for evaluating the three scaling constants is to obtain a set of three independent equations with three unknowns, which are then solved to obtain the k's. These equations can be generated from certainty considerations, probabilistic considerations, or a combination of both. For example, using certainty scaling, if consequences (y_1,z_1) and (y_2,z_2) are indifferent, then equating these utilities using (5.71), we have

$$k_Y u_Y(y_1) + k_Z u_Z(z_1) + k_{YZ} u_Y(y_1) u_Z(z_1) = k_Y u_Y(y_2) + k_Z u_Z(z_2) + k_{YZ} u_Y(y_2) u_Z(z_2) \quad . \tag{5.72}$$

Both u_Y and u_Z are known, since we are assuming they have already been assessed, so (5.72) is one equation with at most three unknowns. Using probabilistic scaling, suppose (y_3,z_3) is indifferent to the lottery $<(y_1,z_1);\ p;\ (y_2,z_2)>$ which yields (y_1,z_1) with probability p and (y_2,z_2) with the complementary probability 1-p. Then, equating expected utilities

$$u(y_3,z_3) = pu(y_1,z_1) + (1-p)\ u(y_2,z_2), \tag{5.73}$$

which when combined with (5.71) yields an equation involving

k_Y, k_Z, and k_{YZ} as the only unknowns. Clearly, using either or both certainty and probabilistic scaling, one can generate three independent equations with the three k's as unknowns. Let us illustrate this.

Consider again the multilinear utility function (5.71) where the origins of u, u_Y, and u_Z are

$$u(y^o, z^o) = 0, \; u_Y(y^o) = 0, \text{ and } u_Z(z^o) = 0. \tag{5.74}$$

And suppose we are interested in preferences over the consequence space where $y^o \leq y \leq y^*$ and $z^o \leq z \leq z^*$. For illustration purposes, let us further assume that preferences are increasing in both Y and Z so the utility functions can be scaled by

$$u(y^*, z^*) = 1, u_Y(y^*) = 1, \text{ and } u_Z(z^*) = 1. \tag{5.75}$$

Using (5.75) to evaluate (5.71) at (y^*, z^*), one finds that for consistency

$$1 = k_Y + k_Z + k_{YZ}. \tag{5.76}$$

Furthermore evaluating (5.71) at (y^*, z^o) and (y^o, z^*) respectively gives us

$$u(y^*, z^o) = k_Y \text{ and } u(y^o, z^*) = k_Z. \tag{5.77}$$

As a starter, let us first try to see which is larger: k_Y or k_Z? This could be done, for instance, by asking the decision maker if he preferred (y^*, z^o) or (y^o, z^*). From (5.77), if the former is preferred, then $k_Y > k_Z$; if the latter is preferred, then $k_Z > k_Y$; and if they are indifferent, $k_Y = k_Z$. Suppose we find $k_Y > k_Z$. Then we can look for an amount y' such that decision maker is indifferent between (y', z^o) and (y^o, z^*). Equating their utilities using (5.71) yields

$$k_Z = k_Y u_Y(y'), \tag{5.78}$$

where $u_Y(y')$ is known. To help the decision maker identify y', we might present him with a specific choice between (y, z^o) and (y^o, z^*) with y fixed. If the first consequence were preferred to the second, y would be decreased and the binary choice reoffered; if the second consequence were preferred, y would be increased and the procedure repeated. With this approach, one should soon converge to y'. Equation (5.78) is the result of certainty scaling.

For an example of probabilistic scaling, suppose using techniques discussed in Chapter 4 we assessed the indifference probability π_Y such that (y^*, z^o) is indifferent to $\langle (y^*, z^*);\ \pi_Y;\ (y^o, z^o) \rangle$. Using (5.71) and equating expected utilities, we find

$$k_Y = \pi_Y. \tag{5.79}$$

The system of equations (5.76), (5.78), and (5.79) has three unknowns, which can be solved for k_Y, k_Z, and k_{YZ}. For $\pi_Y = 0.5$ and $u_Y(y') = 0.8$, one easily concludes that

$$k_Y = 0.5,\ k_Z = 0.4, \text{ and } k_{YZ} = 0.1. \tag{5.80}$$

Let us generalize the ideas in this subsection. All our two-attribute formulations in this chapter express u(y,z) in terms of conditional utility functions over the individual attributes and scaling constants. Thus, if there are N and M conditional utility functions over Y and Z respectively and if there are R scaling constants, we can write

$$u(y,z) = f[u_Y^1(y),\ldots,u_Y^N(y),u_Z^1(z),\ldots,u_Z^M(z),k_1,k_2,\ldots,k_R]. \tag{5.81}$$

where f is specified. The utility functions in (5.81) can all be scaled zero to one since the scaling constants provide overall consistency.

Thus to evaluate the R scaling constants, we must generate R independent equations and solve. As illustrated, each equation can be generated from certainty considerations or probabilistic considerations.

One operational problem of concern is how to generate independent equations, or said another way, how is redundancy in the equations avoided? In practice, one's understanding of the problem and knowledge of the functional form of the utility function are probably the best guards against a large number of redundant equations. When a redundancy does occur, we need to empirically generate another equation that is not redundant to substitute for a redundant one. To illustrate this, let us return to the proceeding example.

Suppose that after (5.78) had been determined, we assessed a y'' and z'' such that (y'',z^o) and (y^o,z'') were equally desirable. Then equating expected utilities,

$$k_Y u_Y(y'') = k_Z u_Z(z''). \tag{5.82}$$

Clearly (5.76), (5.78), and (5.82) are three equations with three unknowns, but (5.78) and (5.82) are not independent of each other. Both are concerned with setting the scaling between Y and Z. To get around this, we can obviously use probabilistic scaling as we did in the original example. If one preferred to use certainty scaling, one could assess a y'' such that (y'',z^*) is indifferent to (y^*,z^o). Then of course,

$$k_Y = k_Y u_Y(y'') + k_Z + k_{YZ} u_Y(y''), \tag{5.83}$$

where $u_Y(y'')$ is known. Now (5.83) is independent of both (5.78) and (5.82). So, for example, equations (5.76), (5.78), and (5.83) can be solved for k_Y, k_Z, and k_{YZ}, and (5.82) can be used to check the consistency of the resulting $u(y,z)$.

5.8.5. Checking for Consistency and Reiterating

There are many different consistency checks which can be used to detect errors in the decision maker's utility function. By an error, we mean that the utility function which we have assessed for him does not represent his preferences when it is tested by hypothetical examples. Three such consistency checks are suggested in this section.

With these as a guide, the decision analyst should have no trouble developing other checks designed to uncover discrepancies in the utility function.

One method to check the validity of a utility function involves paired comparisons of various consequences. Concerning a utility function u(y,z), we might ask the decision maker if he prefers (y_1,z_1) to (y_2,z_2). If so, then $u(y_1,z_1)$ should be greater than $u(y_2,z_2)$ to be consistent. This type of check can be repeated as many times as it is felt useful. It might be wise to start with some easy comparisons and work up to more difficult ones. This acquaints the decision maker with the technique before really pressing his judgment for difficult choices among consequences.

A more systematic way of doing this would be to use the u function on Y x Z to generate a family of indifference curves in the Y x Z plane--we are assuming here, of course, that Y and Z are each unidimensional. Then the decision maker could reflect whether these indifference curves seem reasonable to him.

Another check on the utility function is to empirically determine whether or not he is risk averse on positive rays of the form (y,cy) where c > 0. We might ask the decision maker what consequence (y_1,cy_1) is indifferent to $\langle(y_2,cy_2);\ 1/2;\ (y_3,cy_3)\rangle$. For the case where u(y,cy) is increasing in y, if y_1 is less than $(y_2 + y_3)/2$, we might expect he is ray risk averse. From Section 4.4, we have the theory how to determine if the decision maker is risk averse on this and any other positive rays. If we do decide he is ray risk averse, for the case where preferences are increasing in Y and Z, then using the theory in Section 4.5, it should be clear that to be consistent $u'(y,cy)$ must be positive and $u''(y,cy)$ negative for all y, where u' and u'' denote first and second derivatives with respect to Y. If he is not ray risk averse, then obviously $u(y,z)$ should not possess ray risk aversion.

In cases where the utility function is of a special form, a particular consistency check may be applicable. For example, if $u(y,z)$ is of the multilinear form (5.16) of our preceding example, we can choose any y_1, y_2, z_1, and z_2 such that

$$u(y_2,z_o) > u(y_1,z_o), \tag{5.84a}$$

and

$$u(y_o,z_2) > u(y_o,z_1), \tag{5.84b}$$

and check the sign of parameter k_{YZ} in the following manner. We ask the decision maker whether he prefers $L_1 \equiv \langle (y_2,z_2), (y_1,z_1) \rangle$ or $L_2 \equiv \langle (y_2,z_1), (y_1,z_2) \rangle$. If L_1 is preferred to L_2, then k_{YZ} must be positive. If the lotteries are indifferent, k_{YZ} must be zero, and if L_2 is preferred to L_1 then k_{YZ} should be negative. Also, if L_1 is preferred to L_2 for any set of y_1, y_2, z_1, and z_2 satisfying (5.84), it must hold for all such sets satisfying (5.84). More is said concerning this point in Section 5.4.

In practice, the imagination of the analyst will not be strained in an attempt to develop efficient and effective consistency checks. And as has been mentioned before, if the consistency checks produce discrepancies with the previous preferences indicated by the decision maker, these discrepancies must be called to his attention and parts of the assessment procedure should be repeated to acquire consistent preferences. Once we obtain a utility function which the decision maker and the analyst feel represents the true preferences of the decision maker, we may proceed with the analysis. Of course, if the respondent has strong, crisp, unalterable views on all questions and if these are inconsistent, then we would be in a mess, wouldn't we? In practice, however, the respondent usually feels fuzzier about some of his answers than others, and it is this degree of fuzziness that usually makes a world of difference. For it then becomes usually possible to generate a final coherent set of responses which does not violently contradict any strongly held feelings. Finally, if the decision maker and his analyst remain a bit squeamish about accepting any "compromise" utility function, then they can always embark on a sensitivity analysis.

5.9. Interpreting the Scaling Constants

It is not easy to interpret the scaling constants since they depend on the choices of y^o, y^*, z^o, and z^* which, in turn, depend on the possible consequences of the problem. Let us illustrate our discussion with the additive function

$$u(y,z) = k_Y u_Y(y) + k_Z u_Z(z), \qquad (5.85)$$

where

$$u(y^o,z^o) = 0,\ u_Y(y^o) = 0,\ u_Z(z^o) = 0, \qquad (5.86)$$

and

$$u(y^*,z^*) = 1,\ u_Y(y^*) = 1,\ u_Z(z^*) = 1. \qquad (5.87)$$

Then clearly, for consistency

$$k_Y + k_Z = 1. \qquad (5.88)$$

If the assessment of $k_Y = 0.75$ and $k_Z = 0.25$, one cannot say that Y is three times as important as Z. In fact, one cannot conclude that attribute Y is more important than Z. Going one step further, it is not clear how one would precisely define the notion that one attribute is more important than another. We can say, however, that if starting from the point (y^o,z^o), we would rather change z^o to z^* than change y^o to y^*, then $k_Z > k_Y$, and vice-versa. If, for instance, y^o and y^* are "close together," that is, the range of Y is relatively small, then k_Y may be small but still the Y attribute may be mighty important. Changing the range of y^o to y^* will necessarily change the value of k_Y. Because of the consistency condition (5.88), the scaling constant k_Z will also change. To better illustrate the point, suppose in comparing jobs, attribute Y refers to monetary rewards and all jobs under consideration pay almost the same amount so y^o and y^* are close together; then k_Y may be small but this does not mean that money is unimportant to the decision maker. As y^o and y^* become closer and closer, the value of k_Y approaches zero. Clearly when using an additive utility function in such situations, the pay would have little influence on the final choice of a job, but one can still not conclude money is not important.

To graphically illustrate the point, consider Figure 5.18 which exhibits two consequence spaces*, both of which could be used to evaluate the same problem provided all possible consequences fell within the smaller of the two. Furthermore, suppose the additive utility function is appropriate. If the utility function $u(y,z)$ scaled from zero to one is assessed over (y,z) for $y^o \leq y \leq y^*$ and $z^o \leq z \leq z^*$, we might find, for instance,

$$u(y,z) = k_Y u_Y(y) + k_Z u_Z(z) \tag{5.89}$$

where $k_Y = 0.75$, $k_Z = 0.25$, and u_Y and u_Z are also both scaled zero to one. And since $u_Z(z^+)$ must fall between 0 and 1, let us assume it is 1/3. So from (5.89), note that

$$u(y^o,z^+) = u(y^*,z^o) = 0.25. \tag{5.90}$$

---

*For simplicity, we'll assume preferences are increasing in both Y and Z.

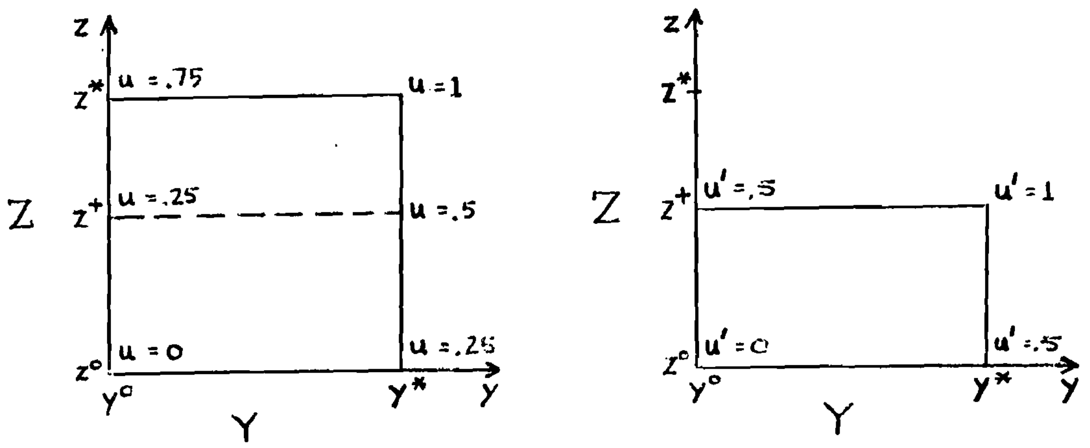


Figure 5.18. An Illustration That Scaling Constants Cannot Be Interpreted As a Indicator of Attribute Importance

Now suppose that we had originally decided to assess preferences over (y,z), where $y^o \leq y \leq y^*$ and $z^o \leq z \leq z^+$ using $u'(y,z)$. The additive utility function

$$u'(y,z) = k_Y' u_Y'(y) + k_Z' u_Z'(z), \tag{5.91}$$

where each of the utility functions can be scaled by

$$u'(y^o,z^o) = 0, \; u_Y'(y^o) = 0, \; u_Z'(z^o) = 0 \tag{5.92}$$

and

$$u'(y^*,z^+) = 1, \; u_Y'(y^*) = 1, \; u_Z'(z^+) = 1, \tag{5.93}$$

then holds. For consistency, clearly

$$1 = k_Y' + k_Z' \tag{5.94}$$

and since (y^o,z^+) and (y^*,z^o) are indifferent from (5.90), utilities can be equated using (5.91), (5.92), and (5.93) to yield

$$k_Y' = k_Z' \; . \tag{5.95}$$

Combining (5.94) and (5.95), we see

$$k_Y' = k_Z' = 0.5.$$

Now if one insists on interpreting the scaling constants as indicators of the importance of their respective attributes, then obviously from u, one must conclude that Y is three times as important as Z. And for the exact same attributes, using u', one would conclude Y and Z are equally important. This may be overemphasizing the point that scaling constants do not indicate the relative importance of attributes, but because this misinterpretation is so common, we thought a little overindulgence might be in order.

5.10. The Assessment of a Utility Function in a Hospital Blood Bank*

This section is meant to tie together the ideas of this chapter and illustrate our method for assessing multiattribute utility functions. This is done using a specific problem--one concerned with blood bank inventory control. The suggestions of Section 5.8 are followed in assessing the preferences of the decision maker in a hospital blood bank over the shortage--outdating consequence space. Although the example involves only two attributes, the general method described is applicable to a wide range of problems requiring multiattribute utility functions. Additional examples are discussed in Chapters 7, 8, and 9.

5.10.1. The Blood Bank Inventory Control Problem

Briefly discussing the blood bank inventory control problem and formulating it in the decision theoretic framework helps to motivate the assessment of the utility function presented below. Jennings [1968] developed a detailed model of a whole-blood inventory system for a blood bank in a hospital and examined the control of such a system. Various operating policies were evaluated in terms of blood

---

*Several parts of this section were adapted from Keeney [1972b].

shortage and blood outdating. Shortage is the blood requested by a doctor which could not be assigned from the hospital inventory. In this situation, a special order for the particular type of blood is placed with a central blood bank, professional donors may have to be called in, an operation may have to be postponed, etc.; but only in extremely rare circumstances would a death result from shortage as defined here. Outdated blood is the blood not used during its legal lifetime, which currently is 21 days in most hospitals.

One basic decision that must be made in hospital blood banks is what type of daily inventory ordering policy is best for each of the blood types. In this section the problem is analyzed for any one blood type. The decision maker must choose among the courses of action denoted by A_i where $i = 1,2,...,n$. For each A_i there is a probability distribution for consequences described in terms of Y and Z which represent shortage and outdating, respectively. More specifically, shortage can be stated in terms of yearly percent of units demanded and not filled from stock, and outdating can be measured in terms of yearly percent of units which outdate. The probability distributions can be obtained by simulation using a model such as Jennings' and from empirical records kept by the blood bank.

The structure of the inventory problem is illustrated in Figure 5.19, where the notation $(\tilde{y}_i, \tilde{z}_i)$ is used to designate the uncertain consequence of act A_i. The decision

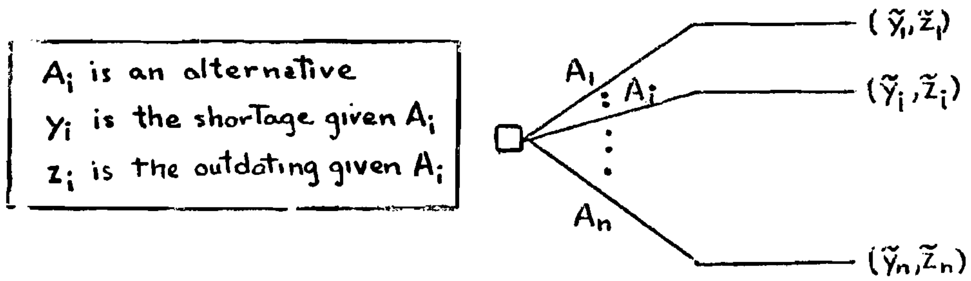


Figure 5.19. The Blood Bank Problem

maker should choose an act based on the assessed probability distributions of the paired random variables and his preferences for the various consequences.

A Perspective. The experiences recounted below are those of one author (Keeney) who contacted the doctor in charge of the blood bank at the Cambridge Hospital in Cambridge, Massachusetts. As part of a doctoral dissertation concerned with utility independence and assessing utility functions, the main purpose of the visit was to see if, in fact, the property of utility independence could be exploited in assessing utility functions. Hence, certain approaches which an analyst might take on a consulting assignment were not followed. Aside from considerations of whether the problem was the 'real problem', etc., one could cite two major shortcomings of this work if it had been a consulting assignment. These are (1) no attempt was made to exploit the value structure, using ideas such as those discussed in Chapter 3, before diving headfirst into the utility structure with probabilistic questioning, and (2) practically no concern was given to whether the decision maker was assessing preferences by considering only impacts to herself, or whether she included her perceived viewpoint of the impacts to patients, doctors, the hospital, and the public.

Notwithstanding the above caveats, we do feel that the process of assessment described below does provide a good indication of the general procedure.

5.10.2. Assessing the Utility Function

Introducing the Terminology and Ideas. On a first visit, Jennings' work was discussed with the doctor and the nurse in charge of ordering blood, and the importance of assessing preferences over the shortage-outdating space was indicated. On a subsequent visit the preferences of the nurse were assessed.* In the interim, she had read Jennings [1968] and developed a good understanding of the purpose of interviews. Before assessing the preferences, the purpose of utility theory was explained to the decision maker, and the meanings of the chosen measures of effectiveness were made clear. Thus, the decision maker realized the purpose of assessing her preferences and was motivated sufficiently to think hard about her feelings concerning the various consequences.

Prior to assessing her preferences, it was determined that shortage would never exceed ten percent of the units demanded and that outdating would not exceed ten percent of the total units stocked during a year. Thus the consequence space was limited, as shown in

---

*The nurse's preferences were used since she had responsibility for ordering whole blood for the blood bank. As indicated, the issue of whether her preferences appropriately represent those of the doctors and patients is ignored. Presumably, the nurse's preferences are influenced by her perceived preferences of the community served by the blood bank.

Figure 5.20. A check was made to ensure that the decision maker knew what was meant by a point (y,z) in the consequence space. When it was clear that the decision maker completely understood the basic ideas, it was possible to begin assessing preferences. At this time, it was stressed that there were no objectively correct or incorrect answers to the questions that would be asked.

Verifying Relevant Independence Assumptions. It was necessary to check whether Y (shortage) was utility independent of Z (outdating). This was done with the aid of Figure 5.21 where P,Q,R,S, etc. represent consequences. And as before, the notation $\langle P,Q \rangle$ will mean a lottery yielding either P or Q with equal probability. The decision maker was asked if she preferred $\langle P,Q \rangle$ or S. The consequence S was chosen by the questioner to make the question relatively simple. She preferred S, as one would expect intuitively. Next she was asked to choose between $\langle P,Q \rangle$ and T, and she chose $\langle P,Q \rangle$; this also was a relatively easy question. Progressively more difficult questions were posed about preferences between $\langle P,Q \rangle$ and N, $\langle P,Q \rangle$ and W, etc., and eventually her preferences "converged" to the fact that $\langle (0,0),(10,0) \rangle$ was indifferent to (6.5,0). Then the same types of questions were repeated using $\langle P',Q' \rangle$ instead of $\langle P,Q \rangle$, and indifference between $\langle (0,6),(10,6) \rangle$ and (6.5,6) was established. In fact, the decision maker stated that she did not see why the 6.5 *should* be different from the previous answer. In

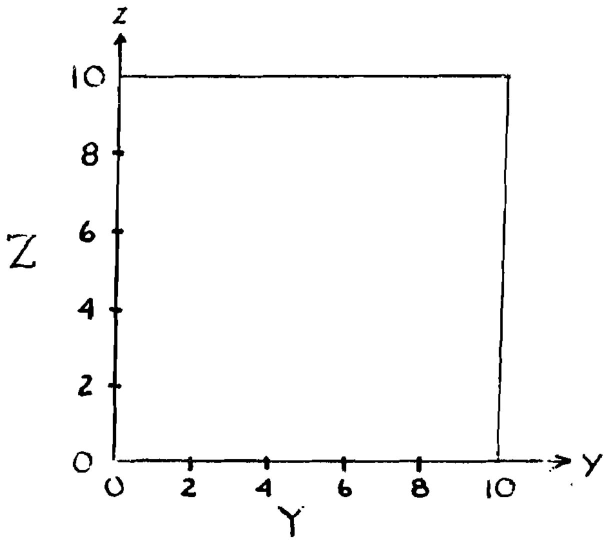


Figure 5.20. The Shortage - Outdating Consequence Space.

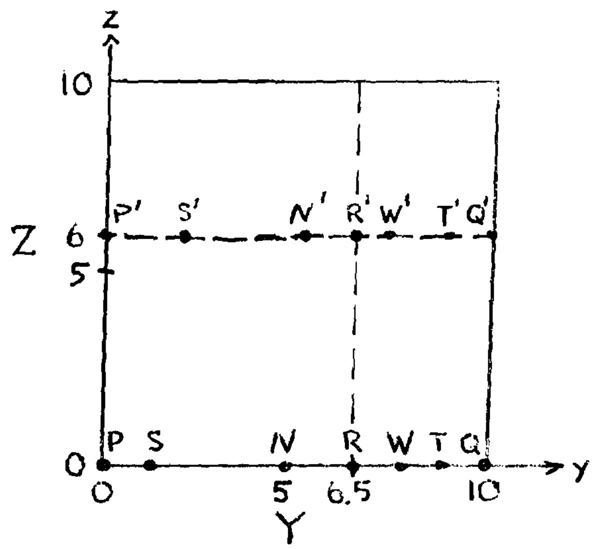


Fig. 5.21. Test Used To Verify Utility Independence

response to a general question she stated that the same was true for any level of Z held constant for all consequences. With this, it was concluded that Y was utility independent of Z. In a similar manner, Z was found to be utility independent of Y. Thus, the attributes were mutually utility independent, and the multilinear utility function previously discussed was applicable.

Assessing Conditional Utility Functions. Next, a conditional utility function for (y,0) was assessed. It was easy to establish that preference was monotonically decreasing in Y. Previously the lottery $\langle(10,0),(0,0)\rangle$ was shown to be indifferent to (6.5,0). In addition $\langle(6.5,0),(0,0)\rangle$ was indifferent to (4,0), and $\langle(10,0),(6.5,0)\rangle$ to (8.5,0). Thus, it was felt that the conditional utility function for (y,0) which will be denoted by $u_Y(y,0)$ was risk averse.

After arbitrarily setting the origin and unit of measure of $u_Y(y,0)$ by

$$u_Y(0,0) = 0 \tag{5.96}$$

and

$$u_Y(10,0) = -1, \tag{5.97}$$

the points on the utility function were plotted as indicated in Figure 5.22A. For simplicity, a utility function of the form $b(1 - e^{cy})$ was chosen. Using $\langle(10,0),(0,0)\rangle \sim (6.5,0)$, the parameter c was specified. Parameter b was

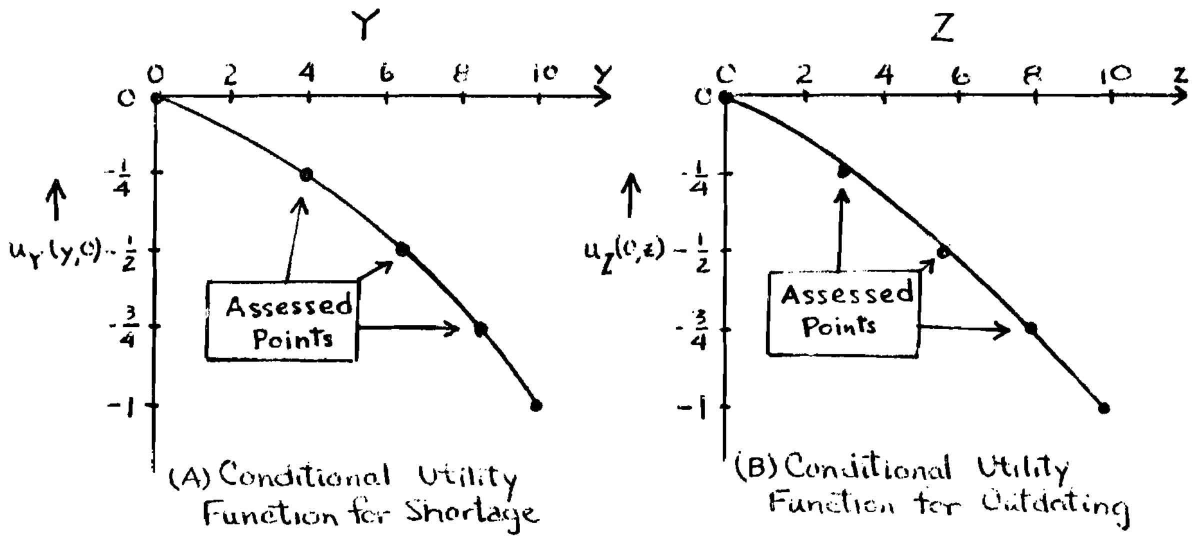


Figure 5.22. The Utility Functions in the Blood Bank Problem

then determined using (5.97) giving the result

$$u_Y(y,0) = \frac{1}{2.67}(1 - e^{.13y}) \tag{5.98}$$

shown in the figure. This form fit the other empirically assessed points very well. Since parameters b and c are positive, this utility function is monotonically decreasing and risk averse.

Similarly, in assessing $u_Z(0,z)$, the conditional utility function for Z, $\langle(0,10),(0,0)\rangle$ was found indifferent to (0,5.5), $\langle(0,5.5),(0,0)\rangle$ indifferent to (0,3), and $\langle(0,10),(0,5.5)\rangle$ indifferent to (0,8). Thus, by scaling

$$u_Z(0,0) = 0 \tag{5.99}$$

and

$$u_Z(0,10) = -1 \tag{5.100}$$

the points on the utility function shown in Figure 5.22B were determined. Again by fitting curves, this utility function was

$$u_Z(0,z) = \frac{1}{0.492}(1 - e^{.04z}). \tag{5.101}$$

Assessing the Scaling Constants. The next step of the assessment involved the consistent scaling of $u_Y(y,0)$ and $u_Z(0,z)$. It was determined that (0,10) was preferred to (10,0), (2,0) was preferred to (0,10), and finally that (0,10) was indifferent to (4.75,0). Now, it is possible to scale the utility function for (y,z), which will be denoted by u(y,z), as follows. First, set

$$u(0,0) = 0 \tag{5.102}$$

and

$$u(10,10) = -1 \ , \tag{5.103}$$

and define k_Y and k_Z by

$$u(10,0) = k_Y \tag{5.104}$$

and

$$u(0,10) = k_Z \ . \tag{5.105}$$

From (5.96), (5.97), (5.102), and (5.104), it follows that

$$u(y,0) = -k_Y u_Y(y,0). \tag{5.106}$$

Likewise, from (5.99), (5.100), (5.102), and (5.105),

$$u(0,z) = -k_Z u_Z(0,z). \tag{5.107}$$

Also, u(4.75,0) = u(0,10) or, by substituting from (5.100), (5.106), and (5.107),

$$-k_Y u_Y(4.75,0) = -k_Z u_Z(0,10) = k_Z. \tag{5.108}$$

Using (5.98), $u_Y(4.75,0) = -0.32$, which can be substituted into (5.108) to yield

$$k_Z = 0.32k_Y. \tag{5.109}$$

Because of mutual utility independence between Y and Z, u(y,z) is of the form*

$$u(y,z) = u(y,0) + u(0,z) + \frac{u(10,10) - u(0,10) - u(10,0)}{u(10,0)\ u(0,10)}\ u(y,0)u(0,z). \tag{5.110}$$

Substituting (5.98), (5.101), (5.106), (5.107), and (5.109) into (5.110), one finds

$$u(y,z) = \frac{-k_Y}{2.67}(1-e^{.13y}) - \frac{0.32k_Y}{0.492}(1-e^{.04z}) - \frac{(1+1.32k_Y)}{(2.67)(0.492)}(1-e^{.13y})(1-e^{.04z}). \tag{5.111}$$

The only parameter needed in (5.111) to completely specify u(x,y) is k_Y. To calculate k_Y it was established that the decision maker was indifferent between <(10,10),(0,0)> and (6,6). Then using (5.102) and (5.103),

$$u(6,6) = 1/2\ u(10,10) + 1/2\ u(0,0) = -1/2. \tag{5.112}$$

Equation (5.111) now can be evaluated at (6,6) and equated to (5.112) to yield

$$k_Y = -0.87. \tag{5.113}$$

One obtains the desired utility function, shown in Figure 5.23, by substituting (5.113) into (5.111):

$$u(y,z) = 0.32\ (1-e^{.13y}) + 0.57\ (1-e^{.04z}) + 0.107\ (1-e^{.13y})(1-e^{.04z}). \tag{5.114}$$

<u>Checking for Consistency</u>. Two types of consistency checks were conducted on this utility function. First, an

*Proof of this result is identical to Theorem 5.2 with the scale of u from minus one to zero rather than zero to plus one as in the theorem.

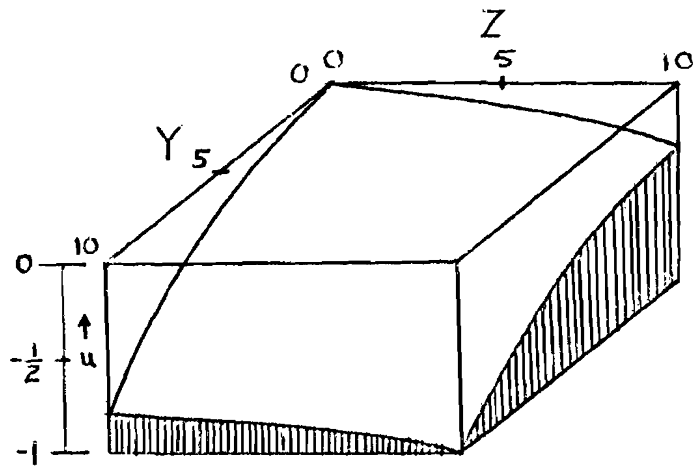


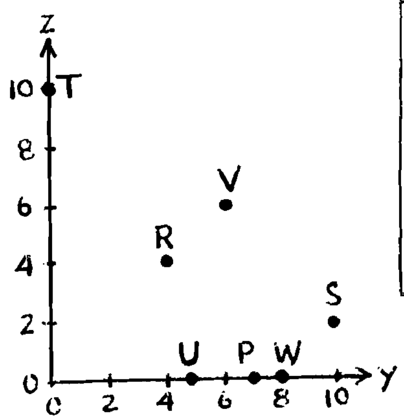
Fig 5.23. The Utility Function for Shortage and Outdating

alternative procedure was used to determine whether the conditional utility functions were risk averse as previously found. The decision maker felt (i,0) was preferred to $<(i+1,0),(i-1,0)>$ for $i = 1,2,\ldots,9$, and thus $u(y,0)$ was indeed risk averse. The same procedure resulted in a similar conclusion for $u(0,z)$.

The second check involved pairwise comparisons of consequences R,S,T,U,V,W, and P as defined in Figure 5.24. In response to questioning, the decision maker said $R \succ S, T \succ R, U \succ R, V \succ W$, and $P \succ V$ where $\succ$ is read "is preferred to." In the table of Figure 5.24, the utilities of these consequences, calculated using (5.114), are shown. A check shows them to be consistent with the decision maker's comparisons. This is true in spite of the fact that only one of the comparisons was simple, that is, an almost obvious choice. The particular outcome of the pairwise comparisons is due, at least partially, to happenstance. Nevertheless, this method for checking consistency is important.

5.10.3. Conclusions

By exploiting general characteristics of the preferences structure, such as utility independence, some of the difficulties of obtaining multiattribute utility functions are overcome. This reduces the actual amount of subjective information needed to specify the utility function. The procedure described here is operational both for identifying the utility independence characteristics



	(y,z)	u(y,z)
R	(4,4)	-.304
S	(10,2)	-.878
T	(0,10)	-.281
U	(4.75,0)	-.272
V	(6,6)	-.499
W	(8,0)	-.585
P	(7,0)	-.475

Fig. 5.24. Consistency Checks

of the preference structure and for assessing the multiattribute utility function.

Two concluding observations of the assessment process are noteworthy. First, the decision maker was interested and enthusiastic about what was being done, and she was willing to think hard about her preferences. This cooperation allowed the assessment procedure to go very smoothly and resulted in a utility function which seemed to represent her preferences accurately. Secondly, the decision maker had a degree in liberal arts, along with her nursing credentials, but no formal education in the quantitative areas; this did not hinder the assessment in any way. One could speculate from this that open-mindedness and willingness to think hard about the consequences are more important for correctly assessing preferences than any formal quantitative education.

CHAPTER 6

MULTIATTRIBUTE PREFERENCES UNDER UNCERTAINTY: MORE THAN TWO ATTRIBUTES

The development of this chapter parallels that of Chapter 5 with the distincion that here we are concerned with multiattribute utility functions with more than two arguments. The problem to be addressed in this Chapter, as well as the last one, was outlined in Section 5.1 and also in that section we discussed procedures for assessing a multiattribute utility function without first specifying the functional form. As indicated there, these procedures are valid for two and "more than two" attribute problems. Therefore, here we illustrate how various sets of additive independence, preferential independence, and utility independence assumptions among the attributes X_i, $i = 1,\ldots,n$ imply a utility function of the form

$$u(x_1,x_2,\ldots,x_n) = f[u_1(x_1),u_2(x_2),\ldots,u_n(x_n)] \quad , \tag{6.1}$$

where x_i is a specific amount of X_i, f is a scalar-valued function, and u_i is a utility function over X_i. These results generalize forms of (6.1) which have been derived for specific sets of preference assumptions by Fishburn [1965a, 1966, 1971], Keeney [1968, 1972a, 1974], Meyer [1970], Pollak [1967], and Raiffa [1969].

6.1. Overview of the Chapter

The results in this chapter are important for applications in that the assumptions are reasonable and operationally verifiable for many problems and furthermore the assessment of the resulting utility function in such a case is greatly simplified. A number of fundamental theoretical relationships between various independence conditions are also presented. These have practical importance in that they provide us with simpler sets of necessary and/or sufficient conditions for specific functional forms of the utility function.

6.1.1 Notation

In this chapter, it will be convenient to introduce a few new bits of notation to streamline the presentation. For reference, the important notation in this chapter is cataloged here:

Attributes. The basic attributes in most of our constructions will be $X_1, X_2, \ldots, X_n$, where X_i may be either a vector attribute or a scalar attribute. Beginning in Section 6.7, we introduce an additional attribute X_o, which plays a role that is distinct from the other X_i, $i = 1,2,\ldots,n$.

Sets of Attributes. The set of attributes X is defined as $\{X_1, X_2, \ldots, X_n\}$. When we use X_o, it also is in X. If Y is a subset of X, we will refer to the set of attributes in Y simply as attribute Y.

Complementary Sets. If two sets of attributes, call them Y_1 and Y_2 partition X, then we will refer to Y_1 and Y_2 as complements of each other. Often the complement of Y will be written $\overline{Y}$.

Preferential Independence and Utility Independence. Rather than repeatedly say that Y_1 is utility independent of its complement $\overline{Y}$ or that Y_2 is preferentially independent of its complement, we will write Y_1 is UI and Y_2 is PI respectively. This is done when no ambiguity can result, and it should be understood that UI or PI is implied relative to the complement.

Consequences. The consequence space $X_1 \times X_2 \times \cdots \times X_n$ represents a rectangular subset of finite dimensional Euclidean space. Consequences are designated by $\underline{x} \equiv (x_1, x_2, \ldots, x_n)$ where x_i designates a specific amount of X_i for $i = 1, 2, \ldots, n$. When referring to a subset Y of X and its complement $\overline{Y}$ we shall often designate $\underline{x}$ by $(y, \overline{y})$; thus for example, if $n = 5$ and $Y = \{X_1, X_3\}$, then $y = (x_1, x_3)$ and $\bar{y} = (x_2, x_4, x_5)$.

Utility Functions. As in previous chapters, we assume throughout Chapter 6 that a set of assumptions, such as von Neumann and Morgenstern's [1947], implying the existence of a utility function, are appropriate. The utility function u is assumed to be continuous in each x_i and bounded. We will write $u(\underline{x})$ or $u(x_1, x_2, \ldots, x_n)$ or $u(y, \overline{y})$ interchangeably.

Scaling. The symbol $\underline{x}^* \equiv (x_1^*, x_2^*, \ldots, x_n^*) \equiv (y^*, \overline{y}^*)$ designates the most desirable consequence and

$\underline{x}^o \equiv (x_1^o, x_2^o, \ldots, x_n^o) \equiv (y^o, \bar{y}^o)$ designates the least desirable. The utility function is scaled by $u(\underline{x}^o) = 0$ and $u(\underline{x}^*) = 1$. Rather than repeat many superscript zeroes, we sometimes take the liberty and write, for instance, $u(x_1, x_2^o, x_3^o, \ldots, x_n^o)$ and $u(x_1^o, x_2, x_3^o, x_4, x_5^o, x_6^o)$ as $u(x_1)$ and $u(x_2, x_4)$, respectively. That is, all attribute levels not explicitly denoted as arguments of a function are at their least desirable level. Granted, the symbolism is not entirely consistent, but the context should dispel any confusion. At least we hope so.

6.1.2 Independence Concepts

Now the concepts of preferential independence and utility independence introduced in previous chapters must be generalized.

Definition. Attribute Y, where $Y \subset X$, is preferentially independent of its complement $\bar{Y}$ if the preference order of consequences involving only changes in the levels in Y does not depend on the levels at which attributes in $\bar{Y}$ are held fixed.

Preferential independence implies the conditional indifference curves over Y do not depend on attributes $\bar{Y}$. The concept concerns the decision maker's preferences for consequences where no uncertainty is involved.

Utility independence, on the other hand, concerns preferences for lotteries which do involve uncertainty.

Definition. Attribute Y is utility independent of its complement $\bar{Y}$ if the conditional preference order for lotteries involving only changes in the levels of attributes in Y does not depend on the levels at which the attributes in $\bar{Y}$ are held fixed.

By definition, it follows that if Y is UI, then Y is PI. The converse is not necessarily true. This relationship can be seen by noting that degenerate lotteries, those involving no uncertainty, are the same things as a consequence. Hence, the preferential independence condition could be stated in terms of the preference order for degenerate lotteries only, and since the utility independence condition holds for all lotteries, the former is implied by the latter. Utility independence is the stronger condition.

If Y is preferentially independent of $\bar{Y}$, it follows that

$$\left[u(y',\bar{y}^{+}) \left\{\begin{matrix} > \\ = \\ < \end{matrix}\right\} u(y'',\bar{y}^{+})\right] \Longrightarrow \left[u(y',\bar{y}) \left\{\begin{matrix} > \\ = \\ < \end{matrix}\right\} u(y'',\bar{y})\right] \quad , \quad \forall\bar{y} \quad , \tag{6.2}$$

where $\bar{y}^{+}$ is any fixed level of $\bar{y}$. Similarly, if Y is utility independent of $\bar{Y}$, then since utility functions are unique up to positive linear transformations

$$u(y,\bar{y}) = f(\bar{y}) + g(\bar{y})\ u(y,\bar{y}') \quad , \quad \text{for all } y \text{ and } \bar{y} \quad , \tag{6.3}$$

where g is always positive and $\bar{y}'$ is an arbitrarily chosen specific amount of $\bar{Y}$. Functions f and g, in general, will depend on the specific value of $\bar{y}'$ but not on the variable y.

As indicated, throughout this chapter we choose for simplicity to scale the utility function from zero to one. Hence,

$$u(y^o,\bar{y}^o) = 0 \tag{6.4}$$

and

$$u(y^*,\bar{y}^*) = 1 \ , \tag{6.5}$$

where y^o and $\bar{y}^o$ are least preferred levels of Y and $\bar{Y}$ and y^* and $\bar{y}^*$ are the most preferred levels. Then, by evaluating (6.3) at y^o, we find

$$f(\bar{y}) = u(y^o,\bar{y}) \ ,$$

so condition (6.3) can be written as

$$u(y,\bar{y}) = u(y^o,\bar{y}) + g(\bar{y})\ u(y,\bar{y}^o) \ , \tag{6.6}$$

where we have chosen to set $\bar{y}'$ in (6.3) equal to $\bar{y}^o$. Equations (6.2) and (6.6) will be used in our proofs.*

6.1.3 Organization

The next section presents a number of representation theorems for three attributes. This (i) indicates some of the issues involved in assessing utility functions with more than two attributes, (ii), illustrates the type of results to be expected, and

---

*Preferential independence and utility independence can be generalized to allow for reversals of preferences as indicated in appendix 6A.

(iii) helps motivate the rest of the chapter. Functional forms of n-attribute utility functions which follow from various sets of preferential and utility independence conditions are presented in Sections 6.3 through 6.5, and assessment of such utility functions follows. Sections 6.7 through 6.10 generalize and tie together the concepts of preferential independence and utility independence. The extension of our results to hierarchical structures of attributes using conditional preference assumptions is the topic of Section 6.11.

6.2 Utility Functions With Three Attributes

Here we state and illustrate four results concerning utility functions with three attributes. Proofs are not included as all of these results are special cases of theorems presented and proven later in this chapter. The results are stated with the most restrictive case (in the sense of the strength of the requisite assumptions) first, and then the second most restrictive, the third, and finally the most general case.

Result 1. If preferences over lotteries on X_1, X_2, and X_3 depend only on their marginal probability distributions for these attributes and not on their joint distribution, then

$$u(x_1, x_2, x_3) = k_1 u_1(x_1) + k_2 u_2(x_2) + k_3 u_3(x_3) \quad . \qquad (6.7)$$

This result is the additive utility function of three attributes. The utility functions u, u_1, u_2, and u_3 can all be scaled from zero to one and the k_i's are scaling constants. Using a weaker set of assumptions, we have

<u>Result 2.</u> If X_1 is utility independent of $\{X_2, X_3\}$, and if $\{X_1, X_2\}$ and $\{X_1, X_3\}$ are preferentially independent of X_3 **and X_2**, respectively, then

$$\begin{aligned} u(x_1, x_2, x_3) = k_1u_1(x_1) &= k_2u_2(x_2) + k_3u_3(x_3) \\ &+ kk_1k_2u_1(x_1)u_2(x_2) + kk_1k_3u_1(x_1)u_3(x_3) \\ &+ kk_2k_3u_2(x_2)u_3(x_3) + k^2k_1k_2k_3u_1(x_1)u_2(x_2)u_3(x_3) \; . \end{aligned} \tag{6.8}$$

Each of u, the u_i's, and the k_i's in (6.8) have the same meaning as in (6.7). In addition k is an additional scaling constant. Clearly if $k = 0$, then (6.8) reduces to the additive form (6.7). If $k \neq 0$, then by multiplying each side of (6.8) by k, adding 1, and factoring, one obtains the multiplicative utility function

$$ku(x_1, x_2, x_3) + 1 = \prod_{i=1}^{3} [kk_iu_i(x_i) + 1] \; . \tag{6.9}$$

Two important things to note about Result 2 are that it uses both utility independence and preferential independence assumptions and that these assumptions concern "overlapping" sets of attributes. Both of these characteristics are very important in specifying multiattribute utility functions with many attributes. Since we use the notation u_2 and u_3 in this result, we implicitly imply that it can be proved that X_2 and X_3 are each utility independent of its complementary set of attributes. Becoming more general, one gets

<u>Result 3.</u> If each of X_1, X_2, and X_3 are utility independent of their respective complements, then

$$\begin{aligned} u(x_1,x_2,x_3) &= k_1u_1(x_1) + k_2u_2(x_2) + k_3u_3(x_3) \\ &+ k_{12}k_1k_2u_1(x_1)u_2(x_2) + k_{13}k_1k_3u_1(x_1)u_3(x_3) \\ &+ k_{23}k_2k_3u_2(x_2)u_3(x_3) + k_{123}k_1k_2k_3u_1(x_1) \\ &\cdot u_2(x_2)u_3(x_3) \quad . \end{aligned} \tag{6.10}$$

Again, the utility functions u, u_1u_2, and u_3 and the scaling constants k_1, k_2, and k_3 are defined as before. In addition, one needs to assess the additional scaling constants k_{12}, k_{13}, k_{23}, and k_{123}. Expression (6.10) is referred to as the multilinear utility function in three attributes. It should be clear that both the multiplicative and additive utility functions are special cases of the multilinear.

The most general case considered in this section is

<u>Result 4</u>. If X_2 and X_3 are utility independent of their respective complements $\{X_1, X_3\}$ and $\{X_1, X_2\}$, then

$$\begin{aligned} u(x_1,x_2,x_3) &= k_1u_1(x_1) + f_2(x_1)u_2(x_2) + f_3(x_1)u_3(x_3) \\ &+ f_{23}(x_1)u_2(x_2)u_3(x_3) \quad , \end{aligned} \tag{6.11}$$

where

$$f_2(x_1) = u(x_1, x_2^*, x_3^o) - u(x_1, x_2^o, x_3^o) \quad ,$$

$$f_3(x_1) = u(x_1, x_2^o, x_3^*) - u(x_1, x_2^o, x_3^o) \quad ,$$

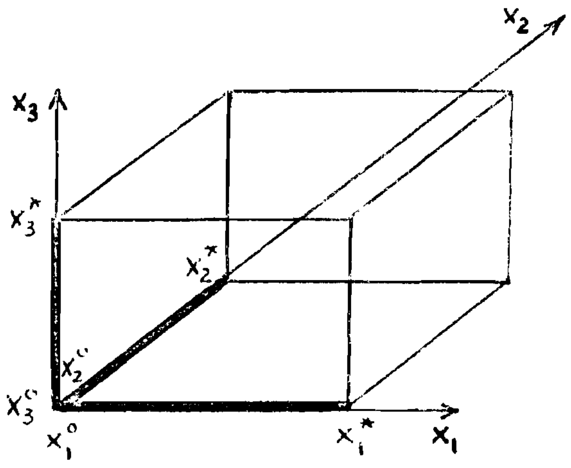
$$f_{23}(x_1) = u(x_1, x_2^*, x_3^*) - u(x_1, x_2^*, x_3^o)$$

$$- u(x_1, x_2^o, x_3^*) + u(x_1, x_2^o, x_3^o) \quad .$$

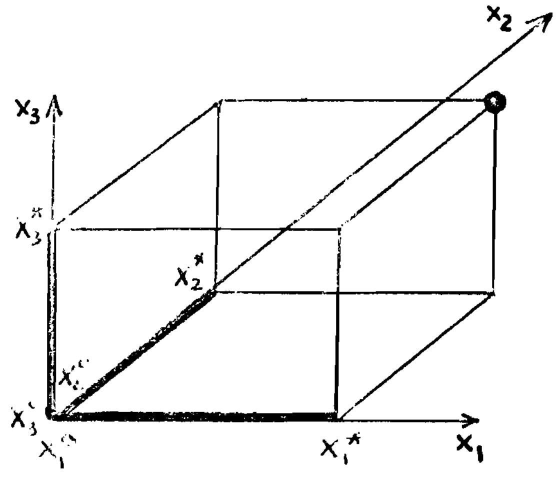
In (6.11), again each of the utility functions is scaled from zero to one, with (x_1^*, x_2^*, x_3^*) being the best consequence and (x_1^o, x_2^o, x_3^o) the worst. If f_2, f_3, and f_{23} are of certain forms, then it is easy to see that Results 1, 2, or 3 could result, and thus the additive, multiplicative, and multilinear utility functions are all special cases of (6.11).

If we consider the attributes as scalar attributes, then we can graphically illustrate what must be empirically assessed using each of the above results. This is done in Figure 6.1, where the dark lines and points indicate consequences which must be assessed on a common scale.

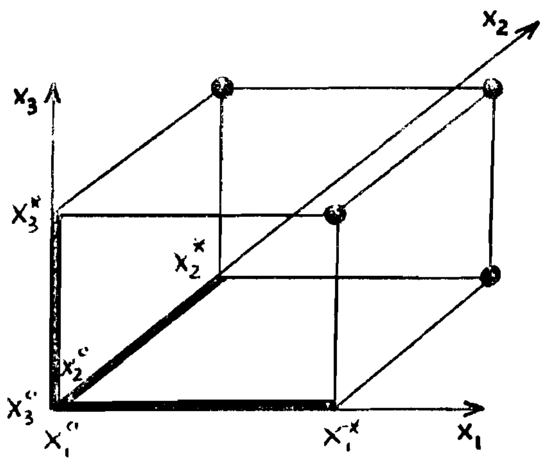
Most of the remainder of this chapter is used to develop results for n-attribute utility functions similar to those in this section. Once there are three or more attributes, it is possible to have overlapping sets of utility independence and preferential independence assumptions without having them contained in each other.



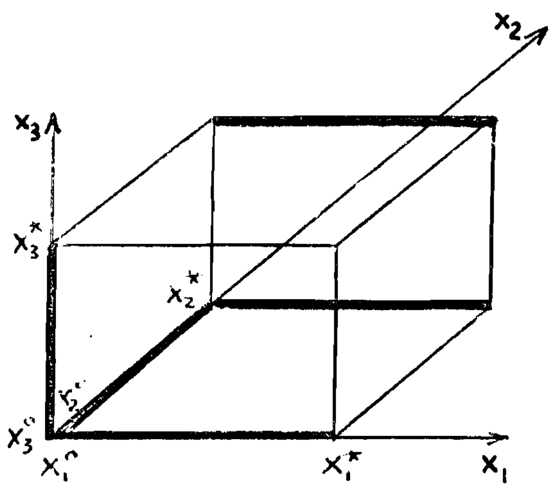
Result 1



Result 2



Result 3



Result 4

Figure 6.1. Empirical Assessment Required to Specify Certain Three-Attribute Utility Functions

This was not possible with just two attributes. As hinted at by the requisite assumptions for Result 2, it turns out to be very fruitful to look at the implications of such overlapping independence conditions. In the next three sections we begin to explore these implications by proving general theorems for n-attribute utility functions.

6.3 The Multiplicative Utility Function*

One of the most important results of multiattribute utility theory specifies conditions that enable one to conclude that a utility function is either multiplicative or additive. Let us first define mutual utility independence which is a sufficient condition for the fundamental result. After stating and proving this result, we will suggest several weaker sets of assumptions which imply mutual utility independence.

Definition. Attributes $X_1, X_2, \ldots, X_n$ are mutually utility independent if every subset of $\{X_1, X_2, \ldots, X_n\}$ is utility independent of its complement.

Theorem 6.1. If attributes $X_1, X_2, \ldots, X_n$ are mutually utility independent, then

$$u(\underline{x}) = \sum_{i=1}^{n} k_i u_i(x_i) + k \sum_{\substack{i=1 \\ j>i}}^{n} k_i k_j u_i(x_i) u_j(x_j) + k^2 \sum_{\substack{i=1 \\ j>i \\ \ell>j}}^{n} k_i k_j k_\ell u_i(x_i) u_j(x_j) u_\ell(x_\ell) \tag{6.12}$$

$$+ \cdots + k^{n-1} k_1 k_2 \cdots k_n u_1(x_1) u_2(x_2) \cdots u_n(x_n) \quad ,$$

*Material in this section is adapted from Keeney [1974].

where

(i) u is normalized by $u(x_1^o, x_2^o, \ldots, x_n^o) = 0$ and $u(x_1^*, x_2^*, \ldots, x_n^*) = 1$,

(2) $u_i(x_i)$ is a conditional utility function on X_i normalized by $u_i(x_i^o) = 0$ and $u_i(x_i^*) = 1$, $i = 1, 2, \ldots, n$,

(3) $k_i = u(x_i^*, \bar{x}_i^o)$,

and

(4) k is a scaling constant* which is a solution to

$$1 + k = \prod_{i=1}^{n} (1 + kk_i).$$

Remark: When $\sum_{i=1}^{n} k_i = 1$, then $k = 0$ and (6.12) reduces to the additive utility function

$$u(\underline{x}) = \sum_{i=1}^{n} k_i u_i(x_i) \quad . \tag{6.13}$$

On the other hand, when $\sum_{i=1}^{n} k_i \neq 1$, then $k \neq 0$, so we can multiply each side of (6.12) by k, add one to each, and factor to obtain

$$ku(\underline{x}) + 1 = \prod_{i=1}^{n} [kk_i u_i(x_i) + 1] \quad . \tag{6.14}$$

When k is positive in (6.14), then $u'(\underline{x}) \equiv 1 + ku(\underline{x})$ and $u_i'(x_i) \equiv 1 + kk_i u_i(x_i)$ are utility functions over the appropriate domains and

$$u'(\underline{x}) = \prod_{i=1}^{n} u_i'(x_i) \quad .$$

When k is negative, note that $u'(\underline{x}) \equiv -[ku(\underline{x}) + 1]$ and $u_i'(x_i) \equiv -[1 + kk_i u_i(x_i)]$ are utility functions over X

---

*Procedures for choosing the correct value of k are given in the appendix 6B at the end of this chapter.

and X_i, respectively, so

$$-u'(\underline{x}) = (-1)^n \prod_{i=1}^{n} u_i'(x_i) \quad .$$

Hence we can refer to form (6.14) as a multiplicative utility function.

<u>Proof</u>. Mutual utility independence by definition implies $\bar{X}_i$ is UI for i = 1,2,...,n - 1 which implies

$$u(\underline{x}) = u(x_i) + c_i(x_i)u(\bar{x}_i), \quad i = 1,2,\ldots,n-1 \quad . \tag{6.15}$$

Setting all $x_i = x_i^o$ except x_1 and x_j, j = 2,3,...,n - 1, we get the equality

$$u(x_1,x_j) = u(x_1) + c_1(x_1)u(x_j) = u(x_j) + c_j(x_j)u(x_1)$$

or

$$\frac{c_j(x_j) - 1}{u(x_j)} = \frac{c_1(x_1) - 1}{u(x_1)} = k \quad , \quad j = 2,3,\ldots,n-1 \quad , \quad u(x_j) \neq 0 \quad , \tag{6.16}$$

where k is some constant. If $u(x_j) = 0$, clearly $c_j(x_j) = 1$, so it follows that

$$c_i(x_i) = ku(x_i) + 1 \quad , \quad \text{for all } i = 1,2,\ldots,n-1 \quad . \tag{6.17}$$

We can repeatedly use (6.15) to obtain

$$u(\underline{x}) = u(x_1) + c_1(x_1)u(x_2, x_3, \ldots, x_n)$$

$$= u(x_1) + c_1(x_1)[u(x_2) + c_2(x_2)u(x_3, x_4, \ldots, x_n)]$$

$$\vdots$$

$$= u(x_1) + c_1(x_1)u(x_2) + c_1(x_1)c_2(x_2)u(x_3)$$

$$+ \ldots + c_1(x_1) \cdots c_{n-1}(x_{n-1})u(x_n) \quad . \tag{6.18}$$

Substituting (6.17) into (6.18) yields

$$u(\underline{x}) = u(x_1) + [ku(x_1) + 1]u(x_2) + [ku(x_1) + 1]$$

$$\cdot [ku(x_2) + 1]u(x_3) + [ku(x_1) + 1]$$

$$\cdot [ku(x_2) + 1] \cdots [ku(x_{n-1}) + 1]u(x_n) \quad . \tag{6.19}$$

When k = 0, (6.19) becomes the additive utility function

$$u(\underline{x}) = \sum_{i=1}^{n} u(x_i) \quad . \tag{6.20}$$

When $k \neq 0$, we can multiply both sides of (6.19) by k, then add 1 to each, and rearrange terms to find

$$ku(\underline{x}) + 1 = \prod_{i=1}^{n} [ku(x_i) + 1] \quad . \tag{6.21}$$

Recall that $u(x_i)$ actually means $u(x_1^o, \ldots, x_{i-1}^o, x_i, x_{i+1}^o, \ldots, x_n^o)$. Since we define

$$u(x_i) \equiv k_i u_i(x_i)$$

so the $u_i(x_i)$ can be scaled from zero to one, (6.20) and

(6.21) become respectively (6.13) and (6.14) which completes the proof. ◀

Notice that the results in Section 5.4 showing that the two-attribute utility function $u(x_1,x_2)$ is either multiplicative or additive if X_1 and X_2 are mutually utility independent is a special case of the above result.

Given that the conditions of Theorem 6.1 do hold, it is important to know whether the utility function is additive or multiplicative. One procedure is to pick any two attributes, say X_1 and X_2. Then, choose two amounts of X_1, call them x_1' and x_1'', between which the decision maker has a preference, and similarly, choose two amounts of X_2, say x_2' and x_2''. Next, fix the amounts of all the attributes other than X_1 and X_2 at some convenient level. Let us designate this as $\bar{x}_{12}^+$. Now we can state without proof the following

Corollary. If in addition to the requisite assumptions of Theorem 6.1, the decision maker is indifferent between a lottery yielding either $(x_1',x_2',\bar{x}_{12}^+)$ or $(x_1'',x_2'',\bar{x}_{12}^+)$ with equal probability or a lottery yielding either $(x_1',x_2'',\bar{x}_{12}^+)$ or $(x_1'',x_2',\bar{x}_{12}^+)$ with equal probability, the utility function must be additive. If he is not indifferent between these two lotteries, then the utility function must be multiplicative.

If the indifference or preference condition between the lotteries holds for one $\bar{x}_{12}^+$, it can be shown to hold for all $\bar{x}_{12}$ because $\{X_1,X_2\}$ is UI. Thus, it is not necessary to worry about the value of $\bar{x}_{12}^{+}$ in ascertaining whether the assumption is appropriate.

<u>6.3.1 Weaker Conditions Implying Mutual Utility Independence</u>

There are several sets of weaker conditions which imply mutual utility independence. These are important because they drastically reduce the number of conditions which need to be verified in order to utilize Theorem 6.1. Given a set of n attributes $\{X_1, X_2, \ldots, X_n\}$ there are $2^n - 2$ subsets which must be UI if mutual utility independence holds. For $n = 10$, this means that in the absence of weaker conditions, 1022 assumptions would need to be verified to ascertain mutual utility independence. The weaker conditions require at most n assumptions.

<u>Theorem 6.2.</u> Given attributes $X_1, X_2, \ldots, X_n$, the following are equivalent:

(1) attributes $X_1, X_2, \ldots, X_n$ are mutually utility independent,

(2) $\bar{X}_i$ is UI, $i = 1,2,\ldots,n$,

(3) $\{X_i, X_{i+1}, \ldots, X_n\}$ is UI, $i = 2,3,\ldots,n$, and $\{X_1, X_2, \ldots, X_{n-1}\}$ is UI,

(4) $\{X_i, X_{i+1}\}$ is UI, $i = 1,2,\ldots,n-1$; $n \geq 3$,

(5) X_1 is UI and $\{X_1, X_i\}$ is PI, $i = 2,3,\ldots,n$; $n \geq 3$.

Notice that by definition, case (1) implies (2) through (5). The reverse implications are proven in Section 6.9. There a procedure for generating sets of assumptions implying mutual utility independence is given. Cases (2), (3), and (4) are all special cases of this general result. The proof that (5) implies (1) also requires a fundamental relationship between preferential independence and utility independence derived in Section 6.7. Pollak [1967] used condition (2) and Meyer [1970] condition (3) to prove that the resulting function must be either multiplicative or additive. For conditions (4) and (5), note that there must be at least three attributes. Otherwise the conditions do not apply as their meaning is transparent.

With conditions (2), (3), (4), and (5), the number of assumptions increases linearly with the number of attributes. However, the sets (2), (3), and (4) require the decision maker to express preferences over lotteries with from two to n-1 attributes varying at a time. This turns out to be a very taxing task for a decision maker. The assumptions of (5) require only preferences over consequences with two attributes varying and preferences over lotteries involving one attribute. These latter conditions seem reasonable for many problems and have proven in practice--see Chapter 7 and 8--to be operationally verifiable.

6.4 The Multilinear Utility Function*

The multilinear utility function with n-attributes is a generalization of the three-attribute Result 3 discussed in Section 6.2 and a generalization of both the additive and

---

*The results of this section have been generalized in Fishburn [1973b] and Farquhar [1974]. The latter work includes decompositions with non-separable interaction terms.

multiplicative utility functions. The result is

<u>Theorem 6.3.</u> Given the set of attributes $X \equiv \{X_1, X_2, \ldots, X_n\}$ with $n \geq 2$, if X_i is utility independent of $\bar{X}_i$, $i = 1,2,\ldots,n$, then

$$u(x) = \sum_{i=1}^{n} k_i u_i(x_i) + \sum_{i=1}^{n} \sum_{j>i} k_{ij} u_i(x_i) u_j(x_j)$$

$$+ \sum_{i=1}^{n} \sum_{j>i} \sum_{\ell>j} k_{ij\ell} u_i(x_i) u_j(x_j) u_\ell(x_\ell)$$

$$+ \cdots + k_{123\ldots n} u_1(x_1) u_2(x_2) \cdots u_n(x_n) \quad , \tag{6.22}$$

where

(1) u is normalized by $u(x_1^o, x_2^o, \ldots, x_n^o) = 0$ and $u(x_1^*, x_2^*, \ldots, x_n^*) = 1$,

(2) $u_i(x_i)$ is a conditional utility function on X_i normalized by $u_i(x_i) = 0$ and $u_i(x_i^*) = 1$,

and

(3) the scaling constants can be evaluated† by

$$k_i = u(x_i^*, \bar{x}_i^o) \quad , \tag{6.23a}$$

$$k_{ij} = u(x_i^*, x_j^*, \bar{x}_{ij}^o) - k_i - k_j$$

$$= u(x_i^*, x_j^*, \bar{x}_{ij}^o) - u(x_i^*, \bar{x}_i^o) - u(x_j^*, \bar{x}_j^o) \quad , \tag{6.23b}$$

---

†To simplify expressions, we will extend our notation so that, for instance, when we write $(x_1, x_2, \bar{x}_{12})$, the symbol $\bar{x}_{12}$ will designate levels of all attributes except X_1 and X_2.

$$k_{ij\ell} = u(x_i^*, x_j^*, x_\ell^*, \bar{x}_{ij\ell}^o) - k_{ij} - k_{i\ell} - k_{j\ell} - k_i - k_j - k_\ell$$

$$= u(x_i^*, x_j^*, x_\ell^*, \bar{x}_{ij\ell}^o) - u(x_i^*, x_j^*, \bar{x}_{ij}^o) - u(x_i^*, x_\ell^*, \bar{x}_{i\ell}^o)$$

$$- u(x_j^*, x_\ell^*, \bar{x}_{j\ell}^o) + u(x_i^*, \bar{x}_i^o) + u(x_j^*, \bar{x}_j^o) + u(x_\ell^*, \bar{x}_\ell^o), \quad (6.23c)$$

and finally

$$k_{123\ldots n} = u(x^*) - \sum_i k_{1\cdots(i-1)(i+1)\cdots n} - \cdots - \sum_{i,j>i} k_{ij} - \sum_i k_i$$

$$= 1 - \sum_i u(x_i^o, \bar{x}_i^*) + \cdots + (-1)^{n-2} \sum_{i,j>i} u(x_i^*, x_j^*, \bar{x}_{ij}^o) + (-1)^{n-1} \sum_i u(x_i^*, \bar{x}_i^o) \quad .$$

(6.23d)

Proof. Because of the utility independence assumptions, from (6.6)

$$u(\underline{x}) = u(\bar{x}_i) + c_i(\bar{x}_i)u(x_i) \quad , \quad c_i > 0 \quad ,$$
$$i = 1,2,\ldots,n \quad . \quad (6.24)$$

where u will be scaled from zero to one. Let us define u_i to be a utility function over X_i scaled from zero to one. Then, noting that $u(x_i) = k_i u_i(x_i)$ for some positive constant k_i, we can define $d_i(\bar{x}_i) = k_i c_i(\bar{x}_i)$ and rewrite (6.24) as

$$u(\underline{x}) = u(\bar{x}_i) + d_i(\bar{x}_i)u_i(x_i) \quad , \quad d_i > 0 \quad , \quad (6.25)$$
$$i = 1,2,\ldots,n \quad .$$

To evaluate the d's, set x_i at its most desirable level x_i^* and evaluate (6.25), yielding

$$u(x_i^*,\bar{x}_i) = u(x_i^o,\bar{x}_i) + d_i(\bar{x}_i)u_i(x_i^*) \quad ,$$

and since $u_i(x_i^*) = 1$,

$$d_i(\bar{x}_i) = u(x_i^*,\bar{x}_i) - u(x_i^o,\bar{x}_i) \quad , \; i = 1,2,\ldots,n \quad . \tag{6.26}$$

Substituting (6.26) into (6.25) and rearranging, we find

$$u(\underline{x}) = u_i(x_i)u(x_i^*,\bar{x}_i) + [1 - u_i(x_i)]u(x_i^o,\bar{x}_i) \quad , \quad i = 1,2,\ldots,n \quad . \tag{6.27}$$

The proof is conceptually simple, although algebraically tedious, from here. One repeatedly substitutes (6.27) into itself for $i = 1,2,\ldots,n$ and rearranges terms to get the result. We indicate the first step, where (6.27) with $i = 2$ is substituted into (6.27) with $i = 1$:

$$\begin{aligned}
u(\underline{x}) &= u_1(x_1)u(x_1^*,\bar{x}_1) + [1 - u_1(x_1)]u(x_1^o,\bar{x}_1) \\
&= u_1(x_1)\Big[u_2(x_2)u(x_1^*,x_2^*,\bar{x}_{12}) + [1 - u_2(x_2)]u(x_1^*,x_2^o,\bar{x}_{12})\Big] \\
&\quad +[1 - u_1(x_1)]\Big[u_2(x_2)u(x_1^o,x_2^*,\bar{x}_{12}) + [1 - u_2(x_2)]u(x_1^o,x_2^o,\bar{x}_{12})\Big] \\
&= u(x_1^o,x_2^o,\bar{x}_{12}) + [u(x_1^*,x_2^o,\bar{x}_{12}) - u(x_1^o,x_2^o,\bar{x}_{12})]u_1(x_1) \\
&\quad +[u(x_1^o,x_2^*,\bar{x}_{12}) - u(x_1^o,x_2^o,\bar{x}_{12})]u_2(x_2) \\
&\quad +[u(x_1^*,x_2^*,\bar{x}_{12}) - u(x_1^*,x_2^o,\bar{x}_{12}) - u(x_1^o,x_2^*,\bar{x}_{12}) \\
&\quad +u(x_1^o,x_2^o,\bar{x}_{12})]u_1(x_1)u_2(x_2) \quad .
\end{aligned} \tag{6.28}$$

Repeating the procedure, we get the desired result (6.22) and (6.23). ◀

That (6.22) is a generalization of the multiplicative and additive utility functions can be seen by comparing the results of Theorem 6.1 and Theorem 6.3. With the multilinear utility function (6.22), there are $2^n - 1$ scaling constants, but since $u(\underline{x}^*) = 1$, we know the sum of all these constants must equal 1 so only $2^n - 2$ are independent. Using (6.23), these can be evaluated from the utilities of the "corner" consequences in X.

6.5. The Additive Utility Function

In this section, we are interested in the n-attribute additive utility function. Much of the work in additive utility theory has been done by Fishburn [1964, 1965a, 1965b, 1966, 1967a, 1967b, 1967c, 1970, 1971, 1972]. He has derived necessary and sufficient conditions for additive utility functions in many situations, including whole product sets, denumerable products sets, incomplete product sets, and interdependencies among some attributes. Pruzan and Jackson [1963] and Pollak [1967] also present necessary and sufficient conditions for a utility function to be additive.

With n attributes, Fishburn's additive independence condition can be defined as follows.

Definition. Attributes $X_1, X_2, \ldots, X_n$ are additive independent if preferences over lotteries on $X_1, X_2, \ldots, X_n$

depend only on their marginal probability distributions and not on their joint probability distribution.

With this condition we can state a fundamental result of additive utility theory.

<u>Theorem 6.4. (Fishburn).</u> The n-attribute additive utility function

$$u(\underline{x}) = \sum_{i=1}^{n} u(x_i, \bar{x}_i^o) = \sum_{i=1}^{n} k_i u_i(x_i) \tag{6.29}$$

is appropriate if and only if the additive independence condition holds among attributes $X_1, X_2, \ldots, X_n$, where

(1) u is normalized by $u(x_1^o, x_2^o, \ldots, x_n^o) = 0$ and $u(x_1^*, x_2^*, \ldots, x_n^*) = 1$,

(2) u_i is a conditional utility function on X_i normalized by $u_i(x_i^o) = 0$ and $u_i(x_i^*) = 1$, $i = 1, 2, \ldots, n$,

(3) $k_i = u(x_i^*, \bar{x}_i^o)$, $i = 1, 2, \ldots, n$.

<u>Proof</u>. The proof follows from repeated use of the derivation of the two-attribute additive utility function in Theorem 5.1. If we define Y as $\{X_2, X_3, \ldots, X_n\}$, from Theorem 5.1,

$$u(x_1, x_2, \ldots, x_n) = k_1 u_1(x_1) + k_Y u_Y(x_2, x_3, \ldots, x_n). \tag{6.30}$$

Then to break down u_Y, we define $Z \equiv \{X_3, X_4, \ldots, X_n\}$ and invoke Theorem 5.1 again to yield

$$u_Y(x_2, x_3, \ldots, x_n) = k_2 u_2(x_2) + k_Z u_Z(x_3, x_4, \ldots, x_n) . \tag{6.31}$$

We proceed in this manner and then substitute (6.31) into (6.30), etc., to yield the result (6.29). To insure proper scaling, all the utility functions can be scaled from zero to one. The converse follows directly from calculating the expected utility of any lottery using the additive utility function. ◀

Pollak's [1967] formulation of necessary or sufficient conditions for additive utility functions leads to

Theorem 6.5 (Pollak). An individual's utility function is additive if and only if his preference between any two lotteries

$$L_1 \equiv \langle (x_i,\bar{x}_i'),(x_i^a,\bar{x}_i^a) \rangle \text{ and } L_2 \equiv \langle (x_i,\bar{x}_i''),(x_i^b,\bar{x}_i^b) \rangle$$

is the same for all x_i for any $\bar{x}_i', \bar{x}_i''$, $x_i^a, \bar{x}_i^a, x_i^b$, and $\bar{x}_i^b$. [Remark before proof. Pollak's basic assumption is illustrated in Fig. 6.2, where L_1 is $\langle A,B \rangle$ and L_2 is $\langle C,D \rangle$. Note that consequences A and C have the same amount of attribute X_i. Pollak's assumption says that whatever preference we have between L_1 and L_2, we must also have if the level x_i in A and C is changed. That is, if A and C are slid horizontally to A' and C', the preference between $L_1' \equiv \langle A',B \rangle$ and $L_2' \equiv \langle C',D \rangle$ must be the same as between L_1 and L_2].

Proof. If u is additive, the expected utilities of the above lotteries (using 6.29) are, respectively,

$$E[u(L_1)] = \frac{1}{2}\,u(x_i,\bar{x}_i^o) + \frac{1}{2}\,u(x_i^o,\bar{x}_i') + \frac{1}{2}\,u(x_i^a,\bar{x}_i^a)$$

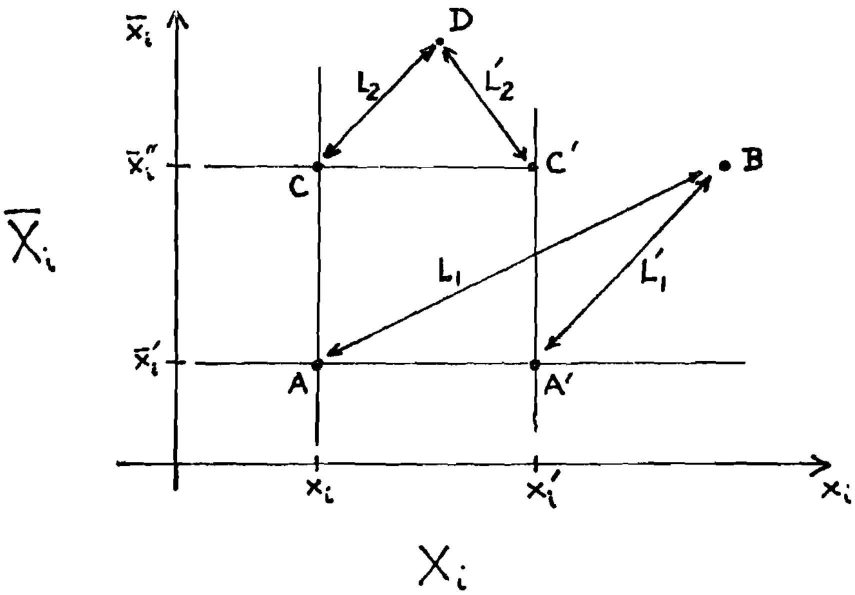


Figure 6.2. An Illustration of Pollak's Additivity Condition

and

$$E[u(L_2)] = \frac{1}{2}\, u(x_i, \bar{x}_i^o) + \frac{1}{2}\, u(x_i^o, \bar{x}_i'') + \frac{1}{2}\, u(x_i^b, \bar{x}_i^b)$$

Subtracting $E[u(L_2)]$ from $E[u(L_1)]$, it is clear that one's preferences between L_1 and L_2 would not depend on x_i.

Now assume preferences between L_1 and L_2 do not depend on x_i. Let us set $\bar{x}_i^b = \bar{x}_i', \bar{x}_i^a = \bar{x}_i''$, and $x_i^a = x_i^b = x_i^o$. Then $L_1 = \langle (x_i, \bar{x}_i'), (x_i^o, \bar{x}_i'') \rangle$ and $L_2 = \langle (x_i, \bar{x}_i''), (x_i^o, \bar{x}_i') \rangle$. For one value of x_i, namely x_i^o, lotteries L_1 and L_2 are equally preferable. Therefore, from the condition of the theorem, this must be the case for any x_i. It now follows directly from repeated application of Theorem 5.1 that u is additive. ◀

The main advantage to the additive utility function is its relative simplicity. The assessment of the n-attribute utility function is reduced to the assessment of n one-attribute utility functions and n - 1 independent scaling constants. Any of the techniques discussed in this book could be used for assessing the one-attribute utility functions. The problem of evaluating scaling constants is addressed in the next section.

A major shortcoming of the additive utility function is the restrictiveness of the necessary assumptions. We could often expect the utility of a lottery to depend not only on the marginal probability distributions of the respective attributes, but also on their joint probability distribution. Another problem is that it is difficult to determine whether or not the requisite assumptions would be reasonable in a specific real-world problem. This difficulty arises because the assumptions are stated in terms of the decision maker's preferences for probability distributions over consequences, with more than one attribute simultaneously varying.

6.6. Assessing Multiattribute Utility Functions*

With the additive, multiplicative, and multilinear utility functions, preferential independence and utility independence have been used to reduce the assessment of an n-attribute utility function to the assessment of n one-attribute utility functions, which can be designated as $u_1, u_2, \ldots, u_n$, and some scaling constants k_j, $j = 1,2,\ldots,r$. And so we have

$$u(x_1, x_2, \ldots, x_n) = f[u_1(x_1), u_2(x_2), \ldots, u_n(x_n), k_1, k_2, \ldots, k_r] , \tag{6.32}$$

where f is a scalar function. Each of the u_i can be assessed independently, since the scaling constants are meant to insure consistent scaling among the u_i's. Thus, except for the fact that there are more of them, the problem of assessing the u_i's is no more involved than

---

*Three publications with discussions concerning the assessment of multiattribute utility functions are Fishburn [1967d], Huber [1974b], and Kneppreth, et al. [1974]. Boyd [1970,1973] discusses an interactive routine for multiattribute utility assessment.

in the two-attribute case considered in Section 5.8. Therefore, we won't address this topic here. However, the problems of verifying the independence conditions and assessing the k_i's are more involved with more attributes. The approach remains the same as in the two-attribute case, but operationally things get a little messier.

6.6.1. Verifying Preferential Independence and Utility Independence Conditions

Partition X into Y and $\bar{Y}$. To check whether Y is preferentially independent of $\bar{Y}$, we might proceed as follows. First choose $\bar{y}^+$ with all components at a relatively undesirable level and choose y' and y'' such that $(y', \bar{y}^+)$ is indifferent to $(y'', \bar{y}^+)$. Then pick another point $\bar{y}'$ with all components at a relatively desirable level and ask the decision maker if $(y', \bar{y}')$ is indifferent to $(y'', \bar{y}')$. This must be true if Y is preferentially independent of $\bar{Y}$. If the decision maker's answer was affirmative, repeat the same procedure for other pairs of y consequences with $\bar{y}$ fixed at various levels. If the answers to these questions still indicate preferential independence, then ask the decision maker, "If you are indifferent between $(y', \bar{y})$ and $(y'', \bar{y})$ for some particular $\bar{y}$, does this imply the same indifference would hold for every choice of $\bar{y}$?" A positive answer implies Y is preferentially independent of $\bar{Y}$.

An obvious way to check whether Y is utility independent of $\bar{Y}$ is to assess conditional utility functions over Y given different amounts of $\bar{Y}$. If they are positive linear transformations of each other, the utility independence assumption would be appropriate. More specifically, one could assess certainty equivalents $\hat{y}$ such that $(\hat{y},\bar{y})$ is indifferent to a lottery yielding either $(y',\bar{y})$ or $(y'',\bar{y})$ with equal probability. If the certainty equivalent for any lottery did not depend on the amount $\bar{y}$, then Y would be utility independent of $\bar{Y}$. In practice, if such a condition held for three or four fifty-fifty lotteries covering the range of Y for approximately four different values of $\bar{y}$ covering the range of $\bar{Y}$, one would usually be justified in assuming Y is utility independent of $\bar{Y}$.

6.6.2. Evaluating Scaling Constants

The basic objective to be followed to evaluate k_1 to k_r should be obvious from Section 5.8. We want to obtain a set of r independent equations which have the k_j's as r unknowns. These are then solved to get the k_j's. The set of equations can be generated from certainty considerations, probabilistic considerations, or a combination of both. For example, if consequences $\underline{x}$ and $\underline{y}$ are equally preferred by the decision maker, then clearly $u(\underline{x}) = u(\underline{y})$, or from (6.32)

$$f[u_1(x_1),\ldots,u_n(x_n),\ k_1,\ldots,k_r] = f[u_1(y_1),\ldots,u_n(y_n),\ k_1,\ldots,k_r] \quad . \qquad (6.33)$$

Once the u_i's have been assessed, $u_i(x_i)$ and $u_i(y_i)$ are just numbers, so (6.33) is one equation with at most r unknowns.

Also, for example, if $\underline{x} \sim \langle \underline{y}, p, \underline{z} \rangle$, then substituting (6.32) into

$$u(\underline{x}) = pu(\underline{y}) + (1 - p)u(\underline{z})$$

gives us another equation with at most r unknowns.

Two operational problems of concern are (1) how does one guarantee the equations are independent and (2) what should one do with more than r independent equations when they are inconsistent.

In practice, one's understanding of the problem and knowledge of the functional form of the utility function are probably the best guards against a large number of redundant equations. Even so, it is interesting to think about one approach which can be used to avoid any redundancy with the multilinear utility function since it involves the most scaling constants of any of our functional forms. Recall that for this case we need 2^n-2 scaling constants, where n is the number of attributes. There are 2^n "corner" consequences of the form $(x_1', x_2', \ldots, x_n')$, where $x_i' = x_i^*$ or x_i^o, and where $u(x_1^*, x_2^*, \ldots, x_n^*) = 1$ and $u(x_1^o, x_2^o, \ldots, x_n^o) = 0$ are used to scale u. If each corner consequence is evaluated in terms of these two reference consequences, or other previously assessed consequences, we will get an independent set of 2^n-2 equations. The most obvious, although not necessarily the best, way to do this is equate each corner consequence to a lottery of the form $\langle (x_1^*, x_2^*, \ldots, x_n^*)$, p, $(x_1^o, x_2^o, \ldots, x_n^o) \rangle$ by assessing the appropriate p. When redundancies do occur, in a set of equations, one needs to empirically generate additional equations relating the k_j's until we do have a set of r independent equations. An example illustrating this is given in Section 5.8.

Concerning overdetermination with inconsistencies, the desire is clearly to have the decision maker reflect on the inconsistencies--which perhaps can be illuminated by the analyst--and change some responses to imply a consistent set of preferences. If, due to time considerations or whatever, this is impossible, then perhaps sensitivity analysis using the different sets of implied scaling factors would indicate the same alternative was best. Or at least, it may be possible to drop some options from further consideration. Of these remaining, we should be able to identify which parameters are critical to the decision, and from this, develop a procedure to specify these parameter values.

6.6.3. Scaling the Conditional Utility Functions

As will become apparent in this section, the problem of scaling conditional utility functions is very similar to that of scaling conditional value functions addressed in Section 3.7. The techniques discussed for assessing scaling factors in the value function context are directly applicable to our current problem. However, in the utility context, the additional possibility of scaling by using probabilistic questioning is appropriate.

The additive, multiplicative, and multilinear utility functions can be written

$$u(x_1,x_2,\ldots,x_n) = \sum_{i=1}^{n} k_i u_i(x_i) + \text{POT} \quad , \tag{6.34}$$

where POT designates "possible other terms." With the additive form, there are no other terms, whereas the opposite is true with the multiplicative and multilinear forms. In each case, when the u_i and u are scaled from zero to one by

$$u(x_1^*,x_2^*,\ldots,x_n^*) = 1, \ u(x_1^o,x_2^o,\ldots,x_n^o) = 0 \tag{6.35}$$

and

$$u_i(x_i^*) = 1, \ u_i(x_i^o) = 0 \quad , \quad \text{for all } X_i \quad , \tag{6.36}$$

then

$$u(x_i^*,x_{\bar{i}}^o) = k_i \quad , \quad i = 1,2,\ldots,n \quad . \tag{6.37}$$

The problem of interest in this subsection is assessing the scaling factors for the conditional utility functions in (6.34) which is done by specifying the k_i's for $i = 1,\ldots,n$. This requires n independent equations with the n unknown k_i's to be generated.

Because the difficulty in manually solving n equations, which are not necessarily linear, with n unknowns is to say the least tedious, current practice in assessing the k_i's usually requires   sets of equations that are simple to evaluate. This basically limits the questions to two types.

Question I. For what probability p are you indifferent between

(1) the lottery giving a p chance at $\underline{x}^*$ and a 1 - p chance at $\underline{x}^o$, and

(2) the consequence $(x_1^o,\ldots,x_{i-1}^o,x_i^*,x_{i+1}^o,\ldots,x_n^o)$.

If we define the decision maker's answer as p_i, then using (6.35), the expected utility of the lottery is p_i, and from (6.37) the utility of the consequence is k_i. Equating the expected utilities, we find

$$k_i = p_i \quad .$$

One could then clearly generate the values of each of the k_i's in this fashion.

The second type of question is illustrated by

Question II. Select a level of X_i, call it x_i', and a level of X_j, call it x_j', such that, for any fixed levels of all other attributes, you are indifferent between

(1) a consequence yielding x_i' and x_j^o together, and

(2) a consequence yielding x_j' and x_i^o together.

Using (6.35) and (6.36), the utilities of these two indifferent consequences can be equated to yield

$$k_i u_i(x_i') = k_j u_j(x_j') \quad . \tag{6.38}$$

Once the single attribute utility functions u_i and u_j are assessed, both $u_i(x_i')$ and $u_j(x_j')$ are easily found, so (6.38) is a simple linear equation. Suppose in addition, for example that $x_i' = x_i^*$, then by (6.36), the relationship between k_i and k_j given by (6.38) is even simpler.

A major shortcoming of questions of both types I and II is the use of the extreme levels of the attributes, that is the x_i^*'s and x_i^o's. Since the range from x_i^o to x_i^* must cover all the possible x_i's, the implications of, and hence preferences for, the extreme levels are usually very difficult for a decision maker to assess. A further difficulty with Question I is the fact that the effect due to varying all n attributes simultaneously must be considered. Hence for computational ease, we must force the decision maker to respond to questions much more difficult to evaluate than would be theoretically necessary. A computer program developed to eliminate this necessity is discussed in Appendix 6C.

A common practice in assessing the k_i's would be to first rank them, then to use question I to evaluate the largest k_i, and finally to use type II questions to evaluate the magnitude of the other k_j's relative to the largest k_i. Once we have the k_i's, the additive form must hold if they sum to one. Otherwise, the k_i's are substituted into (6.34) to use in evaluating k for the multiplicative form or the other constants in the multilinear form. This task in itself can be difficult.

It should not be a particularly difficult task to order the k_i's. This can be done, for instance, by asking the decision maker if he prefers $(x_1^*, \bar{x}_1^o)$ or $(x_2^*, \bar{x}_2^o)$. Using

(6.37) if the former is preferred, then $k_1 > k_2$; if the latter is preferred, then $k_2 > k_1$ and if they are equally preferable, then $k_1 = k_2$. Repeating this for other binary comparisons, one would get a complete ranking of the k_i's. At most it would require $(n^2-n)/2$ such comparisons for a complete ranking, but judicious choice of the order could reduce this to only n-1 comparisons. For instance, it isn't unreasonable to present the decision maker with a list of consequences and ask him to rank them. Using this ranked list as a beginning, we could check its consistancy by asking the n-1 binary choices between now adjacent consequences. The ideas of how one might order the k_i's should now be clear*. Asking a decision maker to rank the k_i's before actually assessing them serves to introduce him to the tradeoff considerations which he must make without initially overwhelming him with complexity.

Example 6.1. Suppose we ascertain that $k_1 > k_2 > k_3$, and actually, for future purposes in this example, we only need to know the largest k_i. Next we ask the decision maker for an x_1, call it x_1', such that (x_1', x_2^o, x_3^o) and (x_1^o, x_2^*, x_3^o) are equally preferable. From (6.34), it follows that

$$k_1 u_1(x_1') = k_2 \quad , \tag{6.39}$$

where $u_1(x_1')$ is just a number between zero and one.

---

*As discussed in Section 5.8, the k_i cannot be interpreted as indicators of the relative importance of the attributes X_i.

Similarly, we ask for another x_1, call it x_1'', such that $(x_1'', x_2^o, x_3^o) \sim (x_1^o, x_2^o, x_3^*)$. Equating utilities gives us

$$k_1 u_1(x_1'') - k_3 \quad . \tag{6.40}$$

Note that our information is identical to that used in Chapter 3 to scale value functions.

If our utility function is additive, then from (6.29) for consistency, we know

$$k_1 + k_2 + k_3 = 1 \quad . \tag{6.41}$$

The set of equations (6.39), (6.40), and (6.41) can easily be solved to yield the appropriate values of the k_i's. And of course for consistency, k_3 must be less than k_2.

If our utility function is multiplicative, then from (6.14) for consistency,

$$k + 1 = (kk_1 + 1)(kk_2 + 1)(kk_3 + 1) \quad . \tag{6.42}$$

Equations (6.39), (6.40), and (6.42) together have four unknowns: k_1, k_2, k_3, and k, so we must generate another equation. Using probabilistic scaling, we might determine p_1 such that (x_1^*, x_2^o, x_3^o) is indifferent to $\langle (x_1^*, x_2^*, x_3^*),\ p_1,\ (x_1^o, x_2^o, x_3^o) \rangle$. Using the three attribute multiplicative utility function and equating expected utilities, we find

$$k_1 = p_1 \quad . \tag{6.43}$$

This equation, together with (6.39), (6.40), and (6.42) can be solved to yield the scaling constants. $\blacksquare$

6.6.4. Scaling the Additive Utility Function

To make a specific point, let us turn our attention to the k_i's in an n-attribute additive utility function. Note that one needs only to assess empirically n-1 of the k_i factors since the n^{th} can be specified from these and the consistency requirement

$$\sum^{n} k_i = 1 \quad . \tag{6.44}$$

Although the use of questions of type I and II for assessing the k_i's is simple procedurally, it may not be so simple operationally. The decision maker may in fact find some of the questions asked of him using this method very difficult. Unfortunately we must introduce more notation before discussing ways to get around such difficulties. For any subset T of the indices $\{1,2,\ldots,n\}$, let x_T be the $\underline{x}$ point where the i^{th} attribute is x_i^* if i belongs to T and is x_i^o if i does not belong to T. Thus if $n = 5$ and $T = \{1,2,4\}$, then

$$x_T = x_{\{1,2,4\}} = (x_1^*, x_2^*, x_3^o, x_4^*, x_5^o). \tag{6.45}$$

Also let us define

$$k_T = \sum_{i \in T} k_i , \tag{6.46}$$

and p_T as the probability such that the decision maker is indifferent between lottery $L_T \equiv \langle x^*, p_T, x^o \rangle$ and consequence x^T. Using this notation, if we assess p_T for any particular T, then $u(x_T)$ must equal the expected utility of L_T. The expected utility of L_T is clearly p_T and $u(x_T) = k_T$, so

$$k_T = p_T \qquad \text{for all } T \tag{6.47}$$

From (6.46), one sees that k_T can be obtained for any subset T from the individual k_i's. However, one motivation here is that it may be easier for the decision maker to go the other way, that is, to get the k_i's from the k_T's.

This can be done using (6.46) and (6.47).

Example 6.2. Again let $n = 5$ and suppose $T = \{1,2,4\}$ and $R = \{1,2\}$. Then if we empirically assess p_T and p_R, it follows from (6.46) and (6.47) that

$$k_T = k_1 + k_2 + k_4 = p_T$$

and

$$k_R = k_1 + k_2 = p_R \quad .$$

Clearly then for this example

$$k_4 = p_T - p_R \quad . \tag{6.48}$$

There are obviously many consistency checks which can be performed to verify our assignment of k_4. For instance, suppose we assess p_Q for $Q = \{1,2,3,5\}$. Then since from (6.47), it follows that

$$k_1 + k_2 + k_3 + k_5 = p_Q$$

and from (6.45)

$$k_1 + k_2 + k_3 + k_4 + k_5 = 1 \quad ,$$

we know

$$k_4 = 1 - p_Q \quad .$$

Another obvious consistency check on k_4 is to assess it directly, as previously indicated, by obtaining p_4. ∎

An alternative approach to assessing the k_i factors is suggested by the following idea from probability theory. In assigning probabilities to a finite set of mutually exclusive and collectively exhaustive events $\{E_1, E_2, \ldots, E_r\}$, it is often natural to make an assignment first to a subset of these events and then to use conditional probability considerations to further subdivide this assignment. We might find it helpful to proceed in an analogous manner in the present context. To this end suppose S is a subset of T. We want to find what portion of the weight of k_T should be assigned to S. Letting $p_{S|T}$ be the probability such that x_S is indifferent to $\langle x_T, p_{S|T}, \underline{x}^o \rangle$, and equating expected utility,

$$u(x_S) = p_{S|T} u(x_T) \quad .$$

From this, we establish the rule that

$$k_S = p_{S|T} k_T, S \subset T \quad , \tag{6.49}$$

which is analogous to the multiplication rule of probability theory.

How one finally chooses to assess the k_i's, whether directly, or indirectly by using (6.49), depends on which procedure seems most natural in the context of the real problem under consideration.

6.6.5. Scaling the Multiplicative Utility Function*

Techniques for evaluating the scaling constants k_i in the multiplicative utility function were addressed in subsection 6.6.3. However, the scaling constant k is special to the multiplicative form so it is considered now. Given Theorem 6.1 holds and $\sum_{i=1}^{n} k_i = 1$, then the additive utility function is appropriate. If $\sum_{i=1}^{n} k_i \neq 1$, the utility function is multiplicative and the additional constant k in (6.14) can be found from the k_i values.

In this case, we can evaluate (6.14) at $\underline{x}^*$ to find

$$1 + k = \prod_{i=1}^{n} (1 + kk_i) \quad . \tag{6.50}$$

If $\sum_{i=1}^{n} k_i > 1$, then using (6.14) and (6.50), the utility independence properties of the utility function (6.14) can only be preserved given that $-1 < k < 0$. In this case, by iteratively evaluating (6.50) given the k_i, $i = 1,2,\ldots,n$, one can converge to the appropriate value of k, call it k*. First set k = k' and substitute this into (6.50). If the right-hand side is smaller than the left-hand side, then k* < k'. If the r.h.s. is greater than the l.h.s., then k* > k'.

When $\sum_{i=1}^{n} k_i < 1$, it follows from similar reasoning that k* > 0. Let us arbitrarily set k = k' in (6.50). If the r.h.s. > l.h.s., then k* < k', whereas if the l.h.s. > r.h.s., then k* > k'.

---

*The assertions in this subsection are proven in Appendix 6B at the end of this chapter.

6.6.6. An Example

To illustrate some of the ideas of this section, let us consider the problem of selecting a job. And just to keep matters simple, let us assume there are three attributes to be considered about each job, namely, monetary compensation, commuting travel time, and degree of urbanization of the area. These will be designated by X_1, X_2, and X_3, respectively. Furthermore, we will assume monetary compensation is broken down into starting salary and future prospects for increases, which we will designate as Y_1 and Y_2, respectively. Hence, $X_1 = Y_1 \times Y_2$. The measurement scales for each of the attributes are summarized in Table 6.1.

Table 6.1

Measurement Scales for Attributes

		Attribute	Measurement Scale
X_1	Y_1	starting salary	starting annual income in dollars
	Y_2	future increases in salary	annual salary after five years in dollars
X_2		commuting travel time	door-to-door travel time from work to job in minutes
X_3		degree of urbanization	metropolitan area population

Now suppose that the additive independence assumptions have been verified for all the X_i terms, and furthermore that this condition does <u>not</u> hold for Y_1 and Y_2. Then according to Theorem 6.4, the utility function $u(x_1,x_2,x_3)$ is additive. To make use of this we must first establish best and worst possible outcomes under any job for each of the attributes, i.e., the x_i^o and x_i^* values. Let us assume these are found as shown in Table 6.2.

Table 6.2

Range of Measurement Scales

			Range	
			Best	Worst
X_1	Y_1	starting salary	\$18,000	\$12,000
	Y_2	future salary increases	\$25,000	\$12,000
X_2		commuting travel time	10 min.	60 min.
X_3		degree of urbanization	15 million	0.5 million

Now we can define $y_1^o = 12{,}000$ and $y_2^o = 12{,}000$ so that $x_1^o \equiv (y_1^o, y_2^o) = (12{,}000;12{,}000)$. Similarly, let $x_2^o = 60$ and $x_3^o = 0.5$. Then choose $x_1^* = (18{,}000;25{,}000)$, $x_2^* = 0$, and $x_3^* = 15$. Notice that x_2^* was set equal to zero minutes travel time even though the best condition required at least ten minutes. This is legitimate for our purposes since the only condition on x_2^* was that it be at least as good as the best possible consequence.

Now that we have specified our x_i^o and x_i^* amounts, we can write from (6.29) that

$$u(x_1,x_2,x_3) = k_1u_1(x_1) + k_2u_2(x_2) + k_3u_3(x_3) \quad , \tag{6.51}$$

where

$$u_i(x_i^o) = 0 \quad , \quad u_i(x_i^*) = 1 \quad , \quad \text{for all } i \quad . \qquad (6.52)$$

The manner in which one would assess the u_2 and u_3 functions subject to the convention of (6.52) was extensively covered in Chapter 4, so nothing more will be said about this here. But the utility function u_1 is a utility function for more than one scalar attribute, namely, Y_1 and Y_2. And as we stated, the decision maker is unwilling to accept the additive independence assumption for these two attributes, so a simple additive function is not appropriate. Perhaps we could use some of the substitution schemes discussed in Chapter 3, which essentially reduce a two-dimensional representation to a unidimensional representation before the conversion to utilities is effected. But what if we're not so lucky? Alternatively, some of the methods to assess two-attribute utility functions discussed in Chapter 5 may be appropriate.

Now let us return to the scaling constants of (6.51). It is possible to ask the decision maker some meaningful qualitative questions about k_i's to get some "feeling" for their values. For instance: "Imagine that each of the performance measures are at the state x_i^o. Would you rather have attribute X_1 pushed to x_1^* than both attributes X_2 and X_3 pushed to x_2^* and x_3^*?" A yes answer would imply $k_1 > k_2 + k_3$, which means $k_1 > 0.5$. We could then ask: "Would you rather have attribute X_2 pushed from x_2^o to x_2^* than X_3 pushed from x_3^o to x_3^*?" And if this question received an affirmative answer, we would know $k_2 > k_3$. If there exists a subset T of attributes such that $x_T \sim x_{\bar{T}}$, then we can infer that $k_T = k_{\bar{T}} = \frac{1}{2}$.

In many ways the richer the set of attributes, the easier it becomes to group attributes in a way that permits the analyst to infer properties of the k_i's without asking probabilistic questions. At any stage of an anlysis such as this, the sophisticated analyst would use sensitivity checks to determine whether he need probe any further. Perhaps the crude qualitative measures already obtained suffice to resolve the original problem.

Another methodological point that needs clarification is the notion of consistency. When questions are asked one way, it might turn out that $k_1 > 0.3$, say, and when asked another way $k_1 < 0.3$. This will happen, and when it does the decision maker will just have to think harder about the issues and modify some of his assumptions or evaluations in order to attain consistency. This is psychologically painful and time-consuming and once again this step should be preceded by a sensitivity analysis to determine whether the inconsistency is worth resolving.

Let's proceed. Suppose that we assess $k_1 = 0.6$, that is, the decision maker is indifferent between (x_1^*, x_2^o, x_3^o) and the lottery $\langle (x_1^*, x_2^*, x_3^*), 0.6, (x_1^o, x_2^o, x_3^o) \rangle$. Then of course

$k_2 + k_3 = 0.4$ and we can ask, for instance, "What is the value of p such that you are indifferent between (x_1^o, x_2^*, x_3^o) and $\langle (x_1^o, x_2^*, x_3^o)$, p, $(x_1^o, x_2^o, x_3^o)\rangle$?" If the decision maker's response os 0.7, from (6.49) we have

$$k_2 = p(k_2 + k_3) = (.7)(.4) = 0.28.$$

Then clearly, $k_3 = 0.12$ so the utility function is

$$u(x_1, x_2, x_3) = 0.6\ u_1(x_1) + 0.28\ u_2(x_2) + 0.12\ u_3(x_3)\ , \tag{6.53}$$

where each of the utility funcitons is scaled from zero to one. Expression (6.53) is then appropriate for evaluating decisions under uncertainty. Of course, one might want to run sensitivity tests on those aspects of the assessment procedure that appear to be most unstable.

6.6.7 Consistency Checks

As with all phases of assessing utility functions, it is important to include consistency checks to develop some confidence in our representation of the decision maker's preferences. Clearly, when we check the consistency of the overall utility function, we are also checking the appropriateness of the scaling constants. It is also prudent to include consistency checks specifically for these scaling constants. In all of these checks, we simply set up additional equations which have some scaling constants in them. But since we have already evaluated these constants, we can plug in their values to check the original assessment. The different approaches for evaluating scaling constants can obviously be used as checks of each other. In most situations, the imagination of the analyst will not be stretched in an attempt to develop efficient and effective consistent checks for the scaling constants.

6.7. A Fundamental Relationship Between Preferential Independence and Utility Independence

We now begin to introduce general results which allow us to weaken the assumptions necessary to invoke Theorems such as those in Sections 6.3 to 6.5. The result of this section relates our two independence conditions concerning cardinal and ordinal preferences over the consequence space X. It allows us to build "higher order" utility independence conditions from the weaker preferential independence conditions of the same order and lower order utility independence conditions.* Decision makers find it very difficult to think about lotteries involving more than one attribute because they must consider simultaneously both tradeoffs between different levels of the attributes and the probabilities that the various consequences occur. However, one can reasonably specify a preference order for lotteries involving only one attribute. Also, although it is not easy, one can fairly accurately indicate the tradeoffs between two attributes

---

*Let $Y \subset X \equiv \{X_1,\ldots,X_n\}$. If Y is UI or PI, the order of the assumption is the number of X_i's in Y. Thus, for example, the assumption that $Y \equiv \{X_2,X_3\}$ is UI is a second order assumption.

under certainty with all the levels of the other attributes fixed. By completing each of these tasks separately, the assumptions necessary to invoke Theorems 6.1 to 6.4 implying specific forms of the multiattribute utility function can be verified.

The general case of our main result can be proven with $X \equiv \{X_o, X_1, X_2\}$ since the X_i's can be vector attributes. The attribute X_o introduced is distinct from the other X_i's. Throughout this chapter, it will never be assumed or implied that X_o is either PI or UI. Hence, it will never be the case that preferences for X_o will ever be independent in any sense of $\bar{X}_o$. We are interested in the utility function $u(x_o, x_1, x_2)$, which is assumed to be continuous, with each argument of u having a definite effect on preferences. It is assumed that preferences over X are bounded and (x_o^*, x_1^*, x_2^*) will designate the most desirable and (x_o^o, x_1^o, x_2^o) the least desirable consequence.

Theorem 6.6.* Given three attributes $\{X_o, X_1, X_2\}$, if $\{X_1, X_2\}$ is preferentially independent of X_o and if X_1 is utility independent of $\{X_o, X_2\}$, then $\{X_1, X_2\}$ is utility independent of X_o.

Remark: This result says preferential independence of $\{X_1, X_2\}$ from its complement can be strengthened to

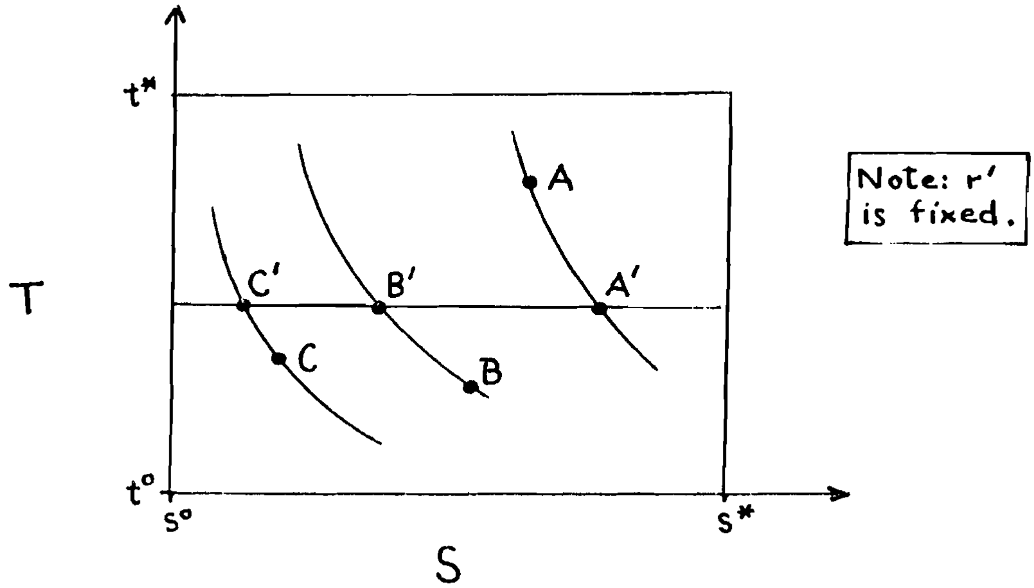
---

*This result does not require boundedness and the independence conditions can be weakened to allow for reversals of preferences over various attributes as proven in Fishburn and Keeney [1974].

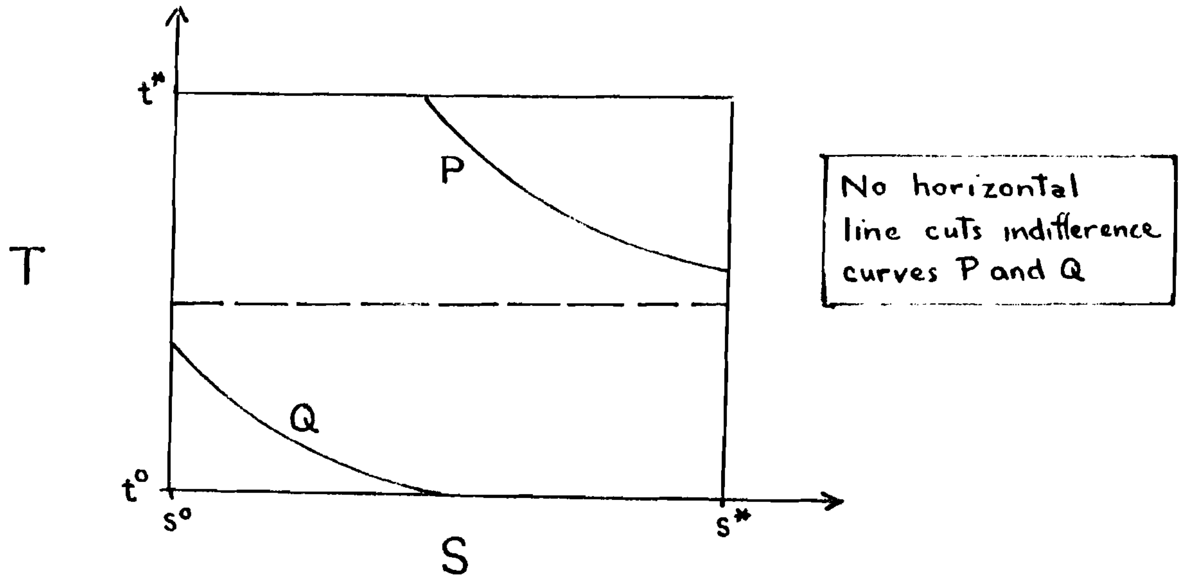
utility independence, provided that either X_1 or X_2 is UI. Theorem 6.6 provides necessary conditions for a "second order" utility independence assumption in terms of a "second order" preferential independence assumption and a "first order" utility independence assumption.

The proof of Theorem 6.6 is fairly involved, but the presentation can be simplified with a special notation. To avoid subscripts where unnecessary, we will define attributes $R \equiv X_0, S \equiv X_1$, and $T = X_2$. Thus for instance, s will be a specific amount of S, and the utility function of interest will be written as u(s,t,r).

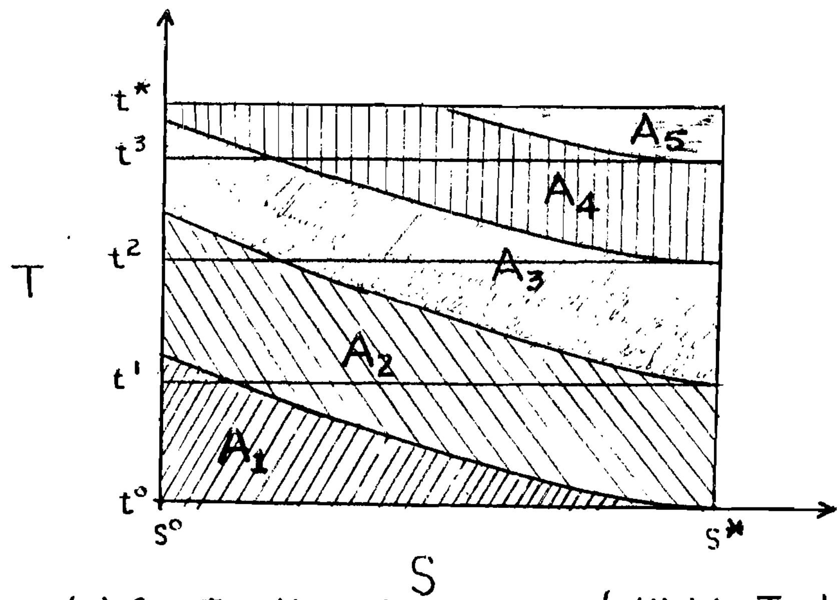
[Idea of Proof. The mode of the proof can be illustrated by taking s and t as scalars. Let r' be any fixed value of R and consider the three iso-preference or indifference curves shown in Figure 6.3**A.** These same conditional indifference curves are appropriate for any other value, r" say, since {S,T} is PI. Suppose that given r', we know that $B \sim \langle A,C \rangle$. The essence of the proof is to show that given any other value of R, r" say, then it still holds that $B \sim \langle A,C \rangle$. But one of the fundamental axioms of utility theory is the substitution principle: a lottery is not made better or worse by substitutiong equivalent (or indifferent) prizes. Hence given r' or r" we know that $\langle A,C \rangle \sim \langle A',C' \rangle$ and $B \sim B'$. But since S is UI we know that $B' \sim \langle A',C' \rangle$ given r' implies that it still is valid given r'. This demonstrates the essence of the proof and it would be simple to clean up the details of the proof if each indifference curve would cut a single horizontal line. But what happens if we have



(A) Conditional Indifference Curves in S×T



(B) A Complicating Case



(C) Constructing the Domain of Utility Independence

Figure 6.3. Illustration of The Proof for Theorem 6.6

two indifference curves P and Q as shown in Figure 6.3B? In this case we have to modify our previous argument and build up the domain of applicability in stages. It is first shown (see Figure 6.3C) that the condition for {S,T} to be utility independent of R holds for all r and (s,t) pairs in A_1. Then, because the line $t = t^1$ overlaps with A_1, we can show the condition holds for all (s,t^1). Since each (s,t) pair in A_2 is indifferent to some pair (s,t^1), the necessary utility independence condition can be extended to include A_2. Then an amount t^2 is chosen such that the line $t = t^2$ overlaps A_2, and the procedure is repeated. Eventually, one of the A_i's (A_4 in Figure 6.3C) will overlap with the line $t = t^*$ so the utility independence condition can be proven valid on that line and extended to A_5. Since the A_i's cover all (s,t) pairs, the utility independence condition is valid for all s,t, and r.]

Proof.* We can represent the condition that S is utility independent of {T,R} by

$$u(s,t,r) = u(t,r) + b(t,r)\,u(s) \quad , \quad b > 0. \tag{6.54}$$

Also, since {S,T} is preferentially independent of R, from (6.2) we know

$$\left[u(s,t,r^o) = u(s^+,t^+,r^o)\right] \Rightarrow \left[u(s,t,r) = u(s^+,t^+,r)\right] \quad , \forall r. \tag{6.55}$$

---

*In this proof, when an attribute is at its least desirable amount, designated as s^o, for example, we may delete it in the function when no ambiguity will result. Thus, rather than write $u(s^o,t^o,r)$, $u(s^o,t,r)$, and $b(t^o,r)$, we will use $u(r)$, $u(t,r)$, and $b(r)$.

For each pair (s,t) in A_1 defined by

$$A_1 \equiv \{(s,t) : u(s,t,r^o) \leq u(s^*,t^o,r^o)\} \quad , \tag{6.56}$$

there exists an s' such that

$$u(s,t,r^o) = u(s',t^o,r^o) \quad , \quad (s,t) \in A_1 \quad . \tag{6.57}$$

From (6.55) and (6.57), it follows that

$$u(s,t,r) = u(s',t^o,r) \quad , \quad \forall r, \ (s,t) \in A_1 \ . \tag{6.58}$$

Evaluating (6.58) with (6.54), we find

$$u(s,t,r) = u(r) + b(r)\, u(s') \quad , \ \forall r, \ (s,t) \in A_1 \quad , \tag{6.59}$$

which combined with (6.57) gives us

$$u(s,t,r) = u(r) + b(r)\, u(s,t) \quad , \ \forall r, \ (s,t) \in A_1 \quad . \tag{6.60}$$

Equation (6.60) says $\{S,T\}$ is utility independent of R for $(s,t) \in A_1$ and all r. We want to extend this condition to all possible (s,t) pairs.

Choose a t^1 such that

$$u(s^o,t^o,r^o) < u(s^o,t^1,r^o) < u(s^*,t^o,r^o) \quad .$$

Since $(s^o,t^1) \in A_1$, from (6.60)

$$u(s^o,t^1,r) = u(r) + b(r)\, u(t^1) \quad , \quad \forall r \quad . \tag{6.61}$$

Evaluating (6.54) at $t = t^1$ and $r = r^o$ yields

$$u(s,t^1) = u(t^1) + b(t^1)\, u(s) \quad , \quad \forall s \quad . \tag{6.62}$$

Setting $t = t^1$ in (6.60) gives us

$$u(s,t^1,r) = u(r) + b(r)\, u(s,t^1) \quad , \quad (s,t^1) \in A_1 \quad , \qquad (6.63)$$

which can be combined with (6.62) to yield

$$u(s,t^1,r) = u(r) + b(r)[u(t^1) + b(t^1)\, u(s)]$$
$$= u(t^1,r) + b(r)\, b(t^1)\, u(s) \quad , \quad \forall r, \quad (s,t^1) \in A_1 \ . \qquad (6.64)$$

Comparing (6.64) to (6.54) with $t = t^1$ shows

$$b(t^1,r) = b(r)\, b(t^1) \quad , \quad \text{all } r \ . \qquad (6.65)$$

Substituting (6.61) and (6.65) into (6.54) with $t = t^1$ yields

$$u(s,t^1,r) = u(t^1,r) + b(t^1,r)\, u(s)$$
$$= u(r) + b(r)[u(t^1) + b(t^1)\, u(s)] \quad , \quad \forall s,r \quad , \qquad (6.66)$$

which can be combined with (6.62) to give

$$u(s,t^1,r) = u(r) + b(r)\, u(s,t^1) \quad , \quad \forall s,r \ . \qquad (6.67)$$

To extend result (6.67), let us define A_2 by

$$A_2 \equiv \{(s,t) : u(s^*,t^o,r^o) < u(s,t,r^o) \leq u(s^*,t^1,r^o)\} \ .$$

For any $(s,t) \in A_2$, there exists an s" such that

$$u(s,t,r^o) = u(s'',t^1,r^o) \quad , \quad (s,t) \in A_2 \quad , \qquad (6.68)$$

so from (6.55), it follows that

$$u(s,t,r) = u(s'',t^1,r) \quad , \quad \text{all } r \quad , \quad (s,t) \in A_2 \ . \qquad (6.69)$$

Evaluating the right-hand side of (6.69) with (6.67) gives

$$u(s,t,r) = u(r) + b(r)\ u(s'',t^1)$$

which, combined with (6.68), gives us

$$u(s,t,r) = u(r) + b(r)\ u(s,t) \quad , \quad (s,t) \in A_2 \quad . \tag{6.70}$$

Equation (6.70) says {S,T} is utility independent of R for (s,t) pairs in A_2.

The process from (6.61) on is now repeated by choosing an amount t^2 such that

$$u(s^o,t^1,r^o) < u(s^o,t^2,r^o) < u(s^*,t^1,r^o) \quad ,$$

and then proving that (6.67) holds with t^2 substituted for t^1. Then (6.70) is extended to include all (s,t) pairs such that

$$u(s^*,t^1,r^o) < u(s,t,r^o) \le u(s^*,t^2,r^o) \quad .$$

Because of the continuity assumptions on u and the fact that S is essential (i.e., u(s) is not a constant and b(t,r) is positive), by repeating this process with a more preferred t on each iteration, we will eventually prove

$$u(s,t^*,r) = u(r) + b(r)\ u(s,t^*)$$

so that for any (s,t) pair such that

$$u(s,t,r^o) = u(s',t^*,r^o)$$

for some s', the utility independence expression similar to (6.70) will follow, More formally, let us define

$$h = \min_{t,r} [u(s^*,t,r) - u(s^o,t,r)]$$

so that it follows from (6.54) that

$$h = \min_{t,r} [b(t,r)]$$

which is positive. Then in choosing the series $t^1, t^2, \ldots,$ if t^k is such that $u(s^o,t^*,r^o) < u(s^*,t^k,r^o)$, choose $t^{k+1} = t^*$. Otherwise select t^{k+1} such that

$$u(s^o,t^{k+1},r^o) = u(s^o,t^k,r^o) + \frac{h}{2} \quad .$$

Since $u(s^o,t^*,r^o)$ must by definition, be less than one, the series $t^1, t^2, \ldots, t^*$ will require at most $\frac{1}{h/2} = \frac{2}{h}$ members.

By the manner in which the A_i's are defined, collecting all the equations similar to (6.60), (6.70), etc., will prove that

$$u(s,t,r) = u(r) + b(r)\, u(s,t) \quad , \quad \forall s,t,r \quad , \qquad (6.71)$$

which is the desired result. ◀

6.8 Relationships Among Utility Independence Assumptions

Let us look at some implications of different sets of utility independence conditions. In particular, we will be interested in implying higher order utility independence conditions from lower order conditions. The results included in this section are requisite for the general theorems which follow in the next sections. Here our result concerns the implications of two overlapping utility independence assumptions.

Definition. Let Y_1 and Y_2 be subsets of $X \equiv \{X_1, X_2, \ldots, X_n\}$. Attributes Y_1 and Y_2 overlap if their intersection is not empty and if neither contains the other.

Theorem 6.7. Let Y_1 and Y_2 be overlapping attributes contained in $X \equiv \{X_0, X_1, \ldots, X_n\}$. If Y_1 and Y_2 are each UI then

(i) $Y_1 \cup Y_2$, the union of Y_1 and Y_2, is UI,

(ii) $Y_1 \cap Y_2$, the intersection of Y_1 and Y_2, is UI,

(iii) $(Y_1 \cap \bar{Y}_2) \cup (\bar{Y}_1 \cap Y_2)$, the symmetric difference of Y_1 and Y_2, is UI,

(iv) $Y_1 \cap \bar{Y}_2$ and $\bar{Y}_1 \cap Y_2$, the differences, are each UI.

Proof. Since X_i can designate a vector attribute, the general case can be proven by considering the special case where $X = \{X_o, X_1, X_2, X_3\}$, $Y_1 = \{X_1, X_2\}$, and $Y_2 = \{X_2, X_3\}$, and where Y_1 and Y_2 are each assumed to be UI.

We must show in this case that (i) $\{X_1, X_2, X_3\}$ is UI, (ii) X_2 is UI, (iii) $\{X_1, X_3\}$ is UI, and (iv) X_1 is UI and X_3 is UI. From (6.6), our hypotheses can be written respectively as

$$u(\underline{x}) = u(x_o, x_1, x_2, x_3) = u(x_o, x_3) + c(x_o, x_3)\, u(x_1, x_2) \quad , \tag{6.72}$$

and

$$u(\underline{x}) = u(x_o, x_1, x_2, x_3) = u(x_o, x_1) + d(x_o, x_1)\, u(x_2, x_3) \quad , \tag{6.73}$$

where as before we have taken the liberty to delete arguments of u,c, and d when they are at their least preferred levels and no misunderstanding can result, that is, when $x_i = x_i^o$. Hence, for instance, $u(x_1,x_2)$ and $d(x_o)$ will denote $u(x_o^o,x_1,x_2,x_3^o)$ and $d(x_o,x_1^o)$ respectively. Note, however, from (6.72) and (6.73) that

$$c(x_o^o,x_3^o) = 1 \text{ and } d(x_o^o,x_1^o) = 1 \quad . \tag{6.74}$$

Part (i). Substituting (6.73) into (6.72) and then (6.72) into (6.73) gives us, respectively,

$$u(\underline{x}) = u(x_o) + d(x_o)\ u(x_3) + c(x_o,x_3)[u(x_1) + d(x_1)\,u(x_2)] \quad , \tag{6.75}$$

and

$$u(\underline{x}) = u(x_o) + c(x_o)\ u(x_1) + d(x_o,x_1)[u(x_3) + c(x_3)\,u(x_2)] \quad . \tag{6.76}$$

Equating (6.75) and (6.76) with $x_3 = x_3^o$ indicates

$$d(x_o,x_1) = c(x_o)\ d(x_1) \quad , \tag{6.77}$$

which, together with (6.76) implies

$$u(\underline{x}) = u(x_o) + c(x_o)[u(x_1) + d(x_1)\ u(x_2,x_3)] \quad . \tag{6.78}$$

Evaluating (6.73) at $x_o = x_o^o$, we see

$$u(x_1,x_2,x_3) = u(x_1) + d(x_1)\ u(x_2,x_3) \quad ,$$

and so from (6.78) it follows that

$$u(\underline{x}) = u(x_o) + c(x_o)\ u(x_1,x_2,x_3) \quad . \tag{6.79}$$

Expression (6.79) says $\{X_1,X_2,X_3\}$ is utility independent of X_o.

Part (ii). Substituting (6.73) into (6.72) yields

$$u(\underline{x}) = u(x_o,x_3) + c(x_o,x_3)[u(x_1) + d(x_1)\ u(x_2)] \quad , \qquad (6.80)$$

which can be evaluated at $x_2 = x_2^o$ giving

$$u(x_o,x_1,x_3) = u(x_o,x_3) + c(x_o,x_3)\ u(x_1) \quad . \qquad (6.81)$$

Combining (6.80) and (6.81) and denoting $c(x_1,x_3)\ d(x_1)$ as $f(x_o,x_1,x_3)$, we find

$$u(\underline{x}) = u(x_o,x_1,x_3) + f(x_o,x_1,x_3)\ u(x_2) \quad , \qquad (6.82)$$

which says X_2 is utility independent of $\{X_o,X_1,X_3\}$.

Part (iii). Setting $x_o = x_o^o$ and $x_2 = x_2^o$ in (6.72) and (6.73) and equating indicates

$$u(x_3) + c(x_3)\ u(x_1) = u(x_1) + d(x_1)\ u(x_3) \quad , \qquad (6.83)$$

which can be rearranged to yield

$$\frac{c(x_3) - 1}{u(x_3)} = \frac{d(x_1) - 1}{u(x_1)} = k \quad , \quad u(x_i) \neq 0 \quad , \quad i = 1,3, \qquad (6.84)$$

where k is a constant since (6.84) has a function of x_3 equal to a function of x_1. If $u(x_1) = 0$, from (6.83), it follows that $d(x_1) = 1$ and similarly $c(x_3) = 1$ when $u(x_3) = 0$. Thus, from (6.84), one sees

$$c(x_3) = ku(x_3) + 1 \quad , \qquad (6.85)$$

and

$$d(x_1) = ku(x_1) + 1 \quad , \qquad (6.86)$$

which can be substituted into (6.75) with $x_o = x_o^o$, yielding

$$u(x_1,x_2,x_3) = u(x_3) + [ku(x_3) + 1][u(x_1) + [ku(x_1) + 1]\ u(x_2)]$$
$$= u(x_2) + [ku(x_2) + 1][u(x_1) + u(x_3) + ku(x_1)\ u(x_3)] \ . \tag{6.87}$$

Combining (6.87) and (6.79), it follows that

$$u(\underline{x}) = u(x_o) + c(x_o)[u(x_2) + [ku(x_2) + 1]\ u(x_1,x_3)]$$
$$= u(x_o,x_2) + g(x_o,x_2)\ u(x_1,x_3) \quad , \tag{6.88}$$

where $g(x_o,x_2) = c(x_o)\ [ku(x_2) + 1]$. Expression (6.88) proves the desired result that $\{X_1,X_3\}$ is utility independent of $\{X_o,X_2\}$.

Part (iv). We are given that $\{X_1,X_2\}$ is utility independent of its complement and part (iii) has shown $\{X_1,X_3\}$ is utility independent of its complement. Hence, from part (ii), it follows that the intersection X_1 is utility independent of its complement $\{X_o,X_2,X_3\}$. Similarly it follows that X_3 is utility independent of $\{X_1,X_1,X_2\}$. ◀

Theorem 6.7, which assumes utility independence conditions and concerns preferences for lotteries, closely parallels a result of Gorman [1968a] concerning preferences for consequences derived from preferential independence conditions. If each designation of the term utility independence in Theorem 6.7 were replaced by preferential independence, we would essentially have Gorman's result, which was presented in Chapter 3.*

6.9 Decomposition of Multiattribute Utility Functions

Roughly speaking, the more utility independence properties we can identify, the simpler the assessment of the utility function becomes. It is important to specify the simplest functional form of the multiattribute utility function consistent with an arbitrary set of utility independence assumptions. With this in mind, we want to generalize the results of Section 6.8 by constructing a "chaining theorem" using Theorem 6.7 as the building block. Let us illustrate with a simple example.

Example 6.3. Let $X \equiv \{X_1, X_2, X_3, X_4, X_5, X_6\}$ and suppose Y $Y_1 = \{X_1, X_2\}$ and $Y_2 = \{X_2, X_3, X_4\}$ are each UI. Then, by invoking Theorem 6.7 repeatedly we can show possible unions of X_1, X_2, and $\{X_3, X_4\}$ are also UI. In particular we know $\{X_1, X_2, X_3, X_4\}$ is UI.

Now suppose that in addition we find out that X_4 is UI. This may eventually help in assessing a utility function, but since X_4 does not overlap--as distinct from being contained in--any of the existing sets of UI attributes, the implications of X_4 being UI can not be further exploited. We may also find out that X_6 is UI, but since it does not overlap any UI attributes, it cannot be used to imply additional utility independence conditions.

---

*Since UI implies PI we could have exploited Gorman's results (see Theorem 3.7) in the proof of Theorem 6.7. But UI is such a strong condition that it enables us to present a straight forward algebraic proof directly.

However, if in addition it is determined that $Y_3 = \{X_4, X_5\}$ is UI, many implications follow from Theorem 6.7. Because $\{X_3, X_4\}$ is UI also, we know X_3, X_4, and X_5 are each UI. And as can be verified, given Y_1, Y_2, and Y_3 are UI, each possible union of the elements $X_1, X_2, \ldots, X_5$ is also UI. ∎

... To page 495

Given an arbitrary set of utility independence assumptions, say Y_j is UI of $\bar{Y}_j$, $j = 1,2,\ldots,J$, we want to exploit this to the maximum extent possible in structuring the resulting utility function. If $J = 2$, three possibilities relating Y_1 and Y_2 exist:

(1) Y_1 and Y_2 overlap,

(2) Y_1 and Y_2 are disjoint,

(3) Y_1 or Y_2 is contained in the other.

The previous section studied case (1). Here we want to investigate the generalization of this case when $J \geq 3$. The implication of cases (2) and (3) for $J \geq 3$, as well as combination of all three cases, will be considered in the remainder of the chapter.

Definition. A utility independent chain is a collection of sets $\{Y_1,\ldots,Y_R\}$, where (1) Y_j is UI, $j = 1,\ldots,R$, and (2) there is an ordering of Y_1 through Y_R such that each Y_j (other than the first in the ordering) overlaps at least one of its predecessors in the ordering.

We will be interested in finding utility independent chains which consist of as many sets as possible. This will allow us to exploit the utility independence properties to the fullest extent in simplifying the implied functional form of the utility function.

Definition. Let $\{Y_1,\ldots,Y_J\}$ be a set such that Y_j is UI, $j = 1,\ldots,J$ and let $\{Y_1,\ldots,Y_R\}$, $R \leqslant J$ be a utility independant chain. This chain is a maximal utility independant chain if no Y_j, $j = R + 1,\ldots,J$, overlaps any Y_j, $j = 1,\ldots,R$.

To gain some insight into this definition, let us construct a maximal utility independent chain, from Y_j, $j = 1, 2,\ldots,J$ where Y_j is UI. Select a Y_1 such that Y_1 is not contained in Y_j for any $j \geq 2$. Next we search for a Y_j, $j \geq 2$, such that Y_j overlaps Y_1. If no Y_j is found which satisfies this condition, then Y_1 by itself can be considered a maximal utility independent chain. Suppose Y_2 does overlap Y_1, then Y_1 and Y_2 are both members of a utility independent chain. The process now repeats using $\{Y_1,Y_2\}$ rather than Y_1.

We search for a Y_j, $j \geq 3$, such that Y_j overlaps at least one Y_k, $k = 1,2$. If no such Y_j is found, then the collection of sets $\{Y_1,Y_2\}$ is a maximal utility independent chain. If Y_3 does satisfy this condition, then $\{Y_1,Y_2,Y_3\}$ is formed and the process repeats until we have $\{Y_1,Y_2,\ldots,Y_R\}$ and no Y_j, $j \geq R + 1$, exists such that Y_j overlaps at least one Y_k, $k \leq R$. When this situation holds, the collection $\{Y_1,\ldots,Y_R\}$ is a maximal utility independent chain. Note that more than one maximal utility independent chain can exist on the set of attributes X.

<u>Definition</u>. Let $\{Y_1,Y_2,\ldots,Y_R\}$ be a maximal utility independent chain. Each Y_j, $j \leq R$, partitions $X \equiv \{X_1,X_2,\ldots,X_n\}$ into Y_j and $\bar{Y}_j$. There are 2^R possible subsets of X created by taking intersections formed with either Y_j or $\bar{Y}_j$ for each $j \leq R$. Thus, for instance, if $R = 3$, we have $Y_1Y_2Y_3$, $Y_1Y_2\bar{Y}_3$,

$Y_1\bar{Y}_2Y_3$, etc. Each intersection, except for $\bigcap_{j=1}^{R} \bar{Y}_j$, is defined to be an <u>element</u> of the maximal utility independent chain $\{Y_1,\ldots,Y_R\}$ if it is not empty.

An example should help illustrate our definitions.

<u>Example 6.4</u>. Consider the set $X = \{X_1,X_2,\ldots,X_8\}$, and suppose Y_j is UI, $j = 1,2,\ldots,5$, where

$$Y_1 \equiv \{X_1,X_2,X_3\},\ X_2 \equiv \{X_3,X_4,X_5\},\ Y_3 \equiv \{X_2,X_3\},$$
$$Y_4 \equiv \{X_5\},\ \text{and } Y_5 \equiv \{X_7,X_8\}.$$

Note that Y_2 overlaps Y_1 so $\{Y_1,Y_2\}$ is a utility independent chain. Now Y_3 is contained in Y_1 but Y_3 does overlap Y_2. Thus, Y_3 is added to $\{Y_1,Y_2\}$ forming $\{Y_1,Y_2,Y_3\}$, another utility independent chain. Checking Y_4, we see it is completely contained in Y_2 and distinct from both Y_1 and Y_3. Thus, the attribute Y_4 does not overlap any of Y_1, Y_2, or Y_3, so it does not enter the maximal utility independent chain we are constructing. Also Y_5 does not overlap any of Y_1, Y_2, or Y_3 implying that the collection of sets $\{Y_1,Y_2,Y_3\}$ is a maximal utility independent chain on X. In addition, Y_5 is itself another maximal utility independent chain on X.

To identify the <u>elements</u> of the maximal utility independent chain $\{Y_1,Y_2,Y_3\}$, we note $Y_1Y_2Y_3 = \{X_3\}$, $Y_1\bar{Y}_2Y_3 = \{X_2\}$, $Y_1\bar{Y}_2\bar{Y}_3 = \{X_1\}$, $\bar{Y}_1Y_2\bar{Y}_3 = \{X_4,X_5\}$, and $Y_1Y_2\bar{Y}_3$, $\bar{Y}_1Y_2Y_3$, and $\bar{Y}_1\bar{Y}_2Y_3$ are empty. Thus there are four elements of the chain, namely X_1,X_2,X_3, and $\{X_4,X_5\}$. For the maximal utility independent chain Y_5, there is the one element $\{X_7,X_8\}$. ■

Let us return to the general case and state an important result.

Theorem 6.8.* Each possible union of the elements in a maximal utility independent chain defined on $X = \{X_0, X_1, \ldots, X_n\}$ is utility independent of its complement in X.

The proof follows in three parts. Let us assume there are L elements $\{W_1, \ldots, W_L\}$ of the maximal utility independent chain $\{Y_1, \ldots, Y_R\}$ and define $Z \equiv \bigcup_{j=1}^{R} Y_j$, which can be thought of either as a collection of the X_i's, which are members of any Y_j, $j = 1, \ldots, R$, or as the set of elements $\{W_1, \ldots, W_L\}$. We first show the set Z is utility independent of its complement. Next, each subset of L-1 of the elements is shown to be utility independent of its complement. Then, using the intersection part of Theorem 6.7, it follows that each union of the elements is utility independent of its complement in X. The proof concerns maximal utility independent chains with three or more elements. The only other possibility involves chains with one element, in which case the theorem is valid by definition.

---

*This result, which explores the implications of maximal utility independent chains repeatedly invokes Theorem 6.7. A similar construction, using Theorem 3.7 as a basic tool, could have proven analogous implications for what we might have referred to as maximal preferential independent chains and their corresponding elements.

Proof. (Part 1) Let $\{Y_1,\ldots,Y_R\}$ be a maximal utility independent chain. By the manner in which the chain was constructed, Y_{k+1} intersects $\bigcap_{j=1}^{k} Y_j$ and hence, using the union part of Theorem 6.7, it follows that $\bigcap_{j=1}^{k+1} Y_j$ is UI. By induction we see that $Z \equiv \bigcap_{j=1}^{R} Y_j$ is UI.

(Part 2) To prove that each union of L-1 elements of the chain is UI, let us renumber the Y_j's so that the typical element of the chain, call it W, is defined by $W \equiv \bigcap_{j=1}^{r} Y_j \bigcap_{j=r+1}^{R} \bar{Y}_j$, where $1 \leq r \leq R$ and $\left[\bigcup_{j=1}^{t-1} Y_j\right] \cap Y_t$, $t = 2,\ldots,R$, is not empty. This renumbering will always be possible because of the manner in which Z was constructed. We wish to prove that Z-W is utility independent of its complement.

Either $\bigcap_{j=1}^{r} Y_j$ must be equivalent to W or it must be equivalent to $\{W, M_1,\ldots,M_s\}$ where $M_1,\ldots,M_s$ designate other elements. Allowing the M's to be null sets, the general case is $\bigcap_{j=1}^{r} Y_j = \{W, M_1,\ldots,M_s\}$. Consider two cases $r \geq 2$ and $r = 1$.

For $r \geq 2$, define $T_j = (Y_j \cup Y_{j+1}) - (Y_j \cap Y_{j+1})$, for $j = 1,2,\ldots,r-1$. By the symmetric difference part of Theorem 6.7, each T_j, $j = 1,2,\ldots,r-1$, is UI. Also, each T_{j+1} overlaps T_j because of the way they are defined. Hence, the union part of Theorem 6.7 implies $\bigcup_{j=1}^{r-1} T_j$ $= \left[\bigcup_{j=1}^{r} Y_j - \{W, M_1,\ldots,M_s\}\right]$ is UI.

If the M's are null sets clearly $\left[\bigcup_{j=1}^{r} Y_j - W\right]$ is UI. Since no Y_j, $j = r + 1,\ldots,R$ contains W, the construction $\left[\bigcup_{j=1}^{t} Y_j - W\right] \cup Y_{t+1}$ is equal to $\bigcup_{j=1}^{t+1} Y_j - W$, for all $t = r,\ldots,R-1$.

Taking successive unions in this manner and, since $\bigcup_{j=1}^{t} Y_j$ overlaps Y_{t+1}, invoking the union part of theorem 6.7, we find that the final construction $\left[\bigcup_{j=1}^{R} Y_j - W\right]$ is utility independent of its complement.

If $\{M_1,\ldots,M_s\}$ is not the null set, we again take successive unions using the Y_j's, $j = r + 1,\ldots,R$, beginning with the original construction $\left[\bigcup_{j=1}^{r} Y_j - \{W,M_1,\ldots,M_s\}\right] \cup Y_{r+1}$. None of the Y_j's, $j = r + 1,\ldots,R$ can contain W. However collectively $\bigcup_{j=r+1}^{R} Y_j$ must contain $\{M_1,\ldots,M_s\}$ since $W \equiv \bigcap_{j=1}^{r} Y_j \bigcap_{j=r+1}^{R} \bar{Y}_j = \{W,M_1,\ldots,M_s\} \bigcap_{j=r+1}^{R} \bar{Y}_j$, implying $\bigcap_{j=r+1}^{R} \bar{Y}_j$ does not contain $\{M_1,\ldots,M_s\}$. Taking the successive unions as described, we will again find $\left[\bigcup_{j=1}^{R} Y_j - W\right]$ is UI.

<u>For r = 1</u>, we have $W = Y_1 \bigcap_{j=2}^{R} \bar{Y}_j$ and since Y_j must contain at least two elements by the manner in which the chain was constructed, the general case is where $Y_1 = \{W,M_1,\ldots,M_s\}$. Each element M_k, $k = 1,\ldots,s$, must be contained in some Y_j, $j = 2,\ldots,R$. Otherwise, for instance, M_1 would only be in Y_1 so $\{W,M_1\} = Y_1 \bigcap_{j=2}^{R} \bar{Y}_j$ which implies

W is not an element.

Thus each element M_k, $k = 1,\ldots,s$, is in at least two Y_j's, $j = 1,\ldots,R$, and we have shown for this case that $\{Z - M_k\}$ is UI, $k = 1,\ldots,s$. Using the intersection part of Theorem 6.7, $\bigcap_{k=1}^{s} \{Z - M_k\} = \{Z - \bigcup_{k=1}^{s} M_k\}$ is UI. Now $\{Z - \bigcup_{k=1}^{s} M_k\} \cap Y_1 = W$, so by the symmetric difference part of Theorem 6.7, we find $\{Z - W\}$ is UI since $\{Z - \bigcup_{k=1}^{s} M_k\} \cup Y_1 = Z$.

(Part 3) From Part 2, each subset of L-1 of the elements $\{W_1,W_2,\ldots,W_L\}$ is utility independent of its complement in X. Hence, any proper subset of these W's is identical to the intersection of the appropriate sets of size L-1, and so by the intersection part of Theorem 6.7, all subsets of elements are utility independent of their complements. ◀

The relevance of Theorem 6.8 in structuring multi-attribute utility functions will be shown in the next section. To illustrate the power of Theorem 6.8, let us use it to prove Theorem 6.2 given in Section 6.3. For reference, the result is repeated here.

Theorem 6.2. Given attributes $X_1,X_2,\ldots,X_n$, the following are equivalent:

(1) attributes $X_1,X_2,\ldots,X_n$ are mutually utility independent,

(2) $\bar{X}_i$ is UI, $i = 1,2,\ldots,n$,

(3) $\{X_i,X_{i+1},\ldots,X_n\}$ is UI, $i = 2,3,\ldots,n$ and $\{X_1,X_2,\ldots,X_{n-1}\}$ is UI,

(4) $\{X_i, X_{i+1}\}$ is UI, $i = 1,2,\ldots,n-1$, $n \geq 3$,

(5) X_1 is UI and $\{X_1, X_i\}$ is PI, $i = 2,\ldots,n$, $n \geq 3$.

Proof. By definition, (1) implies (2) through (5). To prove the converses, we wish to show that $X_1, X_2, \ldots, X_n$ are each elements of a maximal utility independent chain encompassing $\{X_1, \ldots, X_n\}$ given any condition (2) through (5). Then the result directly follows from Theorem 6.8.

(2) $\Rightarrow$ (1). Note that $\bar{\bar{X}}_i = X_i$. Then $\left(\bigcap_{j\neq i} \bar{X}_i\right) \cap X_i = X_i$ is an element of the chain $\{\bar{X}_1, \bar{X}_2, \ldots, \bar{X}_n\}$.

(3) $\Rightarrow$ (1). The collection of sets $Y_i \equiv \{X_i, X_{i+1}, \ldots, X_n\}$, $i = 2,3,\ldots,n-1$ and the set $Y_n \equiv \{X_1, X_2, \ldots, X_{n-1}\}$ make a maximal utility independent chain. Note that $\left(\bigcap_{j=2}^{i} Y_j\right) \cap \left(\bigcap_{j=i+1}^{n-1} \bar{Y}_j\right) \cap Y_n = X_i$, $i = 1,2,\ldots,n-1$, is an element. Also $X_n = \left(\bigcap_{j=2}^{n-1} Y_j\right) \cap \bar{Y}_n$ is an element.

(4) $\Rightarrow$ (1). Let us define $Y_i = \{X_i, X_{i+1}\}$, $i = 1,2,\ldots,n-1$ so $\{Y_1, Y_2, \ldots, Y_{n-1}\}$ is a maximal utility independent chain. Then clearly $X_1 = Y_1 \cap \left(\bigcap_{j=2}^{n-1} \bar{Y}_j\right)$, and $X_i = Y_{i-1} \cap Y_i \cap \left(\bigcap_{j\neq i,i-1} \bar{Y}_j\right)$, $i = 2,3,\ldots,n-1$, and $X_n = \left(\bigcap_{j=1}^{n-2} \bar{Y}_j\right) \cap Y_{n-1}$ are elements.

(5) $\Rightarrow$ (1). By Theorem 6.6, $\{X_1, X_i\}$ is UI, $i = 2,3,\ldots,n$. Define $Y_i = \{X_1, X_i\}$, $i = 2,\ldots,n$, so that $\{Y_2, Y_3, \ldots, Y_n\}$ is a maximal utility independent chain. Then

$X_1 = \bigcap_{i=2}^{n} Y_i$ and $X_i = Y_i \cap \left(\bigcap_{j\neq i} \bar{Y}_i\right)$, $i = 2,\ldots,n$, are elements of the chain.

6.10. ADDITIONAL REPRESENTATION THEOREMS*

In the last three sections, we have looked at the implications of (i) a utility independence condition together with a preferential independence condition, (ii) two overlapping utility independence conditions, and (iii) an arbitrary number of utility independence conditions. Implications of (i) can be used to invoke (ii) or (iii) and implications of (ii) can be utilized with (iii). In this and the following sections, we'll try to integrate some of these ideas, and present some special results which are important. Proofs will usually not be given in detail since they either follow directly from or are similar to earlier ones. First we will look at extensions of the multiplicative and multilinear utility function.

6.10.1. Extension of the Multiplicative Form

The following is a straightforward extension of Theorem 6.1.

Theorem 6.9. Given the set of attributes $X \equiv \{X_o, X_1, \ldots, X_n\}$ where X_i, $i = 1,2,\ldots,n$, are elements of a maximal utility independent chain (this excludes X_o!), then

$$u(x_o,\bar{x}_o) = u(x_o,\bar{x}_o^o) + [u(x_o,\bar{x}_o^*) - u(x_o,\bar{x}_o^o)]\; u(x_o^o,\bar{x}_o),$$

and either (if $\sum_{i=1}^{n} k_i = 1$),

$$u(x_o^o,\bar{x}_o) = \sum_{i=1}^{n} u(x_i,\bar{x}_i^o) = \sum_{i=1}^{n} k_i u_i(x_i),$$

*This section contains specialized results that can be omitted at first reading.

or (if $\sum_{i=1}^{n} k_i \neq 1$)

$$1 + ku(x_o^o,\bar{x}_o) = \prod_{i=1}^{n} [1 + ku(x_i,\bar{x}_i^o)] = \prod_{i=1}^{n} [1 + kk_iu_i(x_i)],$$

where

(1) $\bar{x}_i = (x_o,x_1,\ldots,x_{i-1}, x_{i+1},\ldots,x_n)$, $i = 0,1,\ldots,n$,

(2) $u(x_o^o,x_1^o,\ldots,x_n^o) = 0$, $u(x_o^o,x_1^*,x_2^*,\ldots,x_n^*) = 1$,

(3) $u_i(x_i^o) = 0$, $u_i(x_i^*) = 1$, $i = 1,2,\ldots,n$,

(4) $k_i = u(x_i^*,\bar{x}_i^o)$, $i = 1,2,\ldots,n$,

and

(5) k is a scaling constant† which is a solution to

$$1 + k = \prod_{i=1}^{n} (1 + kk_i).$$

Proof. Using Theorem 6.1, plus the additional assumption that $\{X_1,X_2,\ldots,X_n\}$ is utility independent of X_o, implying

$$u(x_o,x_1,\ldots,x_n) = u(x_o) + c(x_o)\ u(x_1,x_2,\ldots,x_n),\ c > 0;$$

leads one directly to conclude either

$$u(\underline{x}) = k_ou_o(x_o) + [u(x_o,\bar{x}_o^*) - u(x_o,\bar{x}_o^o)] \left[\sum_{i=1}^{n} k_iu_i(x_i)\right], \tag{6.89}$$

or

$$u(\underline{x}) = k_ou_o(x_o) + [u(x_o,\bar{x}_o^*) - u(x_o,\bar{x}_o^o)] \times\left[\frac{1}{k}\left\{\left[\prod_{i=1}^{n} [1 + kk_iu_i(x_i)]\right] - 1\right\}\right], \tag{6.90}$$

which is the desired result. ◀

---

†Procedures for choosing the correct value of k are given in Appendix 6B at the end of this chapter.

6.10.2. Extension of the Multilinear Form

Expressions (6.89) and (6.90) provide forms of the utility function when there is one maximal utility independent chain. However, there are situations where more than one maximal utility independent chain may exist among the same set of attributes. For instance, let X be partitioned into $\{Z_o, Z_1, Z_2\}$ and suppose Z_i is utility independent of $\bar{Z}_i$, $i = 1,2$. That is, for instance, $Z_i \equiv \bigcup_{j=1}^{R} Y_j$ and $\{Y_1, \ldots, Y_R\}$ is a maximal utility independent chain. One can derive functional forms of utility functions involving more than one such chain by considering sets of utility independence assumptions over nonoverlapping attributes. With regard to this, we have

Theorem 6.10. Let $X = \{X_1, X_2, \ldots, X_n\}$ be partitioned into $\{Z_o, Z_1, \ldots, Z_M\}$ where Z_m, $m = 1,2,\ldots,M$, is utility independent. Then the utility function $u(\underline{x})$ can be represented by

$$u(\underline{x}) = g[z_o, u_1(z_1), u_2(z_2), \ldots, u_M(z_M)], \qquad (6.91)$$

where u_m, $m = 1,2,\ldots,M$, is a utility function over Z_m. The specific result is

$$u(\underline{x}) = u(z_o) + \sum_{m=1}^{M} f_m(z_o)u_m(z_m) + \sum_{\substack{m=1 \\ m<j\le M}}^{M} f_{mj}(z_o)u_m(z_m)u_j(z_j)$$

$$+ \ldots + f(z_o)u_1(z_1) \ldots u_M(z_M), \qquad (6.92)$$

where

$$f_m(z_o) = u(z_o, z_m^*, \bar{z}_{om}^o) - u(z_o, z_m^o, \bar{z}_{om}^o), \qquad (6.93a)$$

$$f_{mj}(z_o) = u(z_o, z_m^*, z_j^*, \bar{z}_{omj}^o) - u(z_o, z_m^*, z_j^o, \bar{z}_{omj}^o)$$
$$- u(z_o, z_m^o, z_j^*, \bar{z}_{omj}^o) + u(z_o, z_m^o, z_j^o, \bar{z}_{omj}^o), \quad (6.93b)$$

and

$$f(z_o) = u(z_o, \bar{z}_o^*) - \sum_{m=1}^{M} u(z_o, z_m^o, \bar{z}_{om}^*) + \sum_{\substack{m=1 \\ m<j\leq M}}^{M}$$
$$u(z_o, z_m^o, z_j^o, \bar{z}_{omj}^*) - \sum_{\substack{m=1 \\ m<j\leq M \\ j<k\leq M}}^{M} u(z_o, z_m^o, z_j^o, z_k^o, \bar{z}_{omjk}^*)$$
$$+ \ldots + (-1)^M u(z_o, z_1^o, \ldots, z_M^o). \qquad (6.93c)$$

Theorem 6.10 is the natural extension of the multilinear utility function. The distinction is that Z_o is not assumed to be utility independent of its complement.

One of the important facts to remember about Theorems 6.9 and 6.10 is that they can be used repeatedly in simplifying the expression for a multiattribute utility function. That is, the attributes designated by X_i in (6.89) and (6.90) and by Z_m in (6.92) may be vector attributes and possess utility independent properties among their respective components. If this is the case, then of course Theorems 6.9 and 6.10 can be used in specifying the relevant utility function $u_i(x_i)$ in (6.89) and (6.90) or $u_m(z_m)$ in (6.92). An example should help illustrate this, in addition to clarifying our definitions.

Example 6.5. Suppose we are interested in assessing a utility function u over the set of attributes $X \equiv \{X_1, X_2, \ldots, X_9\}$. Furthermore, suppose that it has been verified that Y_j is utility independent of $\bar{Y}_j$, $j = 1, 2, \ldots, 6$, where

$$
\begin{aligned}
Y_1 &= \{X_2, X_3\}, \\
Y_2 &= \{X_4, X_5, X_6\}, \\
Y_3 &= \{X_5\}, \\
Y_4 &= \{X_5, X_6, X_7, X_8\}, \\
Y_5 &= \{X_8\}, \\
Y_6 &= \{X_8, X_9\}.
\end{aligned}
$$

By our definition, there are two maximal utility independent chains in X which are $\{Y_1\}$ and $\{Y_2, Y_4, Y_6\}$. Attributes Y_3 and Y_5 are not in the chain Z_2 because $Y_3 \cap Y_j$ for $j = 2,4,6$ is either Y_3 itself or empty. The same is true for Y_5, so by the definition of a utility independent chain, Y_3 and Y_5 are excluded. Thus, we can define $Z_1 = Y_1$ and $Z_2 = Y_2 \cup Y_4 \cup Y_6$ and use (6.92) to write

$$
\begin{aligned}
u(\underline{x}) = u(z_o) &+ f_1(z_o)u_1(z_1) + f_2(z_o)u_2(z_2) \\
&+ f_{12}(z_o)u_1(z_1)u_2(z_2),
\end{aligned}
\tag{6.94}
$$

where u, u_1, and u_2 are scaled from zero to one.

There is clearly only one element $\{X_2, X_3\}$ in Y_1 but $\{Y_2, Y_4, Y_6\}$ has five elements: X_4, $\{X_5, X_6\}$, X_7, X_8, and X_9. We can use Theorem 6.9 to specify $u_2(z_2)$ further. For this purpose, we can assume $x_o = x_o^o$ in (6.89) and (6.90) so $u_o(x_o) = 0$ and $c_o(x_o) = 1$, and either

$$u_2(z_2) = k_4u_4'(x_4) + k_{56}u_{56}'(x_5,x_6) + k_7u_7'(x_7) + k_8u_8'(x_8) + k_9u_9'(x_9), \qquad (6.95)$$

or

$$u_2(z_2) = \frac{1}{k}\left\{\left[\prod_{i\varepsilon T} [1 + kk_iu_i'(x_i)]\right] - 1\right\}, \qquad (6.96)$$

where

$T = \{4, (5,6), 7, 8, 9\}$.

Considering only the attributes $\{X_5,X_6\}$, there is another utility independent chain, namely $Y_3 = \{X_5\}$. Hence, by Theorem 6.10,

$$u_{56}'(x_5, x_6) = u_6'(x_6) + f_6(x_6)u_5'(x_5), \qquad (6.97)$$

which can be substituted back into (6.95) or (6.96). The original assumption that Y_5 is utility independent of $\bar{Y}_5$ is redundant for this problem since $Y_5 = \{X_7\}$ is an element in the maximal utility independent chain $\{Y_2,Y_4,Y_6\}$, and by Theorem 6.8, each element of such a chain is utility independent of its complement. Combining (6.94) through (6.97) permits us to decompose $u(\underline{x})$ as far as possible consistent with the specified assumptions. ■

6.10.3. Special Multilinear Forms of the Utility Function

As one might expect, there are many sets of assumptions which are stronger than the utility independence assumptions of Theorem 6.3 and yet weaker than mutual utility independence assumption of Theorem 6.1. Let us illustrate the usefulness of exploring the additional restrictions placed on the utility function by various assumptions. As we will show, additional assumptions reduce the amount of empirical information necessary to specify u. Related results follow in Section 6.11 when we discuss preferences in hierarchical structures of attributes.

Throughout this subsection, we will assume $X \equiv \{X_1, X_2, \ldots, X_n\}$ and each is UI, $i = 1,2,\ldots n$, so that from Theorem 6.3, we know the utility function $u(x_1, x_2, \ldots, x_n)$ can be assessed from the n one-attribute utility functions $u_i(x_i)$ and $2^n - 2$ scaling constants. We could have been a bit more general and looked at the effect of additional assumptions used in connection with the assumptions of Theorem 6.10. However, since the ideas are analogous to that of adding additional assumptions to those of Theorem 6.3, and since the latter case is notationally more convenient, we chose it for illustration.

Y is Utility Independent of $\bar{Y}$. Let us assume $Y \equiv \{X_1, X_2, \ldots, X_m\}$. If Y is utility independent of $\bar{Y}$, then the attributes $Y, X_{m+1}, X_{m+2}, \ldots, X_n$ are a set of attributes, each of which is UI. Thus, Theorem 6.3 applies so the overall utility function can be assessed from (n-m+1) one-attribute utility functions: $u_Y(y)$, $u_{m+1}(x_{m+1}), \ldots, u_n(x_n)$, and $2^{n-m+1}-2$ scaling constants. But $u_Y(y)$ can be assessed by again applying Theorem 6.3 since each of $X_1, X_2, \ldots, X_m$ is utility independent. Therefore, the m utility functions $u_1(x_1)$, $u_2(x_2), \ldots, u_m(x_m)$ and $2^m - 2$ scaling constants will specify $u_Y(y)$. Putting this together, the original utility function of interest $u(x_1, x_2, \ldots, x_n)$ is now specified by the n one-attribute utility functions and $(2^{n-m+1} + 2^m - 4)$ scaling constants.

The usefulness of the additional assumption should be clear; it allows us to specify u with fewer scaling constants. This is so, since the assumption that Y is UI puts a set of consistency restrictions on the scaling constants of the multilinear form.

Y and $\bar{Y}$ are Mutually Utility Independent. Using the same notation as before, let us assume Y and $\bar{Y}$ are mutually utility independent. From Theorem 5.2, we know the utility function can be specified from $u_Y(y)$ and $u_{\bar{Y}}(\bar{y})$ and two scaling constants. Then from Theorem 6.3, it follows that $u_Y(y)$ is specified by m one-attribute utility functions: $u_1(x_1),\ldots,u_m(x_m)$, and $2^m - 2$ scaling constants. Similarly, $u_{\bar{Y}}(\bar{y})$ can be expressed from $u_{m+1}(x_{m+1}),\ldots,u_n(x_n)$ and $2^{n-m} - 2$ scaling constants. Therefore, subject to the additional mutual utility independence assumption, the utility function $u(x_1,x_2,\ldots,x_n)$ is given by $u_1(x_1)$, $u_2(x_2)$, $\ldots,u_n(x_n)$ and $2^m + 2^{n-m} - 2$ scaling constants.

With the mutliplicative and multilinear utility function, as well as the two cases considered above, the utility function u was specified by the n utility functions $u_1,u_2,\ldots,u_n$ and some number of scaling constants. The additional assumptions above allowed us to assess u with less constants than required in the multilinear case. Table 6.3 compares the number of constants required with and without the additional assumptions for some representative values of n and m, and thus gives an indication of the additional simplification in the assessment of u provided. In all cases we assume that X_i is UI, $i = 1,2,\ldots,n$.

Table 6.3

Number of Scaling Constants Required to Assess N-Attribute Utility Functions Given X_i is UI, $i = 1,2,\ldots,n$

	No additional assumptions (Multilinear Utility Function)	Assuming $Y \equiv \{X_1, X_2,\ldots,X_m\}$ is UI			Assuming $Y \equiv \{X_1, X_2,\ldots,X_m\}$ and $\bar{Y}$ are each UI			Assuming $X_1,X_2,\ldots,X_n$ are mutually utility independent (Multiplicative Utility Function)
		m=2	m=3	m=4	m=2	m=3	m=5	
n	2^n-2	2^{n-1}	$2^{n-2}+4$	$2^{n-4}-28$	$2^{n-2}-2$	$2^{n-3}+6$	$2^{n-5}+30$	n
3	6	4			4			3
4	14	8	8		6	8		4
5	30	16	12		10	10		5
6	62	32	20	32	18	14	32	6
7	126	64	36	36	34	22	34	7
8	254	128	68	44	66	38	38	8
9	510	256	132	60	130	70	46	9
10	1022	512	260	92	258	134	62	10

Other Sets of Assumptions. The additional assumptions which we've considered so far are the ones which would necessarily be "building blocks" for more involved sets of utility independence assumptions. To give just one more illustration, let us define Y as before and define Z as $\{X_r,X_{r+1},\ldots,X_n\}$.

Now let us assume Y and Z are utility independent of their complementary sets of attributes. There are three separate cases to consider, namely those where $r \leq m$, $r = m + 1$, and $r > m + 1$. We will just consider the most involved case where $r \leq m$ and there is some "overlapping" in our utility independence assumptions. Let us define the following notation to simplify our discussion:

$$Y_1 \equiv \{X_1, X_2, \ldots, X_{r-1}\},$$
$$Y_2 \equiv \{X_r, X_{r+1}, \ldots, X_m\}, \text{ and}$$
$$Y_3 \equiv \{X_{m+1}, X_{m+2}, \ldots, X_n\}.$$

In terms of this notation, we are assuming that $\{Y_1, Y_2\}$ is utility independent of Y_3 and $\{Y_2, Y_3\}$ is utility independent of Y_1, in addition to the original assumption that X_i is UI, $i = 1,2,\ldots,n$. When these assumptions hold, it follows from Theorem 6.1 that the utility function $u(y_1, y_2, y_3)$ is either additive or multiplicative, and thus one must assess utility functions over each of the Y_j's and assess three independent scaling constants. But the component attributes of each of the Y_j's are each UI, so the $u_j(y_j)$ terms can be assessed from a utility function for each component X_i in Y_j and $2^{b_j} - 2$ scaling constants, where b_j is the number of X_i's in Y_j.

Example 6.6. To illustrate the power of this result, suppose we have nine original attributes denoted by $X_1, X_2, \ldots, X_9$ and $Y_1 \equiv \{X_1, X_2, X_3\}$, $Y_2 \equiv \{X_4, X_5, X_6\}$, and $Y_3 \equiv \{X_7, X_8, X_9\}$. Then when our assumptions hold, we need to assess $u_1(y_1)$, $u_2(y_2)$, and $u_3(y_3)$ and three scaling constants. But each $u_j(y_j)$ requires we assess the respective three utility functions over the respective X_i's and $2^3 - 2 = 6$ scaling constants. Therefore, the overall utility function u requires we assess the nine component utility functions and 21 scaling constants. This can be compared to the 510 scaling constants necessary when only X_i is UI, $i = 1,2,...,9$. $\square$

6.10.4. The Additive Value Function and Multiplicative Utility Function*

It is interesting to relate the additive value function of Section 3.6 to the multiplicative utility function of Section 6.3 since necessary and sufficient conditions for the additive value function are necessary for the multiplicative utility function.

Theorem 6.11. Given:

(a) preferences over $X_1 \times X_2 \times \ldots \times X_n$ are compatible with an additive value function v,

(b) some X_i is UI (let it be X_1),

(c) $n \geq 3$,

then the utility function u must have one of the following three forms:

$$u(\underline{x}) \sim -e^{-cv(\underline{x})}, \quad c > 0 \quad , \tag{6.98a}$$

$$u(\underline{x}) \sim v(\underline{x}) \quad , \tag{6.98b}$$

$$u(\underline{x}) \sim e^{cv(\underline{x})}, \quad c > 0 \quad . \tag{6.98c}$$

---

*The ideas of the result in this section were generated by Richard F. Meyer and John W. Pratt.

[Note before proof: This result says that the utility function over the scalar attribute V, which measures value by v, must have constant risk aversion.]

Proof: Let us write v as

$$v(x_1,x_2,\ldots,x_n) = \lambda_1 v_1(x_1) + \lambda_2 v_2(x_2) + \ldots + \lambda_n v_n(x_n) \tag{6.99}$$

and scale v by

$$v(x_1^*,x_2^*,\ldots,x_n^*) = 1 \quad , \quad v(x_1^o,x_2^o,\ldots,x_n^o) = 0 \quad , \tag{6.100}$$

and similarly scale the v_i's by

$$v_i(x_i^*) = 1 \quad , \quad v_i(x_i^o) = 0 \quad , \quad i = 1,2,\ldots,n \quad , \tag{6.101}$$

so of course

$$\sum_{i=1}^{n} \lambda_i = 1 \quad . \tag{6.102}$$

The idea of the proof is simple: we deduce a utility function for the V-attribute and show that it must have constant risk aversion (see Theorem 4.15) from which the forms (6.98) follow.

Let $y \equiv (x_2,\ldots,x_n)$; in this notation we have $x = (x_1,y)$ and X_1 is UI of Y. For attribute X_1 let $\hat{x}_1 \sim < x_1^*,x_1^o >$ and therefore $(\hat{x}_1,y) \sim < (x_1^*,y), (x_1^o,y) >$ for all y. In terms of the V-attribute, this means that

$$[\lambda_1 v_1(\hat{x}_1) + \bar{v}(y)] \sim < \lambda_1 + \bar{v}(y), \bar{v}(y) > \quad , \tag{6.103}$$

where

$$\bar{v}(y) = \sum_{i=2}^{n} \lambda_i v_i(x_i).$$

In other words, adding $\bar{v}(y)$ to the prizes of the lottery $< \lambda_1, o >$ increases the certainty equivalent by $\bar{v}(y)$, for all $\bar{v}(y)$. This implies constant risk aversion for V has to be shown. ◀

Theorem 6.11 is interesting for two reasons:

(1) it provides for a simple procedure to obtain a multiattribute utility function given the necessary assumptions hold and given that an additive value function has been assessed, and

(2) the analysts can independently assess both a multiplicative (or additive) utility function and an additive value function and use one as a check against the other.

It is important to note that if the utility function is additive, then (6.98b) must hold, whereas if the utility function is multiplicative, either (6.98a) or (6.98c) must be valid.

Given v, the assessment of u is straightforward. Simply assess the certainty equivalent $\hat{x}_1$ for the lottery $< x_1^*, x_1^o >$. Then if

$$v_1(\hat{x}_1) = \frac{1}{2} v_1(x_1^*) + \frac{1}{2} v_1(x_1^o) \quad , \qquad (6.104)$$

the utility function must be the additive case (6.98b). If

$$v_1(\hat{x}_1) \neq \frac{1}{2} v_1(x_1^*) + \frac{1}{2} v_1(x_1^o) \quad , \qquad (6.105)$$

then (6.98a) is the proper form if the left side of (6.105) is less than the right side and case (6.98c) is the appropriate utility function when the right side is smaller. In either case, by setting the utility of $\hat{x}_1$ equal to $< x_1^*, x_1^o >$ using

(6.98a) or (6.98c) and solving, the scaling constant c is determined.

6.11. HIERARCHICAL STRUCTURES AND CONDITIONAL PREFERENCES

Suppose that the attributes for a particular problem have been structured as shown in Figure 6.4. Furthermore, suppose that Y_1 and Y_2 are mutually utility independent. Then from Theorem 5.2, we know

$$u(y_1,y_2) = k_1u_1(y_1) + k_2u_2(y_2) + k_{12}u_1(y_1)u_2(y_2), \tag{6.106}$$

where all utility functions are scaled from zero to one. Note that by evaluating (6.106) at y_1^o and at y_2^o, the respective least preferable amounts of Y_1 and Y_2, we find

$$u_1(y_1) = \frac{u(y_1,y_2^o)}{k_1} \quad \text{and} \quad u_2(y_2) = \frac{u(y_1^o,y_2)}{k_2}. \tag{6.107}$$

The point is that u_1 and u_2 are actually conditional utility functions over their respective domains given a fixed level of the other attribute. Because of utility independence, the conditional utility function over Y_1, for example, is the same regardless of the level of Y_2. That is why we only need one conditional utility function for Y_1 in specifying $u(y_1,y_2)$.

The logical next step in assessing u would be to try to identify functions f_1 and f_2 such that

$$u_1(y_1) = f_1[u_1'(x_1),u_2'(x_2)]$$

and

$$u_2(y_2) = f_2[u_3'(x_3),u_4'(x_4)u_5'(x_5)],$$

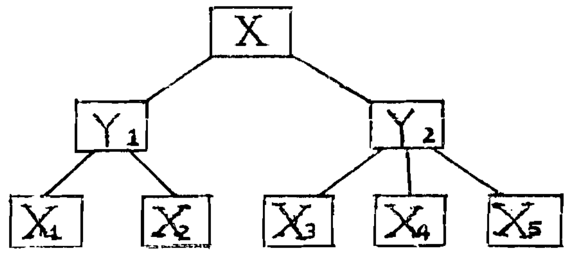


Figure 6.4. A Hierarchy of Attributes for a Hypothetical Problem

where the u's are utility functions over their respective domains. Using previous results, this could be done if X_i is UI, $i = 1,2,\ldots,5$. However, because of the dimensionality, it might be difficult to verify such assumptions. Fortunately, we don't need such strong assumptions. Because Y_1 is utility independent of Y_2, we can just worry about whether X_1 is conditionally utility independent of X_2 given Y_2 is set at y_2^o, for instance.

More generally, in all the formulations of the previous two chapters, once we have determined that $Y \subset X \equiv \{X_1,X_2,\ldots,X_n\}$ is utility independent of $\bar{Y}$, we could then speak of preferences and utility functions over subsets of attributes included in Y without considering the levels of the attributes within $\bar{Y}$. The latter can be specified at some convenient level. With this motivation, we can define a number of useful conditional preference concepts.

6.11.1. Conditional Independence Assumptions

We are interested in concepts of conditional independence for three reasons:

(1) simplifying the structure of a multiattribute utility function provided certain conditional independence assumptions are met;

(2) representing necessary conditions for indpendence assumptions to hold, and thus, in some cases, requiring less empirical questioning to find in fact that they do not hold;

(3) representing sufficient conditions for independence conditions to hold, thus providing for weaker assumptions, and consequently less verification, to imply utility functions of particular forms.

After defining our terms, we will expand on each of these.

To formulize these ideas, consider the set of attributes $X \equiv \{X_1, X_2, \ldots, X_n\}$ which will be partitioned into three non-empty subsets Y_1, Y_2, and Y_3. We will say Y_1 is <u>conditionally preferentially independent</u> of Y_2 given y_3^+ if the preference order for consequences involving only changes of attribute levels in Y_1 does not depend on the level of Y_2 when Y_3 is fixed at y_3^+. Mathematically this condition is

$$\left[u(y_1', y_2', y_3^+) \geq u(y_1'', y_2', y_3^+)\right] \Rightarrow \left[u(y_1', y_2, y_3^+) \geq u(y_1'', y_2, y_3^+)\right],$$

$$\forall y_2. \tag{6.108}$$

Similarly, we define Y_1 to be <u>conditionally utility independent</u> of Y_2 given y_3^+ if the preference order for lotteries involving only changes of attribute levels in Y_1 does not depend on the level of Y_2 when Y_3 is fixed at y_3^+. This condition can be represented mathematically as

$$u(y_1, y_2, y_3^+) = c(y_2) + d(y_2)\, u(y_1, y_2', y_3^+),\ d(y_2) > 0, \tag{6.109}$$

where y_2' is arbitrarily chosen. These definitions follow naturally from our original ones of preferential and utility independence.

There is a generalization of each of these definitions. We will say Y_1 is conditionally preferentially independent of Y_2 given Y_3 if the preference order for consequences involving only changes in attribute levels in Y_1 does not depend on the level of Y_2 when Y_3 <u>is fixed at any level</u>. In the same manner, we define Y_1 to be conditionally utility independent of Y_2 given Y_3 if the preference order for lotteries involving only changes of attribute levels in Y_1 does not depend on the level Y_2 when Y_3 is fixed at any level. These conditionaly preference assumptions can be written respectively as

$$\left[u(y_1',y_2',y_3) \geq u(y_1'',y_2',y_3)\right] \Rightarrow \left[u(y_1',y_2,y_3) \geq u(y_1'',y_2,y_3)\right], \quad \forall y_2,y_3 \tag{6.110}$$

and

$$u(y_1,y_2,y_3) = f(y_2,y_3) + g(y_2,y_3)\, u(y_1,y_2', y_3),\ g(y_2,y_3) > 0, \tag{6.111}$$

where y_2' refers to an abitrary but fixed level of Y_2.

It is clear that (6.110) implies (6.108) and (6.111) implies (6.109), so the latter conditional preference conditions are stronger than the former. Note that the relative preferences over Y_1 given y_3' need not be the same as the relative preferences over Y_1 given y_3'' for condition (6.111) to hold. If Y_1 is conditionally utility independent of Y_2 given Y_3 and if the relative preferences over Y_1 are the same for all values of Y_3, then we find in fact that Y_1 is utility independent of $\{Y_2,Y_3\}$. Hence for any y_2' and y_3',

$$u(y_1,y_2,y_3) = d_1(y_2,y_3) + d_2(y_2,y_3)\, u(y_1,y_2',y_3'). \tag{6.112}$$

In Figure 6.5 we try to graphically illustrate how utility independence and conditional utility independence relate to each other. Condition (6.109), that Y_1 is conditionally utility independent of Y_2 given y_3^+, means the relative preferences over each of the heavy solid lines in Figure 6.5 are strategically equivalent. This means the conditional utility functions over each of these solid lines are the same except for positive linear transformations. This condition does not mean the relative preferences over the heavy dashed lines must be the same. However, they may be. The condition (6.111) that Y_1 is conditionally utility independent of Y_2 given Y_3 means, for instance, that the relative preferences over each of the heavy solid lines must be the same, that the relative preferences over each of the heavy dashed lines must be the same, and that the relative preferences over each of the dot-dash lines must be the same. It does not require that the relative preferences over the solid lines be the same as those over the dashed lines or the dot-dash lines. When in fact, the relative preferences over each of the heavy lines--solid, dashed, and dot-dash--are the same, then it is very likely that Y_1 is utility independent of $\{Y_2,Y_3\}$. "Very likely" is used here since the condition must also hold for all planes not drawn in the figure in addition to those where y_3 is y_3^+, y_3', or y_3'' in order for Y_1 to be utility independent of $\{Y_2,Y_3\}$.

Finally let us define conditional additive independence.

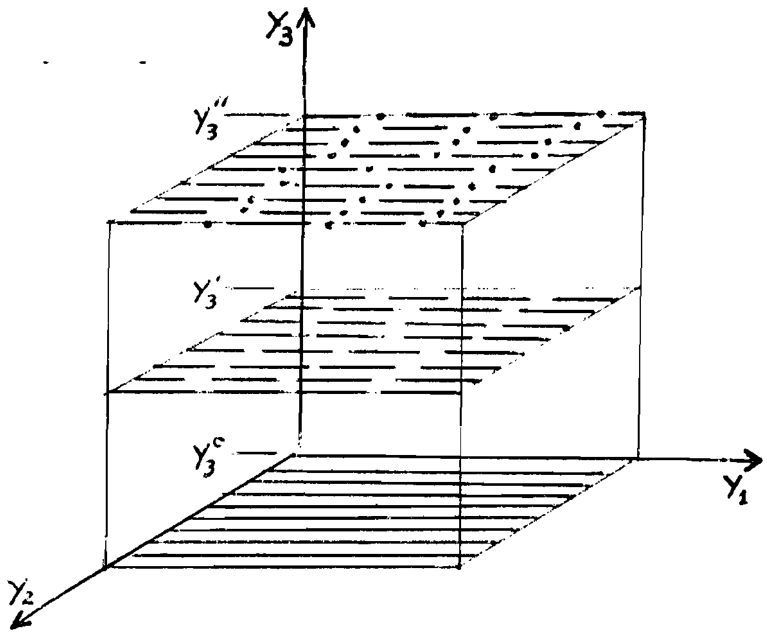


Figure 6.5. Relationships Among Conditional Utility Independence Assumptions and Utility Independence

Attributes Y_1 and Y_2 are conditionally additive independent given y_3' if preferences over lotteries on Y_1 and Y_2 given that Y_3 is fixed at y_3' depend only on their marginal conditional probability distributions and not on their joint conditional probability distribution. And similar to the previous cases, we define attributes Y_1 and Y_2 to be conditionally additive independent given Y_3 if preferences over lotteries on Y_1 and Y_2, given any fixed level of Y_3, depend only on their marginal conditional probability distributions and not on their joint conditional probability distribution.

6.11.2. Simplifying the Structure of Multiattribute Utility Functions

Now we can begin to look at the usefulness of the conditional preference assumptions. For most of the theorems using preferential, utility, or additive independence, analogous results could be derived with the corresponding conditional independence assumptions. We will state a few of these without proofs since they are very similar to proofs found earlier in the book. For example, corresponding to Theorem 5.2, we have the

Theorem 6.12. If Y_1 and Y_2 are conditionally utility independent of each other given y_3^o, then

$$u(y_1,y_2,y_3^o) = u(y_1,y_2^o,y_3^o) + u(y_1^o,y_2,y_3^o) + k u(y_1,y_2^o,y_3^o)\, u(y_1^o,y_2,y_3^o), \tag{6.113}$$

where $u(y_1^o,y_2^o,y_3^o) = 0$ and k is an empirically evaluated constant. The proof is analogous to that of Theorem 5.2.

Aside from those results with direct analogies using utility independence, additional results can be proven, such as Theorem 6.13. If Y_1 and Y_2 are conditionally utility independent of each other given Y_3, then $u(y_1,y_2,y_3)$ can be specified by assessing $u_1(y_1,y_2^o,y_3)$, $u_2(y_1^o,y_2,y_3)$, and $u_3(y_1^*,y_2^*,y_3)$ for arbitrary y_1^o,y_1^*,y_2^o,y_2^*, subject to consistent scaling of the u_i.

The result allows us to assess a three-attribute utility function by assessing three conditional utility functions--one with one attribute and two with two attributes. The reasoning behind Theorem 6.13 is illustrated using Figure 6.6, where it is assumed that Y_1, Y_2, and Y_3 are scalar attributes. The consequences whose preferences must be assessed are shaded in the figure. Suppose we want to get the utility of an arbitrary point (y_1',y_2',y_3'), illustrated as point A. Since Y_1 is conditionally utility independent of Y_2 given y_3', the utility of A can be expressed in terms of the utilities of B and C since the relative preferences between A, B, and C, are the same as those between A', B', and C', and the latter are known. The utility of C is also known, but the utility of B is not known. However, since Y_2 is conditionally utility independent of Y_1 given y_3', the utility of B can be expressed in terms of the utilities of B' and D since the relative preferences between B', B, and D are the same as those between C', C, and D' and the latter are known. Since the utilities of B' and D are assessed, we can calculate the utility of B and hence the utility of an arbitrary consequence A.

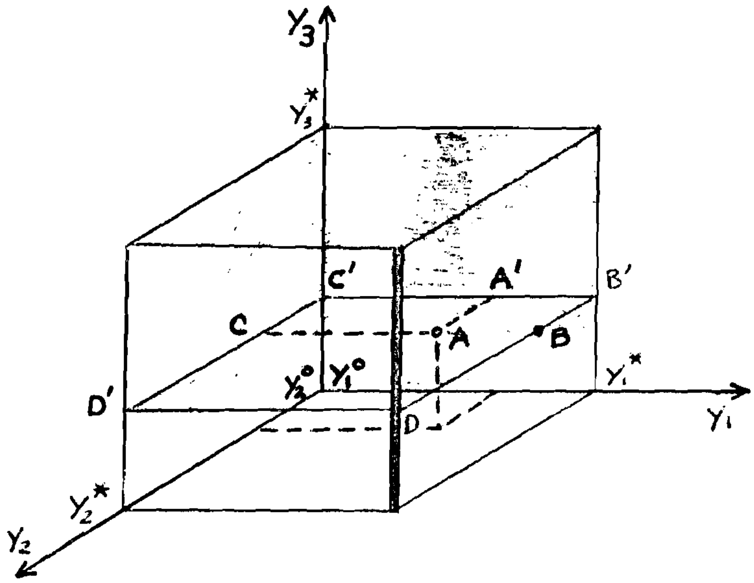


Figure 6.6. Illustration of a Proof for Theorem 6.13

One more result indicative of the usefulness of conditional independence concepts in specifying the structure of a utility function is

Theorem 6.14. If Y_1 and Y_2 are conditionally additive independent given Y_3, then

$$u(y_1,y_2,y_3) = u(y_1,y_2^o,y_3) + u(y_1^o,y_2,y_3) - u(y_1^o,y_2^o,y_3), \tag{6.114}$$

where

$$u(y_1^o,y_2^o,y_3^o) = 0.$$

The proof is very similar to that of Theorem 5.1. This result allows us to specify the three-attribute utility function from two consistently scaled two-attribute utility functions. If Y_1, Y_2, and Y_3 each represent scalar attributes, then (6.114) says we only need to assess a utility function over the two shaded planes of Figure 6.6 in order to completely specify u.

6.11.3. Necessary Conditions of Independence Assumptions

Let us go on to the second area in which conditional independence assumptions are useful. The ideas discussed here are trivial from an analytical point of view, but helpful from a practical point of view, so they are included. It may be quite difficult in some situations to determine whether or not Y_1 is utility independent of $\{Y_2,Y_3\}$. However, if we hold the value of Y_3 fixed and check relative preferences over Y_1 for various values of Y_2 given y_3^+ and if these relative preferences are not the same, then clearly the relative preferences over Y_1 cannot be the same for all (y_2,y_3) pairs. Hence Y_1

could not be utility independent of $\{Y_2, Y_3\}$. To formalize this, we state the mathematically trivial but useful

Theorem 6.15. A necessary condition for Y_1 to be utility independent of $\{Y_2, Y_3\}$ is that Y_1 be conditionally utility independent of Y_2 given Y_3.

In a similar spirit, we state without proof

Theorem 6.16. A necessary condition for Y_1, Y_2, and Y_3 to be additive independent is that Y_1 and Y_2 be conditionally additive independent given Y_3.

6.11.4. Sufficient Condtions for Independence Assumptions

A third use of conditional utility independence is that it provides us with the tools to state sufficient sets of assumptions about preference independence properties. Thus less empirical validation is necessary to verify that a particular form of utility function is appropriate for a particular problem. In this regard we prove

Theorem 6.17. If Y_1 is conditionally utility independent of Y_2 given Y_3 and if Y_1 is conditionally independent of Y_3 given y_2^o, then Y_1 is utility independent of $\{Y_2, Y_3\}$.

Proof: Since Y_1 is conditionally utility independent of Y_2 given Y_3, from (6.111) for arbitrary Y_2, which we will choose as y_2^+,

$$u(y_1, y_2, y_3) = d_1(y_2, y_3) + d_2(y_2, y_3)\, u(y_1, y_2^+, y_3). \qquad (6.115)$$

And since Y_1 is conditionally utility independent of Y_3 given y_2^+, from (6.109)

$$u(y_1,y_2^+,y_3) = c_1(y_3) + c_2(y_3)\ u(y_1,y_2^+,y_3^+). \qquad (6.116)$$

Substituting (6.116) into (6.115), we find

$$u(y_1,y_2,y_3) = d_1(y_2,y_3) + d_2(y_2,y_3)\ [c_1(y_3)+c_2(y_3)u(y_1,y_2^+,y_3^+)]$$

$$= f_1(y_2,y_3) + f_2(y_2,y_3)\ u(y_1,y_2^+,y_3^+), \qquad (6.117)$$

where

$$f_1(y_2,y_3) = d_1(y_2,y_3) + d_2(y_2,y_3)\ c_1(y_3)$$

and

$$f_2(y_2,y_3) = d_2(y_2,y_3)\ c_2(y_3).$$

Equation (6.117) implies Y_1 is utility independent of $\{Y_2,Y_3\}$. ◀

A particularly important class of problems concern those with a hierarchical structure of attributes. Some useful results pertaining to this follow.

Theorem 6.18. If $\{Y_1,Y_2\}$ is utility independent of Y_3 and Y_1 is conditionally preferentially independent of Y_2 given y_3', then Y_1 is preferentially independent of $\{Y_2,Y_3\}$.

Proof: The utility independent condition implies

$$u(y_1,y_2,y_3) = u(y_3) + c(y_3)\ u(y_1,y_2,y_3^o), \qquad (6.118)$$

and the conditional preferential independence assumption means

$$[u(y_1',y_2',y_3') \geq u(y_1'',y_2',y_3')] \Rightarrow [u(y_1',y_2,y_3') \geq (y_1'',y_2,y_3')]. \qquad (6.119)$$

Evaluating the right side of (6.119) with (6.118) yields

$$u(y_1',y_2,y_3^o) \geq u(y_1'',y_2,y_3^o). \qquad (6.120)$$

Hence, it follows by substituting (6.120) into (6.118) that

$$u(y_1', y_2, y_3) \geq u(y_1'', y_2, y_3), \ \forall y_2, y_3, \tag{6.121}$$

which means that Y_1 is preferentially independent of $\{Y_2, Y_3\}$. ◀

There is the analogous result involving utility independence.

Theorem 6.19. If $\{Y_1, Y_2\}$ is utility independent of Y_3 and Y_1 is conditionally utility independent of Y_2 given y_3', then Y_1 is utility independent of $\{Y_2, Y_3\}$.

Proof: The assumptions mean

$$u(y_1, y_2, y_3) = f(y_3) + g(y_3)\, u(y_1, y_2, y_3''), \tag{6.122}$$

where y_3'' is arbitrarily chosen, and

$$u(y_1, y_2, y_3') = c(y_2) + d(y_2)\, u(y_1, y_3'). \tag{6.123}$$

Setting $y_3'' = y_3'$, substituting (6.123) into the right-side of (6.122), and rearranging terms yields

$$u(y_1, y_2, y_3) = f(y_3) + g(y_3)\, c(y_2) + g(y_3)\, d(y_2)\, u(y_1, y_3'). \tag{6.124}$$

Evaluating (6.124) at (y_1, y_2^o, y_3^o), solving for $u(y_1, y_3')$, and substituting this result back into (6.124) proves the desired result. ◀

The previous two results allow us to independently focus on the decision-maker's preferences over utility independent chains and their elements without worrying about the levels of the other attributes once they have been fixed at any convenient level. For instance, in Example 6.5 of Section 6.10, we did not need the condition that X_5 was utility independent of $\bar{X}_5$ in order to arrive at (6.97). By using Theorem 6.19, we only needed that X_5 was conditionally utility independent of X_6 given $\bar{X}_{56}$ was fixed at some convenient level $\bar{x}_{56}$. This latter condition would be much easier to verify than the former.

Concerning additive independence, we have

Theorem 6.20. If (i) Y_1 and Y_2 are conditionally additive independent given Y_3, (ii) Y_1 and Y_3 are conditionally additive independent given y_2^o, and (iii) Y_2 and Y_3 are conditionally additive independent given y_1^o, then Y_1, Y_2, and Y_3 are additive independent.

Proof: Conditions (ii) and (iii) imply, respectively, that

$$u(y_1,y_2^o,y_3) = u(y_1,y_2^o,y_3^o) + u(y_1^o,y_2^o,y_3) \tag{6.125}$$

and

$$u(y_1^o,y_2,y_3) = u(y_1^o,y_2,y_3^o) + u(y_1^o,y_2^o,y_3), \tag{6.126}$$

where

$$u(y_1^o,y_2^o,y_3^o) = 0.$$

Thus, substituting (6.125) and (6.126) into (6.114), which follows from condition (i), we find

$$u(y_1,y_2,y_3) = u(y_1,y_2^o,y_3^o) + u(y_1^o,y_2,y_3^o) + u(y_1^o,y_2^o,y_3). \tag{6.127}$$

Expression (6.127) is the additive utility function from which additive independence directly follows. ◀

6.11.5. An Example Illustrating the Hierarchical Structure

We will propose a simplified version of a regulation problem typical of those facing the various segments of government to illustrate some of the ideas introduced here. In particular, suppose a state is considering passing legislation requiring that seat belts be worn by all state highway users. The overall objective of such a program is state to "improve the well-being" of motorists in the state. Subobjectives are to minimize physical harm to motorists and to keep monetary costs as low as possible. Thus we might define our overall attribute X as "well-being," where Y_1 is "physical harm" and Y_2 is "monetary costs." Furthermore, suppose that Y_1 is broken into attributes X_1 and X_2 representing deaths and serious injuries, respectively, and that Y_2 is broken into attributes X_3 and X_4 representing costs to motorists and costs to the state, respectively. The measures of effectiveness which will be used for each of the attributes are listed in Table 6.4. Figure 6.7 should be useful in illustrating the hierarchical structure of the attributes.

Our next step in an analysis, and the step of interest here, would be to assess a utility function $u(\underline{x})$. Clearly $u(\underline{x})$ can also be written as $u(y_1,y_2)$ or $u(x_1,x_2,x_3,x_4)$. A reasonable place to begin to structure u might be with the additive independence assumption discussed in Section 5.3. The first place we check is whether this condition holds for attribute Y_1 and Y_2 and suppose it does not. But we do verify that Y_1 is preferentially independent of Y_2 and that X_1 is utility independent of $\{X_2,Y_2\}$. Then from Theorem 6.6, Y_1 is utility independent of Y_2. Suppose we also find that Y_2

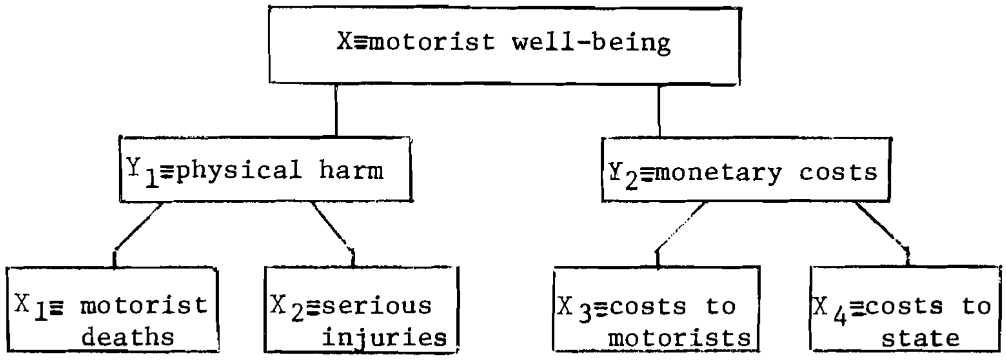


Figure 6.7. Hierarchical Structure of Attributes for the Seat Belt Problem

is utility independent of Y_1, so from Theorem 5.2,

$$u(y_1,y_2) = u(y_1,y_2^o) + u(y_1^o,y_2) + ku(y_1,y_2^o)\, u(y_1^o,y_2),$$

where (6.128)

$$u(y_1^o,y_2^o) = 0.$$

Table 6.4

Attributes and Measures for the Seat Belt Problem

Attribute	Measures of Effectiveness
$X_1 \equiv$ motorist deaths	annual number of highway deaths in state
$X_2 \equiv$ motorist serious injuries	annual number of highway serious injuries
$X_3 \equiv$ monetary costs to motorists	dollar cost to install seat belts in a car
$X_4 \equiv$ monetary costs to state	annual dollar cost to maintain program

We wish to go further if possible and simplify both $u(y_1,y_2^o)$ and $u(y_1^o,y_2)$. Taking these in order, suppose we ascertain X_1 and X_2 are conditionally mutually utility independent given y_2^o. From the fact X_1 was utility independent of $\{X_2, Y_2\}$, we knew this held for X_1. It follows from Theorem 6.12 that

$$u(y_1,y_2^o) \equiv u(x_1,x_2,y_2^o)$$

$$= u(x_1,x_2^o,y_2^o) + u(x_1^o,x_2,y_2^o) + k_1 u(x_1,x_2^o,y_2^o)\, u(x_1^o,x_2,y_2^o),$$

where the origin is still set by $u(y_1^o,y_2^o) \equiv u(x_1^o,x_2^o,y_2^o) = 0.$ (6.129)

Notice the scaling of u has not yet been specified.

Now we go on to $u(y_1^o, y_2)$, which for purposes here will be written $u(y_1^o, x_3, x_4)$. After checking, suppose we can conclude only X_4 is conditionally utility independent of X_3 given y_1^o. This means a result analogous to Theorem 5.6 in Section 5.6 is valid and

$$u(y_1^o,x_3,x_4) = u(y_1^o,x_3,x_4^o)\ [1-u(y_1^o,x_3^o,x_4)] + u(y_1^o,x_3,x_4')\ u(y_1^o,x_3^o,x_4), \tag{6.130}$$

where the origin is set by $u(y_1^o,x_3^o,x_4^o) = 0$ and the scale by $u(y_1^o,x_3^o,x_4') = 1$.

Since the three utility functions of (6.128), (6.129), and (6.130) all have the same origin and the scale is only specified in (6.130), we can directly substitute (6.129) and (6.130) into (6.128) to get an expression for $u(x_1,x_2,x_3,x_4)$ in terms of $u(x_1,x_2^o,x_3^o,x_4^o)$, $u(x_1^o,x_2,x_3^o,x_4^o)$, $u(x_1^o,x_2^o,x_3,x_4^o)$, $u(x_1^o,x_2^o,x_3^o,x_4)$, $u(x_1^o,x_2^o,x_3,x_4')$, k, and k_1 where the origin and scale of u are set by

$$u(x_1^o,x_2^o,x_3^o,x_4^o) = 0$$

and

$$u(x_1^o,x_2^o,x_3^o,x_4') = 1,$$

respectively. Thus, in this example, by exploiting independence and conditional independence conditions, the assessment of the four-attribute utility function u has been simplified to the consistent assessment of five one-attribute conditional utility functions and two additional scaling constants. This means seven scaling constants are required in all--one each to specify a second point on each of the conditional utility functions which already have the same origin plus k and k_1.

6.12. SUMMARY

This chapter develops representation theorems that are suitable for a decision-maker's utility function given various sets of assumptions about his basic preferences. Relationships amongst various assumptions are investigated with two purposes in mind: (1) to weaken the assumptions necessary to imply particular forms of utility function, and (2) to understand and exploit fully all the implications of an arbitrary set of preference assumptions. An oversimplified summary of the results of Chapters 3, 5, and 6 is given in Table 6.5.

As one begins to generalize the two-attribute results of Chapter 5 to more than two attributes, one fact is apparent. In three or more dimensions, the richness of possible sets of preferential and utility independence assumptions increases greatly for two reasons: first, the existence of independence conditions over overlapping attributes, and second, the existence of conditional preferential and utility independence assumptions. We have attempted to illustrate this richness by the results stated.

For most real problems, we would expect that collectively, the techniques presented in chapters 3 through 6 should

Table 6.5. Independence Assumptions in Multiattribute Utility Theory
Given: Consequences $x \equiv (x_1, x_2, \ldots, x_n) \equiv (y,z)$

Assumption	Preferential Independence (PI)	Utility Independence (UI)	Additive Independence (AI)
Concerns	Preferences for consequences (y,z) with z held fixed	Preferences for lotteries over over $(\tilde{y},z)$ with z held fixed	Preferences for lotteries over $(\tilde{y},\tilde{z})$ with both y and z varying
Definition	Y is PI of Z if preferences for consequences (y,z') with z' fixed don't depend on the amount z'	Y is UI of Z if preferences for lotteries on $(\tilde{y},z')$ with z' fixed don't depend on the amount z'	Y and Z are AI if preferences for lotteries $(\tilde{y},\tilde{z})$ depend only on the marginal probability distributions on y and z
Mathematical Implication of Definition	$[u(y^a,z')>u(y^b,z')] \rightarrow [u(y^a,z)>u(y^b,z)]$, all z	$u(y,z) = f(z) + g(z)\, u(y,z^o)$, $g>0$ all z	$\langle(y',z'),(y^o,z^o)\rangle \sim \langle(y',z^o),(y^o,z')\rangle$, all y',z'
Note:	PI is not reflexive; Y PI Z does not imply Z PI Y	UI is not reflexive	AI is reflexive
Main Results: Two attributes		If Y UI Z and Z UI Y, then $u(y,z) = k_Y u_Y(y) + k_Z u_Z(z) + kk_Y k_Z u_Y(y) u_Z(z)$	If Y and Z are AI, then $u(y,z) = k_Y u_Y(y) + k_Z u_Z(z)$
Main Results: n attributes	If $\{X_1,X_i\}$ PI $\overline{X}_{1i}$, $i=2,3,\ldots,n$, then $v(x) = \sum_{i=1}^{n} v_i(x_i)$ where v's are all value functions	If X_i UI $\overline{X}_i$, $i=1,2,\ldots,n$, then $u(x)=\sum k_i u_i(x_i) + \sum_{j>i} k_{ij} u_i(x_i) u_j(x_j) + \ldots + k_{1\ldots n} u_1(x_1) \ldots u_n(x_n)$	If X_i AI $\overline{X}_i$, $i=1,2,\ldots,n$, then $u(x) = \sum_{i=1}^{n} k_i u_i(x_i)$
Combining PI and UI	If $\{X_1,X_i\}$ PI $\overline{X}_{1i}$, $i=2,3,\ldots,n$ and X_1 UI $\overline{X}_1$, then either (1) $1+ku(x) = \prod_{i=1}^{n} [1+kk_i u_i(x_i)]$ or (2) $u(x) = \sum_{i=1}^{n} k_i u_i(x_i)$		

significantly help to specify a "reasonable" representation of the decision-maker's preferences, provided that the problem has been structured with meaningful objectives and attributes in the sense discussed in Chapter 2. A number of researchers, including Yntema and Torgerson [1961], Fisher [1972 , 1973], von Winterfeldt and Edwards [1973a, 1973b], and Dawes and Corrigan [1974], have results which lend support to this claim. The cases discussed in Chapters 7 and 8 provide additional supporting evidence.

Appendix 6A

Generalization of Preferential Independence and Utility Independence

Suppose that we have vector attributes Y and Z, and that given Z is z^o, there is a definite preference for different levels of Y. If this is quantified by the value function $v(\underline{y},\underline{z})$, then if

$$v(\underline{y}',\underline{z}^o) \geq v(\underline{y}'',\underline{z}^o) \Rightarrow v(\underline{y}',\underline{z}) \geq v(\underline{y}'',\underline{z}), \quad \forall\, \underline{z},$$

we define Y to be preferentially independent of Z. Thus, given a $\underline{z}$, the conditional preference order of $\underline{y}$ is the same regardless of the $\underline{z}$ chosen. If the reverse order on $\underline{y}$ occurs for some $\underline{z}'$, that is

$$v(\underline{y}',\underline{z}^o) \geq v(\underline{y}'',\underline{z}^o) \rightarrow v(\underline{y}'',\underline{z}') \geq v(\underline{y}',\underline{z}'),$$

we will say conditional preferences on $\underline{y}$ given $\underline{z}^o$ and $\underline{z}'$ are reversals of each other. One could also have indifference on Y given Z is some other level $\underline{z}''$. We will say that Y is <u>generalized</u> <u>preferentially</u> <u>independent</u> of Z if, given any two levels of Z, say $\underline{z}'$ and $\underline{z}''$, the two orderings of $\underline{y}$ given $\underline{z}''$ are either identical, reversals of each other, or indifference exists among the $\underline{y}$. A bit more mathematically, Y is generalized preferentially independent of Z if,

$$v(\underline{y},\underline{z}) = f(\underline{z}) \left[v(\underline{y},\underline{z}^o)\right],$$

where the only restriction on $f(\underline{z})$ is that it is scalar valued. If $f > 0$, we have the case where Y is preferentially independent of Z.

In a similar manner, we can have reversals of preferences among lotteries over Y for different amounts of Z. If Y is <u>generalized</u> <u>utility</u> <u>independent</u> of Z, then

$$u(\underline{y},\underline{z}) = g(\underline{z}) + h(\underline{z})u(\underline{y},\underline{z}^o),$$

where $\underline{z}^o$ is chosen so that there is a definite conditional preference on Y given $\underline{z}^o$ and h can be negative, zero, or positive. If $h(\underline{z}')$ is negative, then preferences over lotteries on Y given $\underline{z}'$ are a reversal of the order on these lotteries given $\underline{z}^o$. Of course, when $h > 0$, then we have the utility independent case considered in detail in the chapter.

In Fishburn and Keeney [1974, 1975], it is shown that results analogous to many in this chapter can be derived using the weaker generalized preferential independence and generalized utility independent condition rather than preferential and utility independence. Some results, however, do not follow due to the reversals of preferences.

Appendix 6B

Evaluating the Scaling Constant k in the Multiplicative Utility Function

Taking $\underline{x} = \underline{x}^*$ in (6.14) yields

$$k + 1 = \prod_{i=1}^{n} (1 + kk_i). \qquad (6B.1)$$

By evaluating (6.12) at x_i^* and $\bar{x}_i^*$ we have $k_i = u(x_i^*)$, and

$$ku(\bar{x}_i^*) + 1 = \prod_{j \neq 1} (1 + kk_j) = \frac{k+1}{kk_i+1}$$

or

$$1 = k_i + [1 + kk_i]\, u(\bar{x}_i^*). \qquad (6B.2)$$

Since $k_i < 1$ and the u-value above is positive, it follows from (6B.2) that

$$1 + kk_i > 0. \qquad (6B.3)$$

Comparing the signs of the two sides of (6B.1), we infer

$$k > -1. \qquad (6B.4)$$

Now let $S = \sum_{i=1}^{n} k_i$, and introduce the polynomial

$$f(q) = 1 + q - \prod_{i=1}^{n} (1 + k_i q), \quad (-1 < q < \infty), \qquad (6B.5)$$

so that (6B.1) says that $f(k) = 0$; note also that $f(-1) < 0$. Differentiating (6B.5) gives

$$1 - f' = \sum_{i=1}^{n} k_i \prod_{i \neq j} (1 + k_j q), \qquad (6B.6)$$

which shows that 1-f' is an increasing function and hence f' is decreasing.

First suppose $S = 1$, i.e., $f'(0) = 0$. Then since f' is decreasing in $(-1,\infty)$, it is positive in $(-1,0)$ and negative in $(0,\infty)$. Thus $q = 0$ is the only root of $f(q) = 0$ in $(-1,\infty)$, and so $S = 1$ implies $k = 0$ and corresponds only to the additive utility function.

Next suppose $S < 1$, i.e., $f'(0) > 0$. Then since f' is decreasing, it is positive in $(-1,0)$, so that $f(q) = 0$ has no root between (-1) and the root at 0. It follows from (6B.6) that $f'(\infty) = -\infty$, and so $f'(q) = 0$ has a unique root q^* in $(0,\infty)$. Since $f(0) = 0$ and $f' > 0$ in (o,q^*), $f(q) = 0$ has no root in $(0,q^*)$. Since $f(q^*) > 0$, and f' is negative and decreasing to $(-\infty)$ in (q^*,∞), $f(q) = 0$ has a unique root k in (q^*,∞); moreover $f > 0$ in $(0,k)$ and $f < 0$ in (k,∞), so that the iterative method described in the text works provided the search for k is confined to $(0,\infty)$.

Finally, suppose $S > 1$, i.e., $f'(0) < 0$. Since f' is decreasing, it is negative in $(0,\infty)$, so that $f(q) = 0$ can have no root to the right of the root at $q = 0$. Since $f > 0$ immediately to the left of this root, while $f(-1) < 0$, there must be at least one root k of $f(q) = 0$ in $(-1,0)$; since f' is decreasing and $f(0) = 0$, there can be at most one such root, and the iterative method described in the text is valid, provided the search for k is confined to $(-1,0)$.

Appendix 6C

An Interactive Computer Program for Assessing and Using Multiattribute Utility Functions*

Section 6.6 discussed and illustrated the considerations necessary for assessing multiattribute utility functions. The task is difficult and the current state-of-the-art of unaided empirical assesement (i.e., with the lack of direct computer support) has some shortcomings. The most important of these are as follows:

(1) the necessity to ask 'extreme value' questions to keep the computational requirements for specifying a utility function to a manageable level,

(2) the tedium of calculating the component utility functions and scaling constants even in this case,

(3) the lack of immediate feedback to the decision-maker of the implications of his preferences,

(4) the absence of an efficient procedure to 'update' the decision-maker's preferences and conduct sensitivity analysis.

This section describes the major features of a computer package designed to alleviate the above shortcomings with existing methods for the assessment and use of multiattribute utility functions. The package is referred to by the mnemonic MUFCAP standing for "multiattribute utility function

---

* This section was adapted from Keeney and Sicherman [1975]. In Section 9.7, an analogous interactive computer program is described which is designed for intertemporal tradeoffs.

calculation and assessment package." At present, some of the subroutines in the package are rather crude. However, the package is operational, and as a first try, indicates a worthwhile direction to proceed.

6C.1. Applicability of the Functional Forms

In terms of the required assessments and general robustness, the additive and multiplicative utility functions appear to be the practical ones for say $n \geq 4$. Even when the requisite assumptions do not precisely hold over the domains of all the attributes, it may be a good approximation (see von Winterfeldt and Edwards [1973b]) to assume they do or it may be reasonable to integrate different additive and multiplicative utility functions over separate regions of these attributes. Furthermore, by nesting one multiattribute utility function inside another, a technique described in the next paragraph, additional flexibility in the preference structure can be achieved.

The results of Theorem 6.1 (multiplicative utility function) and Theorem 6.4 (additive utility function) are valid regardless of whether the X_i's are scalar attributes or vector attributes. This means that the x_i's can be either scalars or vectors. In the former case, the component utility functions u_i are single-attribute utility functions, whereas in the latter case, u_i is itself a multiattribute utility function. If X_i is a vector attribute, it is possible, subject to satisfying the requisite assumptions, to reuse Theorem 6.1 or 6.4 in structuring u_i. In such a case, we will say u_i is a nested multiattribute utility function. That is, u_i is a multiattribute utility functicn nested within the multiattribute utility function u.

Nesting multiplicative forms provides an extra degree of freedom in the problem by having an extra independent scaling constant. Without nesting, using the multiplicative utility function, the number of independent scaling constants is equal to the number n of attributes. However, suppose the last single utility function u_n is a multiplicative utility function nested within the overall utility function and that u_n has three single attributes. Then one would need n scaling constants for the "outer multiplicative utility function" and three for the "inner multiplicative utility function" for a total of n + 3, even though there are only n + 2 single attributes, $X_1,\ldots,X_{n-1}$ and the three single attributes in u_n. The degree of freedom afforded by the extra parameter permits tradeoffs between two attributes to be dependent on a third. This allows for some violation of the preferential independence conditions. By various nesting schemes, enough extra constants could be provided to model situations in which tradeoffs between many pairs of attributes depend on the level of other attributes. The additive and multiplicative utility functions are simple enough to be tractable and yet, especially with nesting, robust enough to adequately quantify preferences for many problems. In practice, however, assessing and using these multiattribute utility functions is "easier said than done."

In all that follows, we will assume that the assumptions implying that the multiattribute utility function is either additive or multiplicative have been verified. Also, since the procedures for assessing the basic components, the u_i's and k_i's, of both the multiplicative and additive utility functions are essentially the same, there is no need to consider the cases separately. Recall that the extra constant k in the multiplicative form is calculated directly from the k_i's. For illustration, the multiplicative form will be used for both the overall utility function u and any nested utility function. In the remainder of this section, we summarize the MUFCAP package. Details and a listing of the program are found in Sicherman [1975]. The abbreviation MUF will mean a multiattribute utility function of either multiplicative or additive form.

6C.2. Commands to Structure the Utility Function

Structuring a utility function consists of specifying a functional form, its attributes, and the ranges for each of the attributes. MUFCAP has several commands for structuring a utility function. The INPUT command requests a name for the utility function and asks for the number of attributes which are arguments of this function. The package then requests a name and a range for scalar attributes. This consists of two numbers which bound the amounts to be considered for each attribute. To specify a vector attribute, one inputs a range with one bound equal to the other bound such as 0, 0. MUFCAP recognizes this as a signal for a vector attribute and notes that the u_i associated with that attribute is a nested MUF. The package then requests the number of attributes which are arguments of this nested MUF. For each of these a name and range will be solicited. Further levels of nesting could be specified if desired and the information requested would be analogous to the material above. After a nested MUF is completely specified, the program returns to ask for the names and ranges for whatever attributes have not yet been covered in the outer MUF. When all the attributes have been input, the structure is complete and MUFCAP requests a new command from the user.

The INPUT command provides for all the bookkeeping which will be necessary for information to follow. Each k_i and u_i, including those in a nested MUF, can be accessed using the name of the attribute with which it is associated. The INPUT command is quite flexible in having no limit to the degree of nesting allowed.

In addition to INPUT, the package has commands for adding or deleting attributes to or from the utility function. It also has a command for switching the order of the attributes in a utility function. In this way, attributes may be conveniently "regrouped" to alter the model for the problem in terms of different nesting schemes.

6C.3. Commands to Specify the Single Attribute Utility Functions

The next step in assessing a MUF involves specifying the u_i's for the single attributes. As noted in Chapter 4, sophisticated computer programs do exist for assessing single (scalar) attribute utility functions. One could incorporate these into MUFCAP. Initially, however, for simplicity in the current package, a subroutine for assessing undimensional utilty functions, referred to as UNIF was developed.

MUFCAP has available commands to specify conveniently three types of unidimensional utilty functions: linear, exponential, and piecewise linear. As indicated in Chapter 4, the linear utility function implies risk neutrality. This form requires no more information than the range of the attribute. The exponential forms implies constant risk aversion or constant risk proneness. It requires the specification of a certainty equivalent for a single lottery. Given this, the exponential form is fitted and scaled automatically by the program. The piecewise linear utility function is specified by providing the abscissa and ordinate values for n points ($3 \leq n \leq 15$) of the utility function. This form can be used for non-monotonic or S-shaped utility functions. These three types provide the user with the means of conveniently specifying his preferences for many situations. More forms can easily be added to the package in the future.

MUFCAP also has commands which enable a user to display any assessed unidimensional utility function to check its appropriateness. The command UNICAL calculates the utility for one or a series of attribute levels. INVERSE calculates the attribute level corresponding to a given utility. LOTTERY evaluates the certainty equivalent for any lottery with n consequences and their associated probabilities over that attribute, where $2 \leq n \leq 15$. When there are two consequences, LOTTERY can also calculate the probability which will make the lottery indifferent to a given certainty equivalent.

To summarize, MUFCAP has convenient commands to assess u_i's which are unidimensional utility functions and to examine their implications as a check on their reasonableness.

6C.4. Commands to Specify the Scaling Constants

Using the attribute names as identifiers, MUFCAP allows the user to set the scaling constants in the MUF corresponding to each attribute. If X_i is a vector attribute, the u_i associated with it is a MUF with its own internal scaling constants. By referring to the name of this vector attribute, the user can specify the internal scaling constants for the associated nest MUF. When all the k_i's for a particular MUF have been set, the program automatically calculates the corresponding k (see Appendix 6B).

Once u_i's have been evaluated, the package has several commands useful for assessing the k_j's in any particular MUF. The command INDIF2 takes as input two pairs of two indifference consequences each. These consequences can vary only in terms of the two attributes, say X_j and X_m. Their scaling constants k_j and k_m are the object of assessment. Using the MUF and the single-attribute utility functions, the program computes the relative value of k_j and k_m implied by the indifference pairs. With INDIF2, the user is not limited to choosing consequences which have one attribute at a least desirable level in order to determine the relative k_i's.

Given the information from INDIF2, indifference curves over attributes X_j and X_m can be calculated with the command IMAP. IMAP permits a user to get immediate feedback on the implications of the relative k_i's which he has specified. He can quickly see if the points "claimed" to be indifferent really appear so to him. If not, the relative k_i's can be changed until they represent the user's preferences for tradeoffs between those attributes.

Once we know the relative k_i's, the command INDIF1 takes as input a single pair of indifference consequences and computes the k and the absolute magnitude of the k_i's implied by that pair and the relative k_i's. For consistency checks, a new indifference pair of consequences can be input into INDIF1, which then computes the factor by which the current k_i's need to be multiplied to be consistent with the indifference point just given. MUFCAP provides a routine which allows the user to multiply the currently assigned k_i's for any MUF by any factor. In this way, INDIF1 enables the calculation of the magnitude of the k_i's using an indifference relation instead of a lottery over all the attributes at once.

6C.5. Commands for Evaluating Alternatives and Sensitivity Analysis

Once the u_i's and k_i's have been set, the utility function is completely specified and can be used. To help explore the implications of the utility function and to perform 'rough' analysis, MUFCAP has commands for specifying two kinds of alternatives; those with certainty and those with uncertainty. For 'certainty' alternatives, which are simply consequences, uniattribute amounts are solicited until the alternative is completely described. For 'uncertainty' alternatives, at present, MUFCAP assumes probabilistic independence and requests a probability distribution function for each single attribute. The probability distribution function currently used is a piecewise linear approximation to the cumulative probabilty distribution for X_i. The user supplies n abscissa-ordinate pairs, where $2 \leq n \leq 9$ to specify the cumulative distribution. Then MUFCAP calculates the expected utilities for probabilistic alternatives. The cumulative distribution was chosen rather than the probability density function because the fractile method of assessing probabilities (see Schlaifer [1969]) yields points of the cumulative distribution. Other forms of probability distributions such as the Gaussian as well as probabilistic dependencies could be added to the package in the future.*

The specified alternatives are given names by the user. With these names, the user may add, change, or delete alter-

---

*It is easy to use Monte Carlo techniques to find the expected utility values for dependent probability distributions. The Monte Carlo routine would generate a sequence of $\{\underline{x}^{\alpha} : \alpha = 1,2,\ldots\}$ and one would then compute $\frac{1}{T} \sum_{\alpha=1}^{T} u(\underline{x}^{\alpha})$ for T large.

natives. He may also choose the ones which are to be evaluated by listing their names with the appropriate commands about to be described.

The command EVAL is used to evaluate (i.e. compute the expected utility for) any alternative or group of alternatives. By specifying a group of alternatives differing slightly in in some feature, one can conduct a sensitivity analysis of the probabilistic inputs. Also, EVAL will compute the expected utilities for any multiattribute utility function specified in the command. Thus, using EVAL, one can conduct a sensitivity analysis of the preference structure by varying parameters, such as the scaling constants, in the multiattribute utility function. In this same way, different utility functions of members of a decision-making group can be used to evaluate and rank the alternatives. This might help clarify differences of opinion and suggest certain creative compromises if needed.

The command GRAD evaluates the gradient of a utility function at any number of specified consequences. The gradient is defined as the vector $\left(\frac{\partial u}{\partial x_1}, \frac{\partial u}{\partial x_2}, \ldots, \frac{\partial u}{\partial x_n}\right)$ and indicates the direction of the steepest increase in the utility function at a specified point. The gradient components tells us which attribute level changes would yield large increases in utility. This could be useful in generating worthwhile alternatives. Of course, one must keep in mind the scales of the attributes in interpreting the gradient.

In addition to the gradient, GRAD also computes the vector $\left(\frac{\partial u}{\partial u_1}, \frac{\partial u}{\partial u_2}, \ldots, \frac{\partial u}{\partial u_n}\right)$. Each component represents the rate of change of u with respect to a change in the utility u_i. These components reveal the attributes for which an increase in its utility will yield the largest increase in u. The advantage of calculating these quantities in addition to the gradient components are (a) components can be calculated for MUF's as well as unidimensional utility functions, and (b) the unit of measurement for a uniattribute does not distort the magnitude of the component. Thus in some cases, $\frac{\partial u}{\partial u_i}$ might better indicate possible improved alternatives than $\frac{\partial u}{\partial x_i}$. MUFCAP makes both available.

Summarizing, EVAL permits the evaluation of alternatives, and along with routines which alter paramters, provides for sensitivity analysis. GRAD makes use of the analytical formulation of the problem to calculate quantities useful in suggesting alternatives which might be better than the ones currently specified.

6C.6. General Command Format and Commands for Facilitating Use of the Package

MUFCAP has the facility for saving the current status of the multiattribute utility structure and the current alternatives in a file of the user's choosing to be read in at a later time. This gives MUFCAP the capability for filing away several different MUF models as well as a large number of alternatives for the same problem. It also allows the user to build up his model over many different sessions at the terminal and restore any status he has saved away with which he wishes to calculate at any particular time.

Another feature of MUFCAP is the supplying of default settings when the INPUT command is used to structure the MUF for the problem. After INPUT, the default for all MUF's is the additive form, with all the k_i's equal to each other, and for all unidimensional utility functions, it is the linear utility function. With these defaults, the user is set to calculate immediately after input. Thus feedback can begin right away without requiring the user to completely specify everything first. Scaling constants and utility functions can then be altered after observing some feedback to refine the model for the problem.

Finally, MUFCAP provides commands to print out the current status of the assessments. There are routines to display the k_i's and k for any MUF, the range and type for any single-attribute utility function, the probability distribution of any attribute for any alternative, the multiattribute utility function structure (i.e. nesting), and the currently defined alternatives. Commands are also provided for easily changing parameters such as individual k_i's or the components of any alternative.

6C.7. Summary

The current version of MUFCAP provides the basic features necessary to assess and use multiattribute utility functions on complex decision problems. In particular, it permits one to use realistic and simple questions in assessing the decision-maker's preferences, rather than the 'difficult to think about' types of questions previously used for computational reasons. MUFCAP provides for (1) a variety of immediate feedback of implications of the decision-maker's responses, (2) evaluation of alternatives and sensitivity analysis, and (3) analyzing differences of preferences and judgments among various individuals in a decision-making group.

The present MUFCAP should be considered a first edition, a basis on which to improve. Some possible improvements of existing routines have been suggested in this section such as a more sophisticated single-attribute utility function assessment technique and potential for evaluating alternatives where probabilistic independence need not be assumed. The program could then be easily coupled with simulation models producing such probability distributions. Other important improvements would inlcude the addition of new routines (1) to help in verifying preferential and utility independence assumptions, (2) to simplify sensitivity analysis and feedback, perhaps with the aid of graphical displays, and (3) to conduct conflict analyses in problems involving more than one decision-maker.

DECISION ANALYSIS WITH MULTIPLE CONFLICTING OBJECTIVES PREFERENCES AND VALUE TRADEOFFS

(Chapters 7 & 8)

Ralph L. Keeney and Howard Raiffa

May 1975 WP-75-53

CHAPTER 7

ILLUSTRATIVE APPLICATIONS

Certainly we are dealing in this book with a non-vacuous problem: many difficult, real-world decision problems do involve multiple objectives. Consequently, many of the concepts we have introduced are relevant and must be applied in either a formal or informal analysis of the alternatives. If one chooses to analyze multiple objectives and value or utility tradeoffs in a formal manner, then immodestly, we believe the ideas and procedures discussed in this book can often be of considerable use. The purpose of this chapter and Chapter 8 is to support this contention by illustrating many cases where multiattribute preferences have been formalized. The present chapter, in a variety of settings, focuses exclusively on the preference assessments themselves whereas Chapter 8, which concerns the site selection of an airport for Mexico City, presents a complete case including probability assessments, analysis of alternatives, interactions with the decision makers, and so on, as well as multiattribute preference assessments.

The applications discussed in this chapter cover the range of topics presented in Chapter 2 through 6. Section 7.1 discusses the generation of objectives and the specification of measures of effectiveness for an air-pollution problem. Section 7.2 discusses the allocation of resources for an educational program and the value functions of the members of a local school board and other local education officials are formalized. Next, a five-attribute utility function for response times of various fire trucks is assessed. This problem typically arises in planning operations of emergency services. Section 7.4 addresses the problem of structuring corporate preferences. In sections 7.5 and 7.6, we discuss preliminary work on the quantification of multiattribute preferences concerning decisions involving the selection of computer systems and decisions about the siting and licensing of nuclear power facilities.

The first six sections of this chapter relate in some depth experiences that we and others have had in assessing multiattribute preferences. The last section, 7.7, gives brief surveys of a number of other problems where formal analyses have explicitly considered multiple objectives using concepts discussed in earlier chapters. These include: utilization of frozen blood, sewage sludge disposal, safety of landing aircraft, choice of a job, shipments of hazardous materials, medical and surgical treatment of cleft lip and palate.

Our thesis is that the concepts and procedures introduced in this book are not just of theoretical, but also of operational interest and they can be--and have been--utilized to make contributions in a variety of important contexts. Many analysts are currently applying decision analysis to such crucial problems as those discussed in this chapter and the inventory of case studies is growing rapidly.

7.1 AIR POLLUTION CONTROL*

In New York City, the mayor must decide whether he should approve a proposed major addition to Consolidated Edison's electric power generating station in Astoria, Queens. If this addition is approved, City residents would be reasonably assured of receiving the growing quantity of electricity they will demand over the next several years at reasonable cost. However, approval would result in increased air pollution, particularly in terms of sulfur dioxide, particulates, and nitrogen oxides. Should this addition be approved?

In both Boston and New York City, the respective City Councils must decide whether to pass legislation that would place stringent limits on the sulfur content of fuels burned in the city. If passed, the legislation would lead to a definite improvement in the city's air quality--especially in terms of the air pollutant: sulfur dioxide.

---

*This section draws heavily on the dissertation of Ellis [1970] and adapts material from Ellis and Keeney [1972]. A related dissertation by Mead [1973] goes into more depth on the Astoria problem. Both dissertations were supervised in part by Raiffa.

However, passage of this legislation would require residents to incur added annual costs for heating and electricity to pay for the more expensive fuels with low sulfur contents. Should these City Councils pass such legislation?

In Washington, D.C., the U.S. Congress must decide whether to establish very stringent emission standards for carbon monoxide, hydrocarbons, and nitrogen oxides for all motor vehicles manufactured and sold in the United States. Establishment of these standards would contribute toward improving the air quality. On the other hand, they would require the public to pay significantly more money for new automobiles. Should Congress adopt these stringent standards?

Each of these decision problems is faced presently or has been faced recently by public officials. Moreover, they are representative of a host of similar problems that public officials increasingly confront. The basic question is: Should government adopt a specific, proposed program intended to improve the air quality? With each such investigation there is the additional question: "What should the air quality standard be?"

The major focus of this book has been to describe how a decision maker--in this case a public official--can utilize decision analysis to help make up his mind: how to select a desirable cause of action amongst the myriad of alternatives he confronts. In this section, we focus our attention on the selection of a set of objectives and measures of effectiveness for analyzing governmental programs designed to better control air pollution. We draw heavily upon the concepts discussed in Chapter 2.

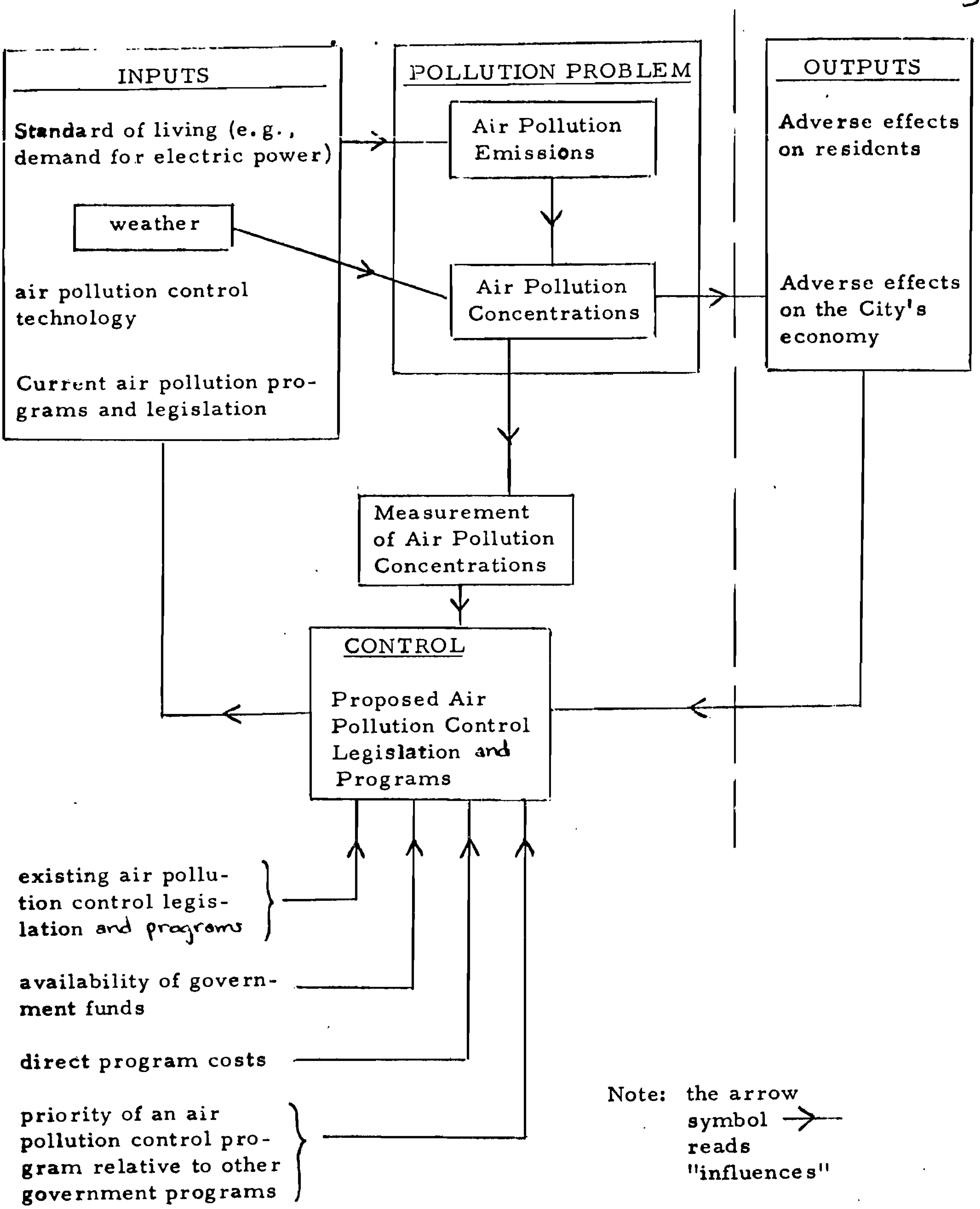
As a vehicle for illustrating or suggestions, focus is placed on one specific problem faced by one particular individual, the Mayor of New York City. Obviously, we would not expect the Mayor of New York to spend his time working on details of the air pollution problem. It would be reasonable, however, to expect members of the Mayor's staff in the Environmental Protection Administration and the Department of Air Resources to work on this problem. These individuals and the Mayor might then review the results and implications of such analyses in formulating and supporting air-pollution control programs for New York City.

In the next subsection, a brief overview of the air pollution control problem in New York City is presented, along with an introduction to the sulfur-dioxide problem. Then, objectives and measures of effectiveness are generated for the analysis of the problem. To avoid leaving the reader in midstream, the final subsection briefly sketches other aspects of this problem that were examined.

7.1.1 The Air Pollution Control Problem of New York City

A general model of the process by which many air pollution control programs are designed and evaluated is shown in Figure 7.1. The main problem with the control process as it is currently practiced in most municipal governments is that the outputs are usually not explicitly considered in choosing air pollution policy. The reason is, of course, understandable. There are simply too many complexities: the difficulties in defining appropriate output measures, in establishing the relationships between pollution concentrations and these measures, and in specifying preferences for the various possible outputs. But since action must be taken in most instances the feed-back loop goes directly from the measured air pollution concentrations to the control mechanism. In a sense, the process can be thought of as being short-circuited at the dashed line in Figure 7.1. Whenever this occurs the decision-making process excludes from formal analysis the most important information necessary for rational control. The suggestions here are meant to eliminate the short-circuit and include the outputs explicitly in the decision-making process. Of course, we do admit that good informal analysis often beats poor formal analysis. But our purpose here is to improve formal analysis.

The Sulfur Decision Problem. A survey of air pollution problems and current air pollution control programs in New York is given in Eisenbud [1970]. In 1970 a major decision still to be made in New York City's air pollution



A General Model for Evaluating
Air Pollution Control Programs

Figure 7.1

control program concerned sulfur dioxide. Table 7.1 presents a breakdown of the estimated 1972 emissions of sulfur dioxide from sources within the City (as viewed from 1969). These estimates accounted for all provisions of existing laws enacted through mid-1971.

Table 7.1: Estimated 1972 Emissions of Sulfur Dioxide in New York City [NYC Department of Air Resources, 1969]

Source of Emissions	Emissions of Sulfur Dioxide (tons)	(per cent of total)
Incineration of refuse	2,500	(0.6%)
Motor vehicles	20,400	(5.1%)
Industrial processes	9,900	(2.5%)
Space heating	195,300	(49.2%)
Power generation	169,500	(42.6%)
TOTAL	397,600	(100.0%)

Since over ninety per cent of these emissions arise from the burning of fuels for space heating and power generation and since the only current, practical way to reduce emissions from these sources is to lower the sulfur content of the fuels burned, one important decision faced by the City was: "should the legal limit on the sulfur content of fuels burned in the City (then one per cent) be lowered?"

7.1.2 Identification of Major Objectives

In almost every decision problem faced by the Mayor of New York City, his most fundamental objective is to improve the well-being of his constituents. However, one must spell out in more detail what is meant by this objective as it pertains to air pollution. Precisely what would the Mayor like to accomplish by his actions concerning air pollution? After some serious thought, an evolutionary process led Ellis to divide the overall objective into five major objectives:

1. Decrease the adverse effects of air pollution on the health of New York City residents.
2. Decrease the adverse economic effects of air pollution on the residents.
3. Decrease the adverse effects of air pollution on psychological well-being of the residents.
4. Decrease the net costs of air pollution to the city government.
5. Achieve as desirable a political "solution" as possible.

These objectives require little justification. However, it should be noted that the second objective is meant to include costs of the air pollution control program in addition to costs of pollution itself. The net costs alluded to in the fourth objective include all the direct costs, such as the costs of an air pollution control program, as well as indirect costs such as those due to migration of businesses and industry from the city, less tourism, and tax revenue losses resulting from employee absences, due to sickness caused by air pollution.

Do these five objectives include all the issues of importance to the Mayor? For instance, nothing has been said about the overall consequences of the various alternatives on New York State, on the Federal government, on businesses, or on non-residents of New York City. Should these factors be included in a complete analysis of proposed air pollution control programs? Of course, the Mayor is concerned about these issues. However, note that some aspects of these consequences, such as economic effects due to tourism and businesses moving to the city, are included in the objective "decrease the net costs of air pollution to the city government." Benefits to non-residents from any air pollution program, for example, are probably highly correlated with the benefits to residents, and therefore in a first approximation could be ignored. All in all Ellis felt that explicit consideration of any of these additional objectives would not alter the optimal strategy, and therefore they were initially excluded from his list of objectives. However, after a preliminary analysis, he did reexamine these exclusions --albeit in an informal manner.

7.1.3 Assigning Attributes to Each Objective

The next task is to identify for each of the objectives suitable attributes that unambiguously indicate the degree to which the associated objective is achieved.

Health Effects on Residents. Several possible attributes immediately come to mind for the objective "decrease the adverse effects of air pollution on the health of residents." These include the annual number of deaths attributable to air pollution, the annual number of man-days of morbidity attributable to air pollution, and some subjectively assessed health index that includes consideration of both morbidity and mortality.

Important objections can be raised against each of these. The annual number of deaths attributable to air pollution is not comprehensive in that it does not account at all for what is believed to be the more prevalent effect of air pollution on health--namely, its effect on morbidity. Similarly, the annual number of man-days of morbidity does not account at all for the extremely serious effect of air pollution on health in terms of mortality.

Thus, it seems clear that no single measure of effectiveness, aside from possibly a subjective health index, can be identified for this objective. However, because such an index lacks a physical interpretation, it is not particularly desirable in terms of the measurability criterion discussed in Chapter 2. Hence, the alternative of specifying the major objective in more detail was chosen. Health considerations were divided into two detailed objectives, "decrease mortality" and "decrease morbidity."

For the first of these, two of the possible measures of effectiveness are the "annual number of deaths attributable to air pollution" and the inversely oriented scale, "per capita increase in the number of days of remaining lifetime due to improved air quality." The first equally weighs the death of an old person and the death of a child, whereas the second measure weighs the death of a young person more heavily. The latter measure was chosen since it was felt that in this case, it more adequately describes the impact of a program alternative with respect to "decrease mortality."

For the objective "decrease morbidity," the "per capita decrease in the number of days of bed disability per year due to improved air quality" was chosen as the attribute. Obviously, this does not include such effects as sore eyes which would not force one to a bed. Part of the consequence of sore eyes is psychological, which can be accounted for by the third major objective. However, the physical aspects of sore eyes intuitively seem important enough to be formally included in the analysis. To do this we would suggest calibrating a number of days of bed disability per year which one would feel is equivalent to having sore eyes of different levels of severity during the year. Then for each program alternative, the effects due to sore eyes would be included in the analysis by adding an "equivalent number of bed-days disability" to our measure of the degree to which "decrease morbidity" is met.

Economic Effects on Residents. No single attribute could be identified for the objective "decrease the adverse economic effects of air pollution on residents of New York City," because the Mayor would want to consider the economic impact on residents at various income-levels. As a compromise Ellis chose to consider a dichotomy: the economic effects on low-income and on other residents. Per capita annual net cost to residents was used as the measure of effectiveness for each group.

Psychological Effects on Residents. There seems to be no direct measure of effectiveness for the objective "decrease the adverse effects of air pollution on the psychological well-being of the residents." One could, however, define some subjective index and perhaps interview residents about their feelings for various levels of air quality. But Ellis chose a simpler approach, which used the daily concentration of sulfur dioxide as a proxy attribute for "psychological well-being."* Since this pollutant can easily be detected both visually and by

*It is important to emphasize that this concentration level is to be viewed as a proxy for psychological well-being only and not for the other objectives.

breathing, it seems reasonable to assume "psychological well-being" is closely related to the concentration levels.

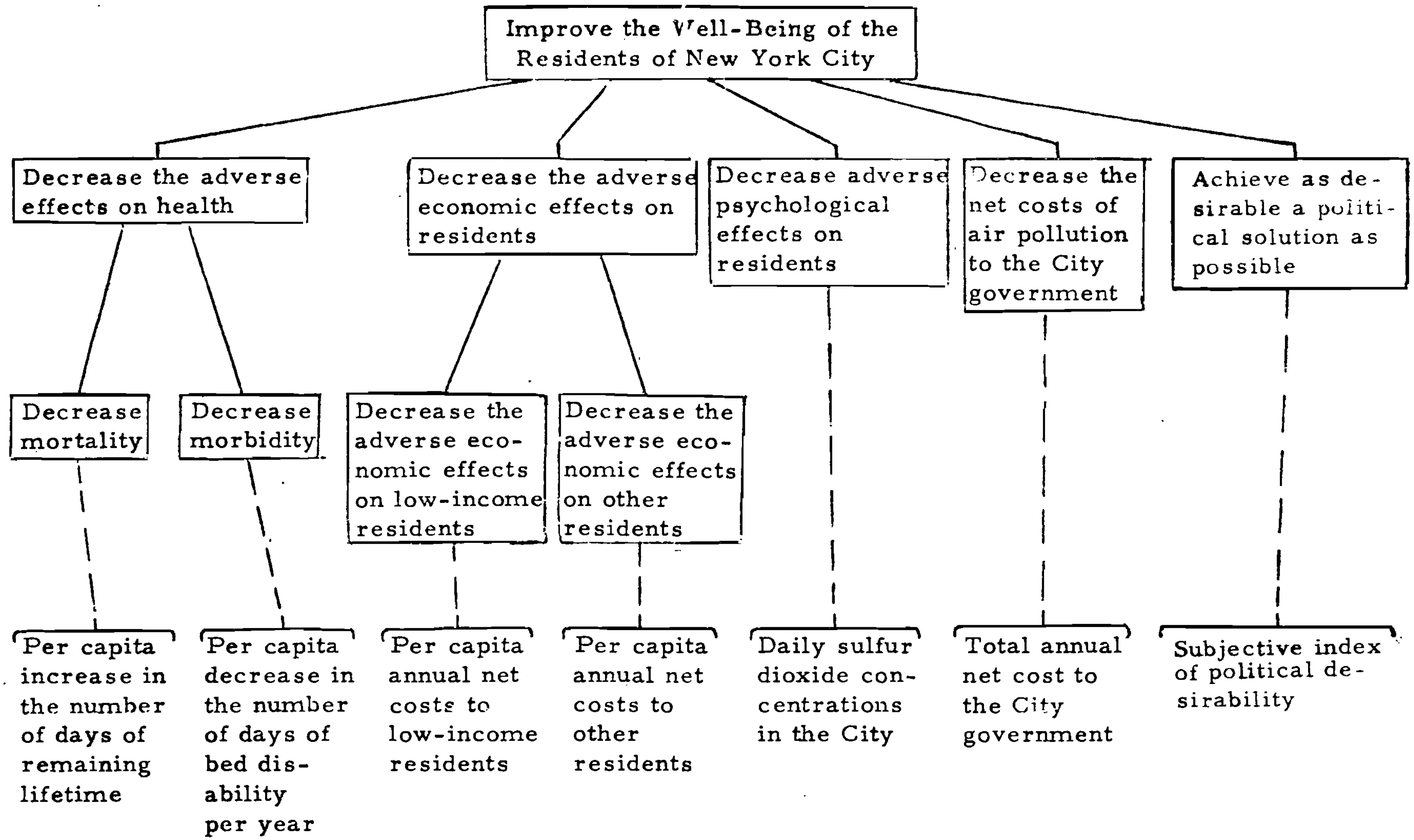
Economic Effects to the City. As a measure of effectiveness for the fourth objective, "decrease the net costs of air pollution to the City government" Ellis used "annual net costs." As mentioned previously, this includes both direct and indirect costs.

Political Implications. The fifth objective, "achieve the best political solution to the air pollution problem," has no nice objective measure of effectiveness and a subjective index was used. Many considerations must be included in measuring the index, such as the possibility of court suits brought by landlords or home-owners who are forced to pay higher fuel prices for heating, the Mayor's relations with the City Council and with Con Edison and with any of the political groups in the city, and the support of the general public for various program alternatives. All of these have a potential effect on the Mayor's political future which also should be taken into account.

7.1.4 The Final Set of Objectives and Attributes

Figure 7.2 exhibits the hierarchy of objectives and their associated measures of effectiveness used by Ellis in his study of air pollution control in New York City.

Of course, there may be important objectives which



The Complete Set of Objectives and Attributes for Choosing an Air Pollution Control Program for New York City

Figure 7.2

Ellis did not think about that are consequently not included in his analysis. However, if one cannot identify such omissions before utilizing the implications of the analysis, the same omissions might have occurred if any less formal procedure for guiding the decision making process were followed. And in this case we would be no worse off using formal analysis than not. Admittedly, with an informal analysis one might think intuitively or subconsciously about objectives that one might not be able to articulate. And also admittedly, a formal analysis may inhibit this mysteriously creative, gestalt way of thinking. But on the other hand, this type of unstructured introspective analysis is so very private that others cannot share in the process and suggest additions or modifications they deem appropriate.

7.1.5 Decision Analysis of the Sulfur-Dioxide Decision Problem

Since the purpose of this section was to develop objectives and attributes for the sulfur-dioxide decision problem, the ensuing analysis will only be briefly mentioned. The interested reader may refer to Ellis [1970] for details of the assessments or to Ellis and Keeney [1972] for an overview.

Since Ellis' work was done as a doctoral thesis designed to illustrate the methodology, only two alternatives were explicitly evaluated. These were the status quo, which entailed maintaining a one percent legal limit on the sulfur content of oil and coal used in New York City, and an alternative which lowered the legal limit to 0.37 percent for oil and 0.7 percent for coal. To analyze the full range of alternatives would require a team of researchers rather than one individual.

The alternatives were evaluated in terms of the seven attributes defined as follows:

$X_1 \equiv$ per capita increase in the number of days of remaining lifetime,

$X_2 \equiv$ per capita decrease in the number of days of bed disability per year,

$X_3 \equiv$ per capita annual net costs to low-income residents,

$X_4 \equiv$ per capita annual net costs to other residents,

$X_5 \equiv$ daily sulfur-dioxide concentrations in parts per million,

$X_6 \equiv$ total annual net cost to the City government,

$X_7 \equiv$ subjective index of political desirability.

Joint probability functions describing the possible impact of the two alternatives were assessed exploiting probabilistic independence, conditional probability assessments, and a small simulation model. Exploiting some utility independence assumptions which were deemed to be appropriate on the basis of discussions with staff members in the Department of Air Resources, a seven-attribute utility function felt to parameterize the Mayor of New York's preferences was structured.

It is interesting to note that Ellis' did _not_ conclude that the Mayor of New York would view each attribute as utility independent of its complement. The main reason for this was the feeling that the Mayor would likely be more risk averse in terms of attributes X_1, X_2, X_4, and X_5 if the political effects were at an undesirable level than he would given desirable political effects. From his interaction with the Department of Air Resources, Ellis did conclude that for the ranges of the possible consequences, the attributes X_3, X_6, and X_7 were each individually utility independent of their respective complement. Also, he felt that given any fixed level of attributes X_7, the attributes X_1, X_2, X_4, and X_5 would each be conditionally utility independent of the remaining attributes. With these assumptions, the assessment of the complete utility function required (1) assessing seven one-attribute utility functions, one over each effectiveness measure, and (2) assessing eighteen scaling constants to insure the seven utility functions were properly scaled. No assessments of the utility function were completed, although details about the functional form of the utility function and the reasonableness of the utility independence are given in Ellis [1970]. Appropriate techniques for performing each of the necessary assessments are found in earlier chapters of this book.

7.1.6 Impact of This Work

The ideas and results expressed in this section may have had some influence on the thinking of individuals responsible for air pollution control programs in New York City. Although no claim can be made concerning causality, the following events have occurred:

The results of this work, concerning the range of possible effects of a program which lowered the legal limits of oil and coal used in the city from the present one percent to 0.37 and 0.7 percent respectively, were made available to the New York City Environmental Protection Administration, which was in the process of preparing a new air pollution control code for the City. This group included, as one of the key provisions in its recommended code to the City Council, a program which was essentially the same program as the one Ellis analyzed.

These same results, as well as the methods of analysis upon which these results are based, were presented by Howard M. Ellis in testimony before the New York City Council in its legislative hearings on the proposed new air pollution control code. The code was approved by the City Council and became law in 1971. Ellis continued to consult with the City after his thesis was completed.

The present authors suspect that, as is the case with many analyses of this type, the detailed quantitative work involved in doing the full-scale study probably helped the investigator to better understand the qualitative implications of the problem, and it was this qualitative understanding which helped him influence the governmental officials. Perhaps this level of sophistication could have come about through other means, but one should not underestimate the important intellectual and emotional impact that arises when one is forced to express vexing tradeoffs in unambiguous quantitative terms. It forces one to think harder than one is ordinarily accustomed to...especially if one then has to defend his assessments in front of other experts.

7.2 PREFERENCE TRADEOFFS AMONG INSTRUCTIONAL PROGRAMS

Roche* considers the problem faced by a decision maker who has to choose among alternate budget allocations to diverse activities which compete for the same scarce resource. He is concerned about the role played by the decision maker caught "in the middle." That is a decision maker who is in the position where he must, on one hand, obtain funds from some approving authority and, on the other hand, approve the budgets for programs directed by professionals in his employ. With a constrained budget he can increase the budget of one program only at the

---

*In this section, we summarize and review the work of Roche [1971]. His doctoral thesis, which was supervised by Raiffa, makes extensive use of the material in Chapter 3 on tradeoffs under certainty.

expense of other programs. He must take from Peter to pay Paul and do it in such a manner to convince his overseers of the reasonableness of it all. Roche was motivated to see if formal preference analysis of the type we are disucssing in this book could help such a man-in-the-middle both to crystallize his own tradeoffs and to communicate this process to the body that controls the dispensation of funds.

Roche chose to study the budget-allocation problem in the context of a small school district. The school superintendent was the decision maker "in the middle," Roche's principal client; the people below the superintendent were the school principal and the coordinators of various educational programs; the people above the superintendent were the school board which acted as the funding agency for the town. School boards in New England have a great deal of fiscal autonomy and can impose financial obligations on the town. But, of course, these school-board members are themselves elected officials so that the ultimate responsibility does reside in the collectivity of town citizens.

Roche was indeed fortunate--but it was far from all luck--to find a chairman of a school board and a superintendent who were initially interested in pursuing a pilot test of Roche's ideas. It is a credit to Roche that the initial curiosity of these cooperating individuals bloomed into full-scale enthusiastic cooperation and, as we shall see, he was skillful enough in his personal relations to involve other individuals in the measurement exercise. Roche, in his thesis, disguises the name of the town, which he fictiously calls "Somerstown," and he disguises as well the names of the characters that participated in the exercise. However, we assure you that many of the dialogues recorded in the thesis are verbatim reports of actual measurement sessions.

7.2.1 Refining the Problem

Somerstown began a program budgeting effort in September 1969, a couple of years before Roche entered the scene. One of the school board members was a business school professor, and it was through his intervention that the superintendent recast the traditional line-item budget into a program format. At the junior high-school level, the basic program format was segregated according to subject matter. The superintendent and the business school professor alluded to above, admitted however, that to their disappointment the program budgeting effort had practically no effect whatsoever on the reallocation of funds to different school-subject programs. Each year, the funds were allocated like the year before except perhaps for a uniform percentage increase. This background may partially explain the receptive audience that Roche received when he approached Somerstown authorities with the idea of examining fundamental tradeoffs among the funding of different subjects. We also point out, in the way of background material, that Somerstown is a small homogeneous community whose educational program was deemed comparatively stable and free of the many frictions that plagued other educational systems at that tumultuous time.

Roche concentrated on the allocation process for four subject programs in the junior-high program:

i. English/Language Arts

ii. Science

iii. Mathematics

iv. Social Studies

The Somerstown Schools have a coordinator for each of these programs and the coordinators prepare an annual budget for their respective domains of responsibility. Each feels a responsibility to do better each year than the year before and each tries to get increases in funding for his or her program--the usual advocacy procedure. When Mr. A asks for an increase, he seldom feels obliged --nor would it be considered good form--to argue that the money he is seeking should come from Mr. B's program. It is the task of the superintendent to juggle these requests and to suggest a compromise among them in a fashion that maintain the loyalty of his staff and at the same time gains the acceptance of the school committee.

The first half of the thesis is concerned with the creation of a suitable production function: the transformation of financial and personnel inputs to educational outputs--no mean task! For a long time before Roche started his probing, the Somerstown school authorities worried about educational indices. Several indicators could be chosen but many are highly correlated and for convenience of the exercise, Roche and his collaborators chose for each of the four subjects the index, "Percentage of students achieving at or above grade level on the standardized achievement test."

In a later chapter of his thesis Roche does discuss the inadequacy of this output measure. He defends his use of it, however, on pragmatic grounds and he does discuss what other researchers might do if they were to choose other output indices. We feel that the chosen index is far from a good surrogate for educational performance and we feel that it is not an elementary task to suggest how Roche's analysis could proceed using a more sensitive set of output indices. But for the time being we are stuck with the index used and let us get on with the story even though it is marred by the exclusive use of this over-simplified output index.

7.2.2 Relating Program Costs to Output

Let us look at the process Roche followed in confronting the science coordinator. The science budget for the existing year was $81,000 and 59% of the students performed at, or better than their specified grade level. Roche first inquired about the effects of dropping the science program altogether. The coordinator did concede that many of the students would continue to perform at or above the hurdle level. He then inquired about the effect of a 10% increase, (i.e., an increase of $8,100). "What would I be allowed to do with the money?", Dave Flaherty queried.

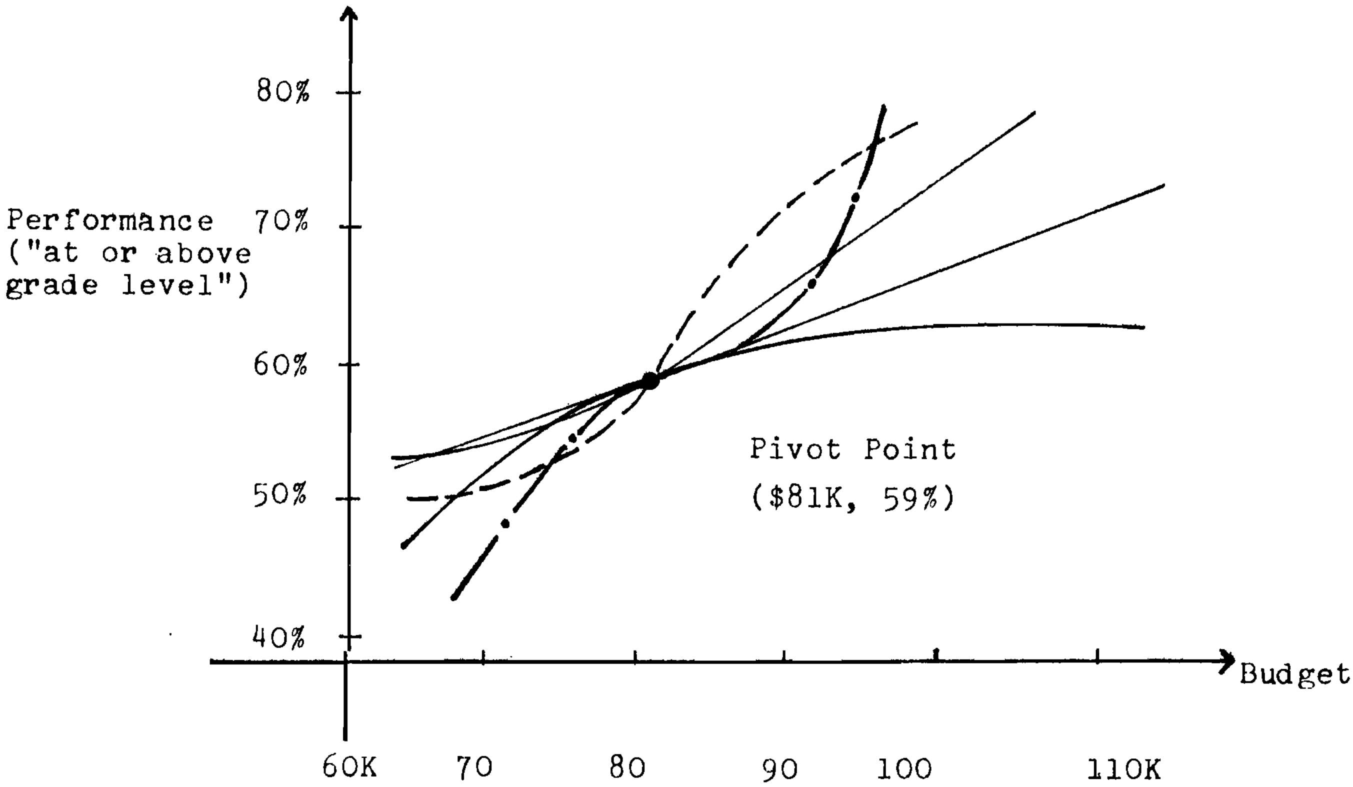
"It's up to you," responded Roche. "The essential point is, Dave, that none of us knows how to use an additional $8,100 in science better than you do. Once you decide what you would do with the $8,100 I will ask you to assess what impact those additional funds would have on the students in the same way we did before. That is, we shall ask in turn: What would you do with the increased funds? Which levels or sections in which grades would be effected? What would you expect the effect would be along the dimension of number of students achieving at or above grade level in science?"

Roche coached Dave Flaherty to think hard about the questions posed. He encouraged the science coordinator to steep himself in the past data, to think about the increased money not in the abstract but in terms of what it would buy in the form of additional teaching help or additional audio-visual facilities, and so on, and to think about the effect on individual students. He posed such questions as: "If you do so-and-so, would this really help Mary Jane over the hurdle?"

The production function ideally should have been probabilistically assessed but all Roche had the courage and time to do was to elicit in each case a median value, i.e. a value for which the assessor thought the true value would be equally likely to fall above or below the estimated value. He formalized the assessment procedure in terms of a written protocol with several pages of work sheets that the coordinator took many hours over a period of days to answer.

The end product that Roche sought from the science coordinator was a curve that plotted estimated performance (% at or above grade level) on the vertical axis against budgetary values on the horizontal axis. This curve, the assessed production function, was meant to go through a pivot point at the status quo level--i.e., a budget of $81,000 produces a performance of 59%.

After Flaherty completed the work sheets prepared by Roche, he was presented with the following task: "Now that you (Flaherty) have completed the assessment questions, we would like to probe your qualitative judgement about the possible shapes of a performance function for the Somerstown Junior High Science Program." Roche then showed Flaherty several shaped curves as shown in Figure 7.3 and they discussed the qualitative meaning of each. After Flaherty seemed to understand the implications of each shape he was asked to select one of the shapes presented or to invent a shape that reflected his true feelings.



Qualitative Shapes of Performance Functions

Figure 7.3

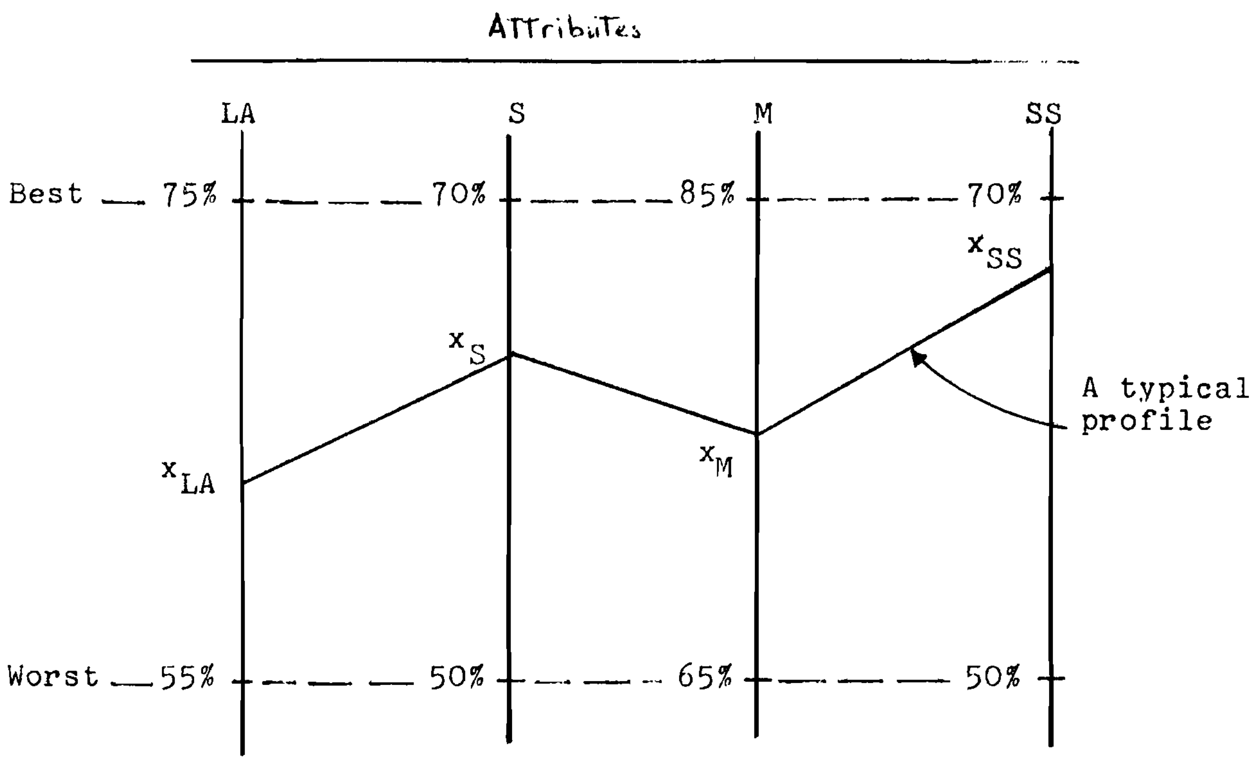
In a gentle manner Roche discussed with Flaherty some of his responses and indicated some inconsistencies amongst the answers he recorded--but he did this corrective procedure with the supportive advice that Flaherty should not be embarrassed at these inconsistencies, since anyone put into his position would be equally inconsistent. The important thing was to have Flaherty reflect and ponder about these inconsistencies and then try to modify some of his earlier assessments so that the revised set of responses would be internally consistent. And what is perhaps more important, the revised answers should be felt to accurately portray the current best assessments Flaherty could make in light of his new level of understanding.

All we can hope to do here in this summary is to give the reader a flavor of the care that Roche took to generate a performance function from each of the four coordinators. The superintendent, Dr. Nelson, had his own views about these performance assessments and felt compelled in some circumstances to modify the assessments of his subordinates. Dr. Nelson remarked, however, that if this assessment process were to be repeated year after year then he would be able calibrate his coordinators on the basis of a track record. The school committee, which monitored the entire exercise felt that it was most appropriate for the superintendent to modify these performance functions in collaboration with his coordinators, since the school committee superintendent had to take full responsibility for the finally recorded performance functions. The committee explicitly stated that their deliberations would be based primarily on the superintendent's own assessments, which, in turn, would be based in part on the inputs he received from his coordinators.

7.2.3 Assessing a Value Function

Now let us turn to the second part of the thesis dealing with preference structures. Roche investigated the preference structures of several concerned indivuduals for different performance profiles. A typical profile is a four-tuple $(x_{LA}, x_S, x_M, x_{SS})$ which refers to performance scores on language-arts, science, mathematics, and social sciences respectively, and where x_{LA}, for example, represents the percentage of students at or above grade level in language/arts.

As is evident in Figure 7.4 each of the performance ranges was restricted to a subinterval of the theoretically feasible range from 0% to 100%. For example, mathematics performance was restricted from the worst case of 65% to the best case of 85%. These restricted ranges were ample to accommodate budgetary changes that could realistically be recommended. It was critical to restrict these ranges so that one could adopt various preferential independence assumptions. We shall expand



Performance Ranges and a Typical Profile

Figure 7.4

on this point shortly.

Due to the considerable support Roche received from Dr. Nelson, the superintendent, and Mrs. Humphrey the chair-woman of the school committee, Roche was able to field test preference assessments with every single administrator and policy maker involved in the decision-making process of the junior high school. These involved the principal and assistant principal of the junior high school, the superintendent and assistant superintendent, and all five members of the Somerstown school committee. Absent from this listing are the citizens and the parents of school children. In addition, the preference procedure was also field tested on a group of 18 doctoral students in educational administration.

It was surprisingly easy to verify the reasonableness of pairwise preferential independence. For example, Roche set x_M and x_{SS} at low levels of 70% and 55% respectively, and then probed conditional preference tradeoffs between X_{LA} and X_S. After he thoroughly engaged his subjects in this problem he asked parenthetically whether any of the tradeoff responses between X_{LA} and X_S would be altered if x_M and x_{SS} were not set at 70% and 55% respectively. Practically all of his subjects felt that these tradeoffs would certainly not be influenced by such modifications of the fixed levels of x_M and x_{SS}. Some subjects, including the superintendent emphasized the point that the tradeoffs would not depend on the fixed levels of X_M and

x_{SS} <u>provided that these levels were within the specified bounds</u>. He felt for example, that if x_M were set at 30% this would be such a shock to the system that his trade-offs between X_{LA} and X_S would be affected.

For all subjects, Roche felt that the necessary pairwise preferential independence assumptions were satisfied to legitimatize adopting a value function of the form

$$v(x_{LA}, x_S, x_M, x_{SS}) = k_{LA}v_{LA}(x_{LA}) + k_S v_S(x_S) + k_M v_M(x_M) + k_{SS}v_{SS}(x_{SS}) \quad (7.1)$$

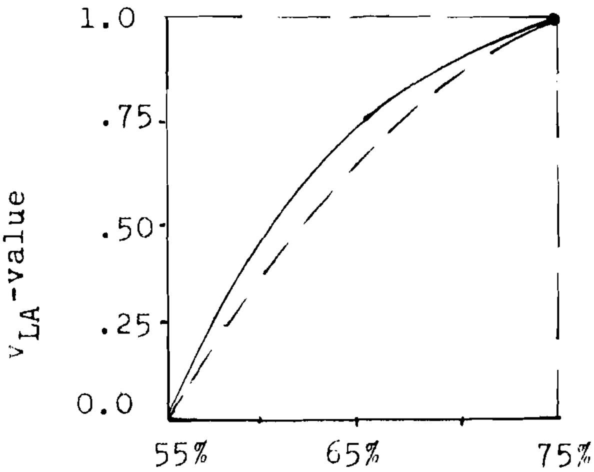
where the component v's were normalized respectively at 0 and 1 for the worst and best alternatives (e.g. $v_{LA}(55) = 0, v_{LA}(75) = 1$, etc.), where the k's were non-negative, and where

$$k_{LA} + k_S + k_M + k_{SS} = 1 \quad . \quad (7.2)$$

Roche followed the assessment procedure described in Section 3.7. He assessed for each subject the component value functions by the mid-value technique: for each component function he first found the .50-value point, next the .25 and the .75 points, then he checked the .50-point against the .25 and .75 points, and finally he discussed the general shape of the v-component functions. Next he sought the k-weights. He asked such questions as: "Suppose we consider a disastrous profile such as (55, 50, 65, 50) where all performance measures are at their worst levels. Now suppose you could push one of these worst scores up from the worst level to the best, which would you choose? Would you prefer to push language/arts up from 55 to 75, or science from 50 to 70, or mathematics from 65 to 85, or social science from 50 to 70?" He thus probed each respondent for rankings of the k's. Next, he followed the technique discussed in Section 3.7 and determined precise numerical values for the k-weights. Figure 7.5 depicts the assessments of Superintendent Nelson and his assistant, Mr. Elliot. Table 7.2 summarizes some salient data collected from the nine principal actors involved in the exercise. Roche not only obtained Nelson's assessments but he had Nelson guess at what some of his associates would record. It's fascinating to read how Nelson rationalized some of the recorded assessments of members of his staff and the school committee members. There are striking differences of opinion!

As regards the 18 students in the doctoral seminar in educational administration, all of whom were subjected to the same assessment procedure, we quote from Roche:

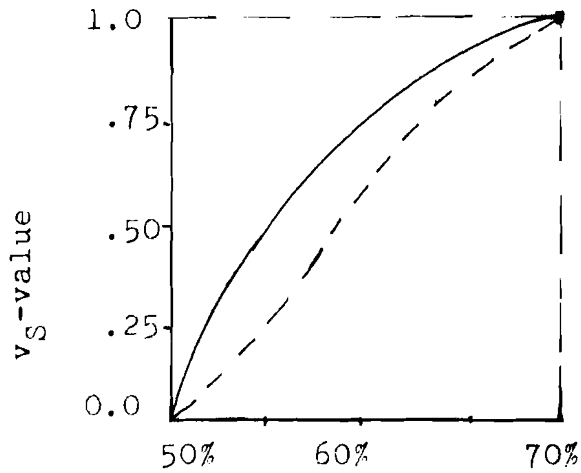
> "There is little to be gained at this point in the study from exhibiting the eighteen structures. However, the following summary information might be of interest.
>
> 1. With respect to the Language Arts program, 11 of the curves were concave, 6 were linear, and 1 was S-shaped about the current



Language Arts Performance

Nelson: k_{LA} = .30

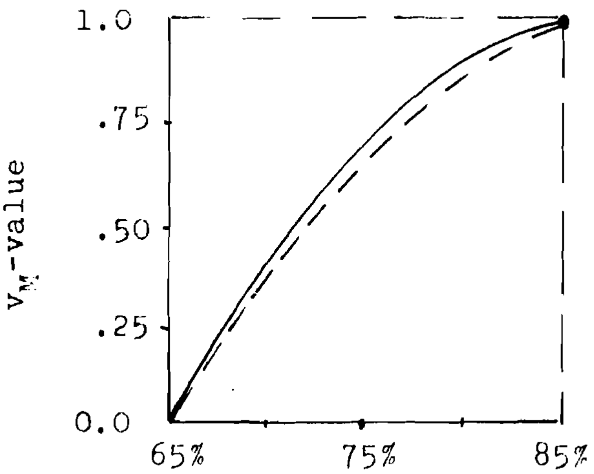
Elliot: k_{LA} = .33



Science Performance

Nelson: k_S = .21

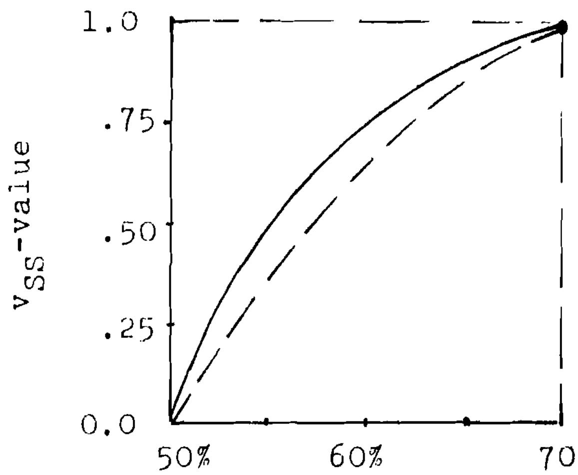
Elliot: k_S = .20



Mathematics Performance

Nelson: k_M = .26

Elliot: k_M = .27



Social Studies Performance

Nelson: k_{SS} = .23

Elliot: k_{SS} = .20

Performance ranges use the evaluator: "percentage of students at or above grade level."

Dr. Nelson, Superintendent: ————

Mr. Elliot, Assistant Super.: ---------

Figure 7.5

Value Functions Assessed by The Superintendent and Assistant Superintendent of The Somerstown School System

TABLE 7.2

Assessed k-Values and .50 Mid-Value Points of Principal Subjects

Name	k-Values LA	S	M	SS	(.50)Mid-Value Points LA	S	M	SS
Administration:								
(1) Mrs. Carter (Principal)	.20	.25	.22	.33	60.5	54.5	70	54
(2) Mrs. MacGregor (Asst. Principal)	.21	.24	.23	.32	61	54.5	68	53.5
(3) Dr. Nelson (Superintendent)	.30	.21	.26	.23	60.5	55	71.5	55
(4) Mr. Elliot (Asst. Superintendent)	.33	.20	.27	.20	62	59	72	57
School Committee:								
(1) Mrs. Humphrey (Chairwoman)	.36	.13	.30	.21	62	63	69	57.5
(2) Mrs. Clark	.22	.26	.23	.29	65	59	67.5	57
(3) Mr. Cowles	.53	.10	.27	.10	65	62	70	63
(4) Mrs. Oscar	.47	.11	.35	.07	65	62	69	60
(5) Mr. MacMillan	.29	.23	.28	.20	59	55	68	54.5
lowest mid-value point assessed:					59	54.5	67.5	53.5
highest mid-value point assessed:					65	63	72	63

Key: Each row contains (1) the scale factor for each program, which indicates the subject's tradeoffs among programs; and (2) the "global" mid-value point for each program, which gives an indication of the subject's tradeoffs within a program. A low mid-value point indicates a strong aversion to poor performance.

performance level.

2. In the Science program, 8 were concave, 2 were linear, 4 were S-shaped about current performance, and 4 were convex. Thus, there was much less concern with poor performance in Science than in Language Arts.
3. Interestingly enough, all 18 of the curves were concave in the Mathematics program. That is, there was unanimous concern with poor performance in Mathematics.
4. In the Social Studies program, 11 of the curves were concave, 3 were linear, 2 were S-shaped about current performance, and 2 were convex.

It is of interest to note that the doctoral students, like the subjects in Somerstown, basically fell into two groups: (1) the "educators" who were essentially concave in all programs (eight of the students fell into this group); and, (2) the "policy makers" who were either concave or linear in Language Arts, concave in Mathematics, and S-shaped or convex in either/or Science and Social Studies (eight of the students fell into this group). Only two of the students did not fall into either of these groups. This was because these two students were S-shaped about current performance in Language Arts. It may be coincidental, but one of the students whose structure very closely approximated the typical school committee member's structure in Somerstown, had just recently run for election to the Boston School Committee.

Of even more interest to the analyst was the fact that no student was linear in all the programs. Therefore, without knowing it, the students demonstrated that the typical "priority list" approach, i.e., the constant linear form, would be inappropriate for analyses of their preferences among programs. When this evidence is added to the data generated in Somerstown, it suggests that the analyst should be extremely careful about using the constant linear form.

With respect to the determination of scale factors during the second part of the assessment procedure, the vast majority of the students behaved as did the Somerstown superintendent and a majority of the Somerstown School Committee. That is, 15 out of the 18 students chose Language Arts as that program they would want to "push-up" first. Science was picked by 2 students, and one chose the Social Studies program. Although none of the students picked Mathematics as the "base" program, 9 of them chose this program as the second program they would like to see "pushed-up." The remaining 9 students all chose Social Studies as the second program."

After Roche obtained the full assessments from his subjects he asked each of four of the School Board members plus the assistant superintendent to suggest budgetary alternatives that would either be most appealing to themselves and would have some chance of being accepted by the group or be of a type that they would welcome seeing evaluated. Five alternatives besides the no-change position were thus generated. Again we quote from Roche:

" The "no-change" alternative for the Junior High School Core Program was as follows: allocate $92,000 to the Language Arts program, $81,000 to the Science program, $76,000 to the Mathematics program, and $75,000 to the Social Studies program. The alternative allocations (expressed as changed to the "no-change" case) are listed below with the names of the individuals who suggested them.

1. The Humphrey alternative. Take $6,000 from Science, and $6,000 from Social Studies. Increase Language Arts by $10,000 and increase Mathematics by $2,000.
2. The Oscar alternative. Take $7,000 from Science, and $2,000 from Social Studies. Increase Language Arts by $6,000, and increase Mathematics by $3,000.
3. The Elliot alternative. Take $7,000 from Science, and $1,000 from Social Studies. Increase Language Arts by $3,000, and

increase Mathematics by $5,000.

4. The Cowles alternative. Take $3,000 from Language Arts, and $6,000 from Mathematics. Increase Science by $4,000, and increase Social Studies by $5,000.

5. The Clark alternative. Take $2,000 from Language Arts, $2,000 from Science, and $1,000 from Mathematics. Apply all $5,000 to Social Studies."

Using the performance functions as generated by the program coordinators and modified by Superintendent Nelson, and using the preferences of each of the four administrators and five Board members, it was possible to evaluate the six suggested proposals. These preferences are shown in Table 7.3 and comparative rankings are shown in Table 7.4. We can see readily that the Humphrey alternative strictly dominates the "no-change" alternative and the Elliot alternative. Furthermore when power realities are also considered, the Humphrey alternative essentially overpowers the Clark and Oscar alternatives as well. This leaves a contest between the Humphrey and the Cowles proposals. Again, however, looking at the personalities and the strengths of preferences one would be tempted to single out the Humphrey proposal as the obvious winner.

Roche raises the question whether the above described procedure could seriously be implemented for group decision making. He writes:

TABLE 7.3

Preference of Each Individual for Each Budget-Alternative Generated by the Educational Value Function

Administrators and Board Members	"No change"	Humphrey	Oscar	Elliot	Cowles	Clark
	Alternative Budget Allocations					
Dr. Nelson	.730	.743	.737	.743	.730	.727
Mr. Elliot	.642	.650	.643	.646	.646	.637
Mr. Carter	.771	.778	.778	.777	.793	.785
Mr. MacGregor	.765	.771	.771	.771	.784	.775
Mrs. Humphrey	.667	.697	.686	.686	.668	.667
Mrs. Clark	.638	.647	.629	.628	.676	.632
Mr. Cowles	.584	.647	.624	.608	.563	.570
Mrs. Oscar	.608	.647	.650	.631	.588	.597
Mr. MacMillan	.813	.816	.807	.809	.816	.809

Key: In each row, the preference or value number as determined by each subject's preference structure is presented for each alternative budget allocation.

TABLE 7.4

Rankings of the Alternatives for Each Individual
Developed from the Preference Numbers in Table 7.4

Administrators and Board Members	Alternative Budget Allocations: "No change"	Humphrey	Oscar	Elliot	Cowles	Clark
Dr. Nelson	3	1	2	1	3	4
Mr. Elliot	4	1	3	2	2	5
Mr. Carter	5	3	3	4	1	2
Mr. MacGregor	4	3	3	3	1	2
Mrs. Humphrey	4	1	2	2	3	4
Mrs. Clark	3	2	5	6	1	4
Mr. Cowles	4	1	2	3	6	5
Mrs. Oscar	4	2	1	3	6	5
Mr. MacMillan	2	1	4	3	1	3

"Under normal conditions, I don't believe it would be reasonable to expect that policy makers would allow their own preference structures to be communicated. Recall that Dr. Nelson said that he would usually want to keep his own preference structure private. The administrators and policy makers in Somerstown are rather unusual. They willingly cooperated in this effort in order to further research on decision making. Additionally, there are no major educational problems in Somerstown. That is, there are no sensitive issues at stake. Therefore, no individual felt threatened by having his or her preference structure recorded. In such a case, decisions would be of the fine tuning variety, rather than the sensitive policy decisions."

Roche developed a computer program that takes the performance functions and the preference structure of a single decision maker--he used Nelson's as an example--and generates the optimum allocation for a given overall budget level. It is essentially a resource-allocation type of dynamic program. Given this program it is easy to generate the program implications of various overall budget levels. Roche however, did not choose to formalize tradeoffs between money and the four indices of scholastic performance. If he had chosen to do this, undoubtedly the set of four scholastic attributes would have been preferential independent of the monetary attribute so that all of Roche's work would also be relevant and appropriate in the extended framework. The computer program also makes it relatively easy to investigate various sensitivity studies: for example, dependence on the k-weights or on changes in performance functions.

We conclude this section with a quote from Roche:

> "Although this research demonstrates that these new techniques could be used to examine budgetary alternatives among programs, the demonstration was within a very narrow context. There may be problems in attempting to use these formal techniques elsewhere. The local educational setting served as a 'laboratory' for the investigation of these techniques. I believe that this setting is representative of numerous nonprofit organizations. However, on the basis of this research we cannot say that these formal techniques should be used <u>everywhere</u>, but, rather, that they could be used <u>somewhere</u>."

7.3 FIRE DEPARTMENT OPERATIONS[+]

In any analysis of fire-department policy a classical question is: "How much is a minute of response time[‡] worth?" Clearly the value for any particular fire depends on the detailed circumstances of that fire. It is not feasible to treat individually each of the several thousand serious fires which the New York Fire Department extinguishes each year. Instead, we will focus on what will be referred to as the "typical New York structural fire."

[+]The work discussed in this section was done for the New York City-Rand Institute by Keeney employed as a consultant. He wishes to thank Deputy Chief Francis J. Ronan of the New York Fire Department and Edward H. Blum of the New York City-Rand Institute for their important contributions to this work. The effort represents part of the joint work by the New York Fire Department and the New York City-Rand Institute to understand and improve the bases for deploying fire department resources. This section utilizes material originally published in Keeney [1973c].

[‡]The response time for a particular piece of equipment is defined as the time elapsed between that apparatus's leaving the fire house and its arrival at the location of the incident.

A general formulation is developed which specifies the value of response time to this "typical fire" as a function of the particular piece of equipment, the response times of the other fire-fighting vehicles, and whether, for example, it is the difference between a 2 and a 3 minute response or the difference between a 7 and an 8 minute response.

An approach to this inherently difficult problem might include (1) engineering research on fire development (e.g., how fast do different materials burn); (2) analyses of data relating losses, damage, etc. to fire department performance; and (3) analysis and distillation of accumulated fire-fighting experience.

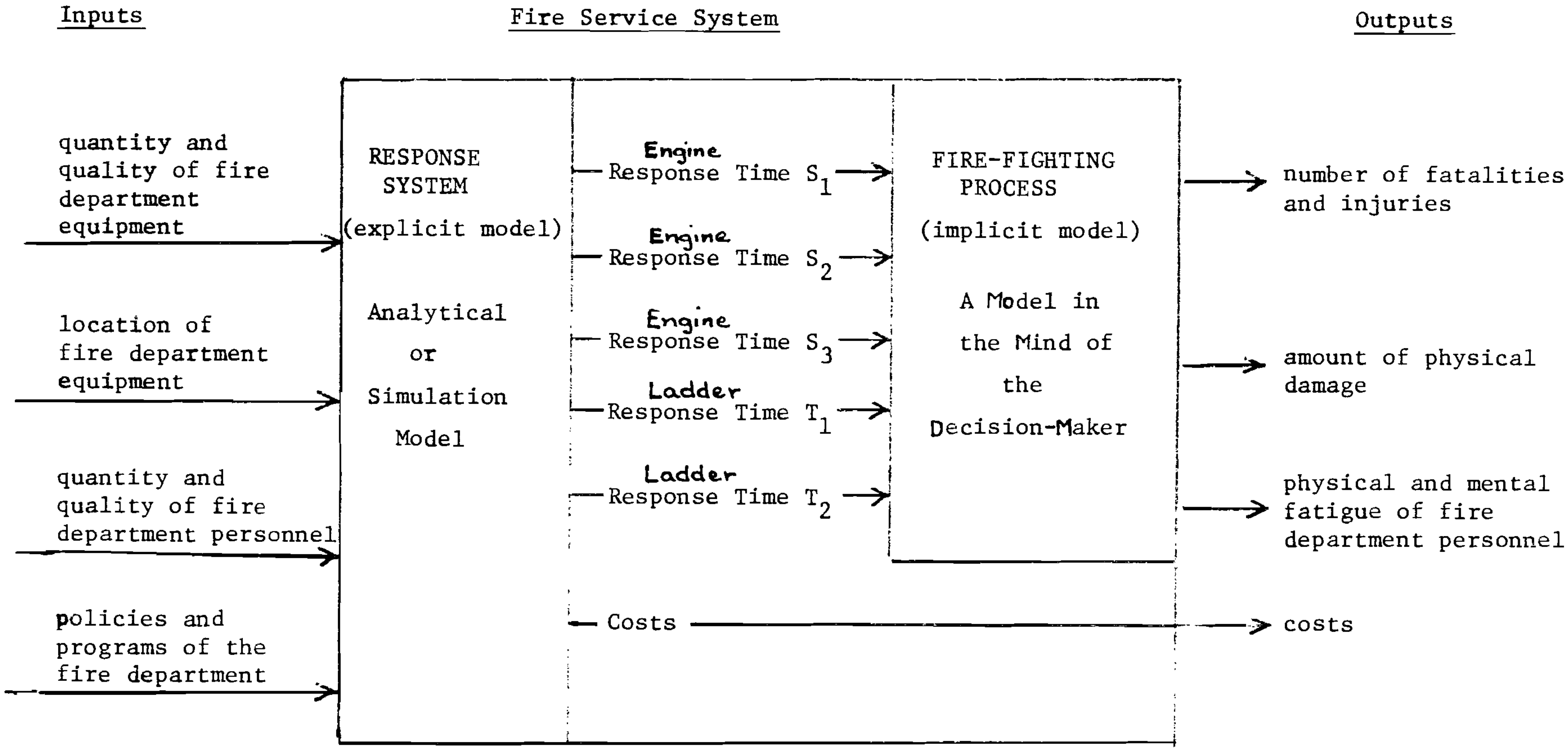
This section, by exploiting the concepts and results of utility theory discussed in earlier chapters, presents an initial attempt to quantify the experience of some New York City Fire Department officials and to investigate means of using this information for evaluating Fire Department policies. This first step involves the preferences of one deputy chief of the New York City Fire Department. A five-attribute utility function is assessed for the response times of the first three engines and first two ladders arriving at a structural fire.

7.3.1 An Overview

Let us step back and try to get an overall picture of where this work fits into Fire Department decision making. It would be desirable to evaluate proposed Fire Department policies and programs in terms of fundamental objectives such as " maximize the quality of fire service provided" and "minimize its cost." Annual cost of the Fire Department measured in dollars may be an appropriate attribute for the second objective, but there is no clear candidate for the first one. Thus, it may be necessary to divide this objective into lower-level objectives such as "minimize loss of life," "minimize injuries," "minimize property damage," "minimize psychological anxiety of the citizens," etc. Reasonable attributes for these first three objectives, are respectively the annual number of deaths, the annual number of injuries caused by fire, and the annual dollar value of lost property, whereas a subjective index would likely be required for the attribute dealing with psychological anxieties. However, these first three attributes are not exactly ideal. It is very difficult to determine what fatalities, injuries, and damage is attributable to the service of the Fire Department and what part is not. For example, an individual who causes a fire by falling asleep while smoking in bed might die before the fire is reported. This and similar fatalities should not reflect on Fire Department services. Such problems with the available data, coupled with the fact that little is known quantitatively about the fire-fighting process, contribute to the non-operational use of these measures. In addition, there are problems about the relative seriousness of different injuries and difficulties of directly placing a value on the life of an individual which further complicates matters.

Fortunately, the response times of the various apparatus responding to fires provide a natural set of proxy attributes for evaluating the level of service for such problems. Figure 7.6 is a simplified model of the fire service system illustrating that response times are inputs to the fire-fighting process, whereas objectives concerning loss of life and property damage relate to outputs.

Firemen are accustomed to thinking in terms of response times in informally evaluating their preferences for various alternative courses of action. In doing this, they use their experience in gauging both the likelihoods of the various possible response times given a particular policy and the effects these response times have on the more fundamental service objectives of the department. Aside from their interpretative appeal, data exist for specifying the probabilities of the response times conditional on a particular course of action. For nearly a decade, the New York Fire Department has kept extensive records on particular aspects of fire occurrence. These data have been analyzed and they provide the necessary input information for developing the simulation model of Fire Department operations, an early version of which is described in Carter and Ignall [1970]. This model is



A Simplified Model of a Fire Department Service System

Figure 7.6

used to generate probability density functions for the response times of any prescribed operational policy.

Our objective here is to relate the various possible response times to the accomplishment of the Fire Department's objectives for fundamental services. We want to distill years of experience of some Fire Department officials by quantifying their subjective preferences for response times to fires in a manner useful for improving the fire-fighter's decision making process. Thus, we are essentially asking the official to consider the implications of a particular set of response times (i.e., the first engine responds in 3 minutes, the second in 5 minutes,..., the first ladder responds in 2 minutes,..., etc.) on the outputs, and then to evaluate his preferences for various sets of response times in light of the respective implications. The result is a subjective model, based on experience, for the fire-fighting process, its consequences, and the relative undesirabilities of these consequences.

Whose preferences should be assessed? By virtue of their experience, it was decided that the operating chiefs were best suited to understand the multitude of implications of various combinations of response times. As a logical first step, the preferences discussed here are those of one deputy chief of the New York Fire Department.

7.3.2 Use of the Response Time Utility Function

The original motivation for assessing a utility function for response times was to develop a model for the escalation of fires. If the originally delegated units cannot control a fire, additional units must be requested, and it is said that the fire escalates. Since such events are very important to the Fire Department's performance, it would be useful to model the escalation phenomenon and include it in the simulation model. Specifically, we would like to know when poor Fire Department service leads to escalation. Since the probability of escalation is clearly related to the quality of deployment and since this quality can be measured by the response-time utility function, it may be desirable to assess the conditional probability of escalation given the quality of the response as summarized by its utility.

The utility function for response times can be useful for guiding decisions concerning operational policy of the department. Examples of such policies concern variation in initial response patterns and dispatching of vehicles, alteration of the areas of responsibility between different pieces of equipment, introduction of "special squads" during high demand hours, and temporary relocation of equipment into areas where resources are almost all working at fires. The simulation mentioned earlier and other models generate, for any given policy, probability distributions for response times. Thus, given an appropriate utility function, one can evaluate policies according to expected utility.

Let $R = \{T_1, T_2, S_1, S_2, S_3\}$ denote the attribute complex dealing with service levels and let C denote the cost attribute. Let

$$u'(c,\underline{r}) = u'(c, t_1, t_2, s_1, s_2, s_3)$$

be the overall utility for cost c and response vector $\underline{r}$. Assuming that R is utility independent of C, a most reasonable assumption, we can define a utility function $u(\underline{r}) = u(t_1, t_2, s_1, s_2, s_3)$ and from the results in Chapter 5,

$$u'(c,\underline{r}) = u'(c, t_1, t_2, s_1, s_2, s_3) = f[c, u(t_1, t_2, s_1, s_2, s_3)] \quad . \tag{7.3}$$

In our discussion, we shall confine our remarks to the assessment of $u(\underline{r})$.

Before proceeding to the assessment of the response-time utility function, let us suggest that the general ideas presented here are relevant to other emergency services such as law enforcement agencies and emergency ambulance systems. In such systems, as discussed and used by Larson [1972], Savas [1969], and Stevenson [1972] to name a few, response times are extensively used to evaluate alternative proposed policies. In all such cases, the question arises, "how much is a minute of response time worth?" The work discussed here is an initial attempt to address such questions.

7.3.3 Assessing the Response-Time Utility Function

During 1970, Deputy Chief Francis J. Ronan of the New York Fire Department and Keeney held a number of discussions to specify Chief Ronan's preferences for response times to fires. These usually lasted between 1 and 2 hours each.

Historically the traditional "standard response" in New York City has involved three engines and two ladders, so it was decided to assess a utility function over five attributes: the response times of the first and second ladders arriving at a fire and the response times of the first three arriving engines. Let us designate these attributes respectively by T_i, $i = 1,2$, and S_j, $j = 1,2,3$, and let t_i and s_j represent specific amounts of T_i and S_j respectively. Thus, we are interested in the response-time utility function $u(t_1,t_2,s_1,s_2,s_3)$.

In discussing the assessment of Chief Ronan's utility function, we will follow the guidelines for the assessment procedure suggested in Chapter 5. Thus, the discussion differentiates into five activities:

(1) familiarization,

(2) verifying the assumptions,

(3) assessing conditional utility functions,

(4) evaluating scaling constants of u,

(5) checking for consistency.

The presentation here will illustrate how the response-time utility function was assessed and what input information was necessary.

Familiarization. Before beginning on this problem, Chief Ronan and Keeney had worked together on a very simple decision-analytic model of a fire-response problem. Also, at an earlier time, Keeney had roughly assessed Chief Ronan's utility function for the response time of the first engine arriving at a fire. The main purpose of this preliminary exercise was to check whether it was reasonable simply to minimize the expected response time of the first engine arriving at a fire. In most analytical studies dealing with emergency services, this linearity assumption (i.e., minimize expected response time) is implicitly used. As a result of this initial exposure, the Chief was introduced to the basic concepts of utility theory. After the first two assessment sessions, which turned out to be learning experiences for both analyst and respondent, the discussions became more productive.

Verifying the Assumptions. To exploit the theory of Chapters 5 and 6, it was necessary to check whether requisite utility independence assumptions were appropriate for this problem.

Specifically, it was verified that it was reasonable* to assume:

(1) engine response times $\{S_1,S_2,S_3\}$ and the ladder response times $\{T_1,T_2\}$ were utility independent of each other,

(2) first ladder response T_1 and second ladder response T_2 were utility independent of each other, and

(3) the j^{th} engine response S_j was utility independent of the other engine responses, for $j = 1,2,$ and 3.

Because of Theorem 5.3 and (1), the assessment of u was broken into two parts: assessments of an engine utility function and a ladder utility function. Analogously, these two utility functions could be broken down into component parts because of (2) and (3).

Let us illustrate the verification procedure with an example. To check if T_1 was utility independent of T_2, Chief Ronan was asked "Given that the response time of the second arriving ladder is fixed at six minutes, what response time t_1 for the first arriving ladder would be indifferent to having a 50-50 chance that the first ladder responds in either one or five minutes?"

---

*The following independence assumptions were deemed to be approximately valid after considerable probing. Some of the dependencies were so slight--contrary to what was first expected--that independence was taken as an innocuous idealization.

Notice that if $t_2 = 6$, then $t_1 < 6$ and this restricts the domain conveniently. A response that $t_1 = 3.4$ minutes was eventually chosen using a "convergence" technique discussed in Section 4.9.

Next we asked the same question only the second ladder response time was fixed at eight rather than six minutes. Again, the indifference response was 3.4 minutes, leading us to believe that the relative preferences for changes in the response time of the first ladder did not depend on the fixed response time of the second ladder. By additional questioning similar to the above, this speculation was confirmed. Thus, it seemed appropriate to assume that T_1 was utility independent of T_2.

Assessing Conditional Utility Functions. Given the assumptions above, utility functions were needed for each of the five response time attributes. Actually these are conditional utility functions since they concern preferences over a single response time given that the other response times are held fixed. However, because of the utility independence conditions, the particular amounts of these other responses are not important, since the utility function should be the same in any case. To illustrate the approach, let us assess $U_1^T(t)$, the conditional utility function for the first arriving ladder.

Through the questioning, we found a 2.2 minute response of the first arriving ladder was indifferent to a 50-50 chance at either a one or three minute response. Similarly, 4.2 minutes was indifferent to a 50-50 chance at three or five minutes, and 6.2 minutes was indifferent to a 50-50 chance at 5 or 7 minutes. In general, a 50-50 chance at either a t or a (t + 2) minute response was indifferent to a (t + 1.2) minute response for certain. As indicated in Chapter 4, such preferences imply the utility function must be of the form

$$u_1^T(t) = d + b(-e^{ct}) \tag{7.4}$$

where d and b, c > 0 are constants. Since utility functions are unique up to positive linear transformations, it was decided to scale u_1^T from minus one to zero. In addition, the response times ranged from zero to twenty minutes, which implied

$$u_1^T(0) = 0 \tag{7.5}$$

and

$$u_1^T(20) = -1 \quad . \tag{7.6}$$

Next, a 4.5 minute response time for the first ladder was found to be indifferent to a 50-50 lottery yielding either one or seven minutes. Hence u_1^T must be such that

$$u_1^T(4.5) = 1/2\ u_1^T(1) + 1/2\ u_1^T(7) \quad . \tag{7.7}$$

Substituting (7.4) into (7.5) through (7.7) yields three equations with three unknowns which can easily be solved to give

$$u_1^T(t) = 0.0998\left(1 - e^{.12t}\right) \quad . \tag{7.8}$$

Similar procedures were used to obtain the other four conditional utility functions.

Evaluating Scaling Constants of u. Given the individual utility functions for the five response times, the next step is to put them together in the appropriate manner to obtain the overall utility function for response times. This requires assessing the scaling constants--that is, the k's--of Theorems 5.3 and 6.3. To illustrate the method, let us use the ladder-response utility function

$$u_L(t_1,t_2) = k_1u_1^T(t_1) + k_2u_2^T(t_2) + [k_1 + k_2 - 1]u_1^T(t_1)\ u_2^T(t_2) \quad . \tag{7.9}$$

Chief Ronan was asked for the response time t_2 of the second ladder such that he would be indifferent between the two ladders arriving in three and eight minutes respectively, denoted by (3,8), and the response $(4,t_2)$. His answer was $t_2 = 5.7$ indicating a willingness to give up one minute of first ladder response in exchange for decreasing second ladder response by 2.3 minutes, given he started from (3,8). This implied

$$u_L(3,8) = u_L(4,5.7) \quad . \tag{7.10}$$

Similarly, we found (2,6) indifferent to (3,4.2) so

$$u_L(2,6) = u_L(3,4.2) \quad . \tag{7.11}$$

Using (7.9) and the individual utility functions to evaluate both sides of (7.10) and (7.11) gives us two equations with two unknowns, the parameters k_1 and k_2, which when solved yields

$$u_L(t_1,t_2) = 0.66u_1^T(t_1) + 0.19u_2^T(t_2) - 0.15u_1^T(t_1)\, u_2^T(t_2) \quad . \tag{7.12}$$

Other parameters of the overall utility function were evaluated in similar ways as covered in Section 6.6. The general idea is to ask questions to obtain equations containing the unknown parameters, and then to solve the set of equations for the parameter values.

Checking for Consistency. Checking the utility function for consistency and reasonableness is obviously very important--both because the assessment is inherently a subjective process and because the synthesis required to obtain the overall utility function can result in the introduction of "errors." It was important to make sure the implications of the utility function agreed with the chief's preferences.

The most important checks concern the conditional utility functions and tradeoffs between the various response times. This involved discussing the implications of the utility function and using the utility function for providing answers to questions like those asked in the assessment process. In all cases where there was a major discrepancy between the implications of the utility function and the chief's preferences, part of the assessment procedure was repeated and his utility function adjusted accordingly. Many parts of the utility function were adjusted in light of consistency checks. The final utility function appears to represent Chief Ronan's responses quite closely.

7.3.4 The Response-Time Utility Function

In this section, we present the final form of the "first-cut" utility function and discuss its implications. From our assessments, we found

$$u(\underline{t},\underline{s}) = 0.24u_L(\underline{t}) + 0.16u_E(\underline{s}) - 0.6u_L(\underline{t})\, u_E(\underline{s}) \quad , \qquad (7.13)$$

where

$$u_L(\underline{t}) \equiv u_L(t_1,t_2) = 0.66u_1^T(t_1) + 0.19u_2^T(t_2) - 0.15u_1^T(t_1)\, u_2^T(t_2) \quad , \qquad (7.14)$$

with

$$u_1^T(t_1) = 0.0998\left[1 - e^{.12t_1}\right] \quad , \qquad (7.15)$$

and

$$u_2^T(t_2) = 0.143\left[1 - 0.5e^{.08t_2} - 0.5e^{.12t_2}\right] \quad , \qquad (7.16)$$

and where

$$u_E(\underline{s}) \equiv u_E(s_1,s_2,s_3) = 0.63u_1^S(s_1) + 0.18u_2^S(s_2) + 0.09u_3^S(s_3)$$

$$- 0.06u_1^S(s_1)\ u_2^S(s_2) - 0.03u_1^S(s_1)\ u_3^S(s_3)$$

$$- 0.01u_2^S(s_2)\ u_3^S(s_3) \quad , \tag{7.17}$$

with

$$u_1^S(s_1) = 0.0845\left[1 - 0.4e^{.1s_1} - 0.6e^{.14s_1}\right] \ , \tag{7.18}$$

$$u_2^S(s_2) = 0.156\left[1 - e^{.1s_2}\right] \tag{7.19}$$

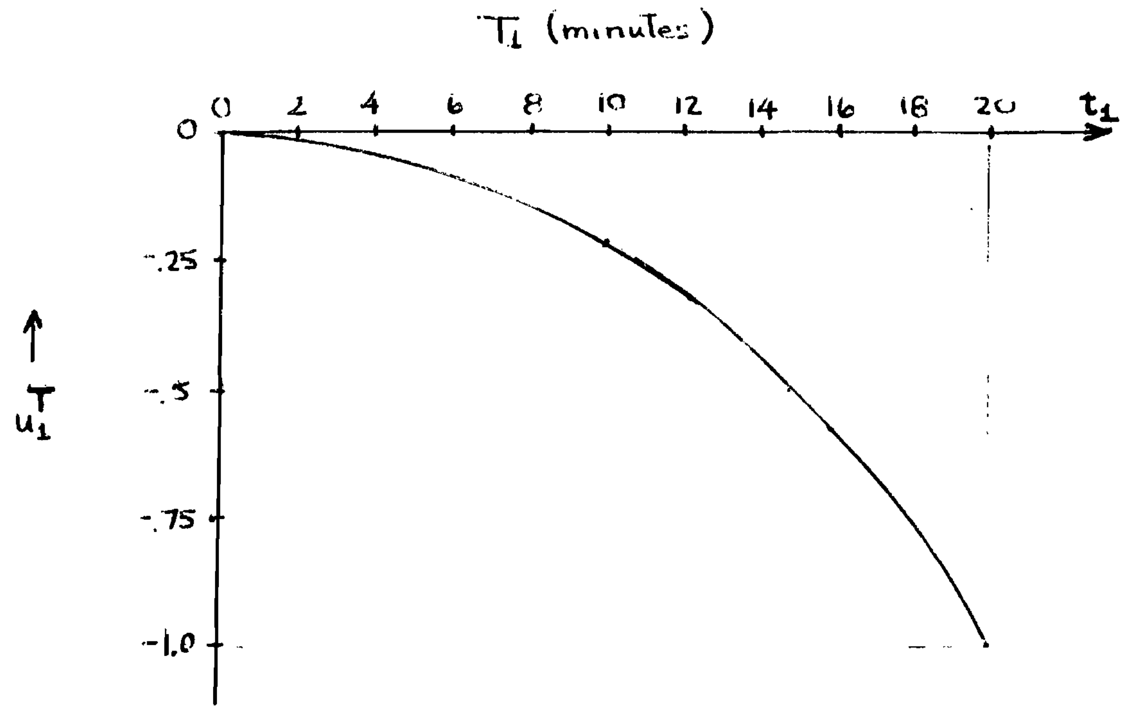
and

$$u_3^S(s_3) = 0.253\left[1 - e^{.08s_3}\right] . \tag{7.20}$$

For illustrative purposes, the utility function in (7.15) is shown in Figure 7.7 and the indifference map implied (7.14) is given in Figure 7.8.

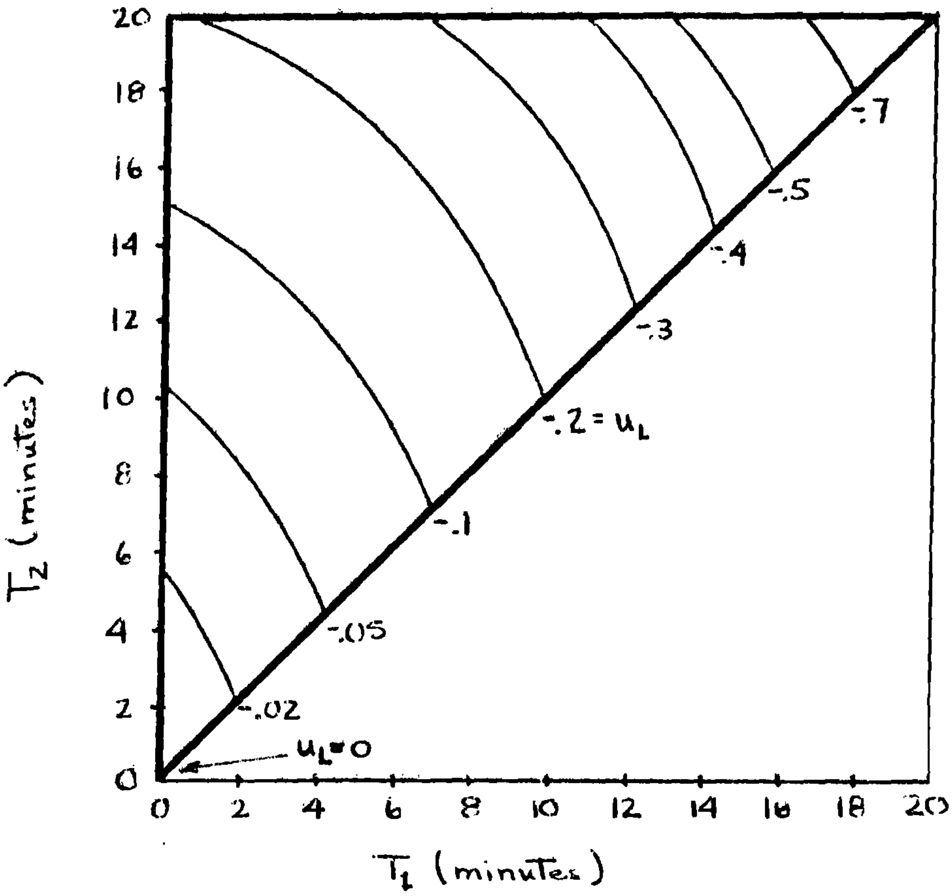
It was decided to evaluate preferences in the unit hypercube from (0,0,0,0,0) to (20,20,20,20,20). Thus, for each of the equations above, the variables may only range from zero to twenty minutes. Furthermore, by definition, we have $t_1 \leq t_2$ and $s_1 \leq s_2 \leq s_3$.

<u>Properties of the Utility Function</u>. The utility function u in (7.13) has several properties which are intuitively appealing and which appear to represent Chief Ronan's preferences. Some of these pertain to u as a whole, some to the utility function for ladders u_L or the utility function for engines u_E, and some to the utility functions of the individual units. Taking the latter first, we have



Utility Function for First Ladder Response Time

Figure 7.7



Indifference Curves for Ladder Response Times
Figure 7.8

(1) u is decreasing in each t_i and s_j. This means the sooner a particular unit arrives, the better, given the response times of other units are fixed.

(2) Each minute of delay of the first arriving engine is more important* than a corresponding minute for the second arriving engine, which in turn is more important than the corresponding delay of the third arriving engine. Similarly, each minute of delay of the first ladder is more important than a corresponding delay of the second ladder. These properties are indicated by the relative values of the coefficients of the u_i^T terms in (7.14) and the u_j^S terms in (7.17).

---

*To clarify the meaning of more important, recall the utility function (7.3) for cost and response times $u'(c,\underline{r}) = u'(c,t_1,t_2,s_1,s_2,s_3)$ and assume the cost attribute C and the set of response attributes $R \equiv \{T_1,T_2,S_1,S_2,S_3\}$ are utility independent of each other. Now select any base level cost c_0 and consider changes $\underline{r}'$ to $\underline{r}''$ and $\hat{\underline{r}}$ to $\hat{\hat{\underline{r}}}$, each of which will be assumed to be for the better. We will say the change $\underline{r}'$ to $\underline{r}''$ is more important than the change $\hat{\underline{r}}$ to $\hat{\hat{\underline{r}}}$ if $c_1 > c_2$, where c_1 and c_2 are defined by $u(c_0,\underline{r}') = u(c_1,\underline{r}'')$ and $u(c_0,\hat{\underline{r}}) = u(c_2,\hat{\hat{\underline{r}}})$. That is, one must be willing to pay more in cost to make the more important change.

(3) The conditional utility function for each attribute is risk averse regardless of the values of the other attributes. This means, for instance, that for T_1, a sure $(t_1'' + t_1')/2$ minute response is preferred to a 50-50 lottery yielding either t_1'' or t_1'. Said another way, the average response time is preferred to the lottery. When this is the case, for each unit, each additional minute of delayed response is more important than the former minute.

Concerning the utility function for ladders, we have:

(4) The relative importance of the response time of the i^{th} ladder increases as the response time of the other ladder increases. Said loosely, this means the slower the first ladder is in arriving, the more important it is that the second ladder arrive soon afterwards. This property is accounted for by the fact that the product term in (7.14) has a negative coefficient.

Similarly, for engines, there is an analogous property:

(5) The relative importance of the response time of the j^{th} engine increases as the response times of the other engines increase. This property is accounted for by the negative coefficients of the product terms in (7.17).

The last two properties concern the entire utility function. They are

(6) A one minute delay in the arrival of the i^{th}

ladder is more important than the corresponding minute delay on the i^{th} engine. Thus, for example, using two minutes responses for both the first engine and the first ladder as a base, we would prefer to have the first ladder respond in two minutes and the first engine in three than to have the first engine respond in two minutes and the first ladder in three. This property is indicated by the fact that the coefficient of u_L in (7.13) is larger than the coefficient of u_E.

(7) The relative importance of the response times of ladders increases as the response times of engines increase. This means the importance of the first arriving engine is less when a ladder has already arrived than it is when no ladders have arrived. The negative coefficient of the product term in (7.13) indicates this property holds.

These properties, each of which is intuitively reasonable, go a long way toward specifying the utility function. That is, the manner in which the shape of the utility function (7.13) can be altered without violating one or more of the above conditions is severely restricted. This fact lends some additional confidence to our assessments.

Although the complexity of assessing a multiattribute utility function increases rapidly as the number of measures of effectiveness increase, the opportunity for "consistency checks" involving properties such as those above also greatly increases. In order to meaningfully represent one's preferences in these complex situations, it is important to exploit such intuitively appealing attitudes toward preference to the fullest extent possible.

7.3.5. Conclusions

The main result of this work is a "first cut" utility function over five response-time attributes, namely, those associated with the first two ladders and the first three engines arriving at a fire. This gives us some means for determining the relative values of a minute of response time for the various pieces of equipment. By looking at the coefficients of the single u_i terms of (7.13), one can get a very rough idea of the relative values of a minute of response time for the different pieces of apparatus. Doing this, if we set the relative value for a minute of response time of the first ladder at 10, the corresponding value for first engine is 7, the second ladder is 3, the second engine is 2, and the third engine is 1.

However, as we have mentioned, the worth of a minute of response time on a specific vehicle depends on the response times of the other pieces of equipment and the time since the alarm was reported. So for instance, using a (2,4;2,4,6) response* as a base, the partial derivatives of u with respect to the five response times are in a ratio of 10:4:5:3:2 implying that if the relative value of a minute of response time of the first ladder is set at 10, the corresponding value of the second ladder is 4, the first engine is 5, the second engine is 3, and the third engine is 2. The point is that the relative values depend on the base response.

The assessment procedure was too time consuming and too complex. Since it was impractical to develop a computer program to help assess one utility function, calculations were done by hand. Thus, there was a lack of immediate feedback to Chief Ronan concerning the implications of his preferences. Often this caused small differences in the chief's responses during different sessions due to the slight variation of his preferences from time to time. But, of course, the involvement over a considerable time span has its merits too. We would like the assessed utility structure to be somewhat stable over time. In the future, an interactive computer program, such as the one discussed in

---

*The first ladder responds in 2 minutes, the second ladder in 4 minutes, the first, second, and third engines in 2, 4, and 6 minutes, respectively.

Appendix 6c, would likely help maintain interest as well as assess the utility function much more quickly with many more consistency checks.

By asking Chief Ronan about his preferences for responses to the "typical structural fire," we essentially asked him to synthesize in his mind all the possible implications of each response aggregated over the possible types of structural fires. This understandably caused some discrepancies in the answers to our questions, because of the tendency to focus on particular types of incidents at different times. Since our major interest in this particular work centers on the first broad cut, rather than details relevant to particular fires, the aggregation requirement may be reasonable.

Our ultimate objective is to obtain a utility function appropriate for the use of the New York Fire Department. This section reports a first step: assessing a utility function of one Deputy Chief of that department. However, the Chief's preferences are his and not necessarily those of the Fire Department, and they should not be interpreted as such. Furthermore, although a serious attempt was made not to lead the Chief to any specific answers, his responses to questions could have in part been shaped by the questioning process, and the results should be interpreted with this possibility in mind.

This assessment exercise was done about five years before the present monograph was sent to the press and if the exercise were to be repeated again, we probably would now proceed somewhat differently. We would attempt to establish some broad, basic, underlying principles, which seemed to govern Chief Ronan's responses and then to deduce more of the structure of his utility function from these basic principles. Essentially we would try to model, to some extent, his motivations based on interviews which would probe more deeply on qualitative matters. This, of course, is easier said than done, and we would like someday to be able to report a good example of this technique. This is the trouble in writing about a subject in its infancy.

7.4 STRUCTURING CORPORATE PREFERENCES FOR MULTIPLE OBJECTIVES*

Every corporation periodically asks itself: "How should we run our business?" More specifically, this raises such questions as: Given the complex social, economic, technological, and political characteristics of our society, which management policies should we adopt now? Are these policies consistent with our personal objectives, with the desires of our shareholders, and with our social value structure? If we choose policy A, will it

---

*We would like to thank the Board of Directors of Woodward-Clyde Consultants for its permission to discuss this work in our book. The assistance of Dr. Keshavan Nair of Woodward-Clyde in writing this section is greatly appreciated. Material in this section is adapted from Keeney [1975].

be possible to account for the contingencies which may arise in the near future and adapt accordingly? How can we best maintain the leadership position in our field and simultaneously, keep the vitality of our organization? All of these are crucial questions which deny the simple dollars and cents answers which are mythically supposed to be appropriate for almost all "business" decisions.

Since early 1972, Woodward-Clyde Consultants, a holding firm for several professional-service consulting firms has used some innovative approaches based on ideas discussed in this book to help them examine questions such as those raised above.* Although this effort is still in progress, it is sufficiently interesting and informative to include here. Two aspects of this effort seem to be unique. First, multiattribute utility functions over attributes measuring fundamental objectives

---

*In November, 1974, Woodward-Clyde made some very broad organizational changes. It is no longer a holder firm but rather one consulting firm with five regional divisions. The work described in this section was done from 1972 through October, 1974, so the organizational structure which prevailed during that period is described. The subsequent organizational changes are briefly summarized at the end of the section.

of the corporation have been assessed for many executives at Woodward-Clyde. Second, this work was done not to evaluate a specific decision, but rather:

- to aid communication among the decision makers,
- to grapple with fundamental issues of the firm,
- to determine and examine differences of opinion in a quantitative fashion, and
- to aid in generating creative alternatives in solving corporate problems.

The affiliate consulting firms of Woodward-Clyde Consultants operate mainly in the geotechnical engineering and environmental areas. Problems they examine include design of earth dams, siting and design of nuclear power plants, geotechnical and environmental studies associated with pipeline systems (e.g., the Trans-Alaska pipeline), and design of structures for earthquake-prone regions. None of the affiliates build any products (e.g. roads, dams, power plants); they are exclusively professional-service consulting firms. Collectively, their fees received in 1973 were approximately 25 million dollars, and historically, this has increased at approximately twenty percent annually. All the shareholders of Woodward-Clyde must be senior professionals on the staff of one of the affiliates.

In 1972, Richard J. Woodward, the Chairman of the Board of Woodward-Clyde Consultants, appointed a long-range planning committee whose assignment included "the development of a long-range plan for Woodward-Clyde Consultants that includes quantified objectives and is responsive to the Statement of Purpose and Standing Policies." After this original committee reported, the 1973 and 1974 Long Range Planning Committees have successively updated the objectives of Woodward-Clyde and examined policy alternatives in terms of these objectives. Douglas C. Moorhouse was the chairman of each of these three committees. Dr. Keshavan Nair, a Vice President of Woodward-Lundgren and Associates, one of the affiliates of Woodward-Clyde was also a member of these committees.

Much of the work discussed here, specifically Sections 7.4.2 through 7.4.5, was done jointly by Dr. Nair and Ralph L. Keeney, working as a consultant to Woodward-Clyde. Section 7.4.1 discusses the original Long-Range Planning Committee's work, which has served as an excellent basis on which to build. The final Section 7.4.6 surveys some of the specific uses being made of Woodward-Clyde's utility function. We just remark here that the purpose in assessing a utility function was <u>not</u> to help management choose amongst action alternatives in a formal manner--e.g., probabilistic analysis was not done in accompaniment of utility analysis--but rather to help management articulate some of its basic assumptions and to facilitate communications amongst the executive group. This, to a large extent, was, and is being, accomplished via the formal assessment procedures described below.

7.4.1 The 1972 Objectives and Measures of Effectiveness

The basic approach taken by the 1972 Long Range Planning Committee to fulfill its mission was (1) to establish the primary objective of the firm, (2) to divide this into subobjectives, and (3) to conduct a deficiency analysis indicating discrepancies between present state and desired state on each objective. By weighting the various objectives, the deficiencies were ranked in order of importance and policies recommended for eliminating these deficiencies.

The overall objective of Woodward-Clyde was provided by a sentence in their Statement of Purpose: "The combined efforts of Woodward-Clyde Consultants and its affiliates are directed toward the creation and maintenance of an environment in which their employees can realize their personal, professional, and financial goals." It was felt that growth was essential in the achievement of this objective.

The hierarchy of objectives developed by the 1972 Long Range Planning Committee is presented in Figure 7.9. This hierarchy has been adaptively revised since that time. The numbers in parentheses in the box with each objective indicates the original division of weight among subobjectives. More will be said about this later. In

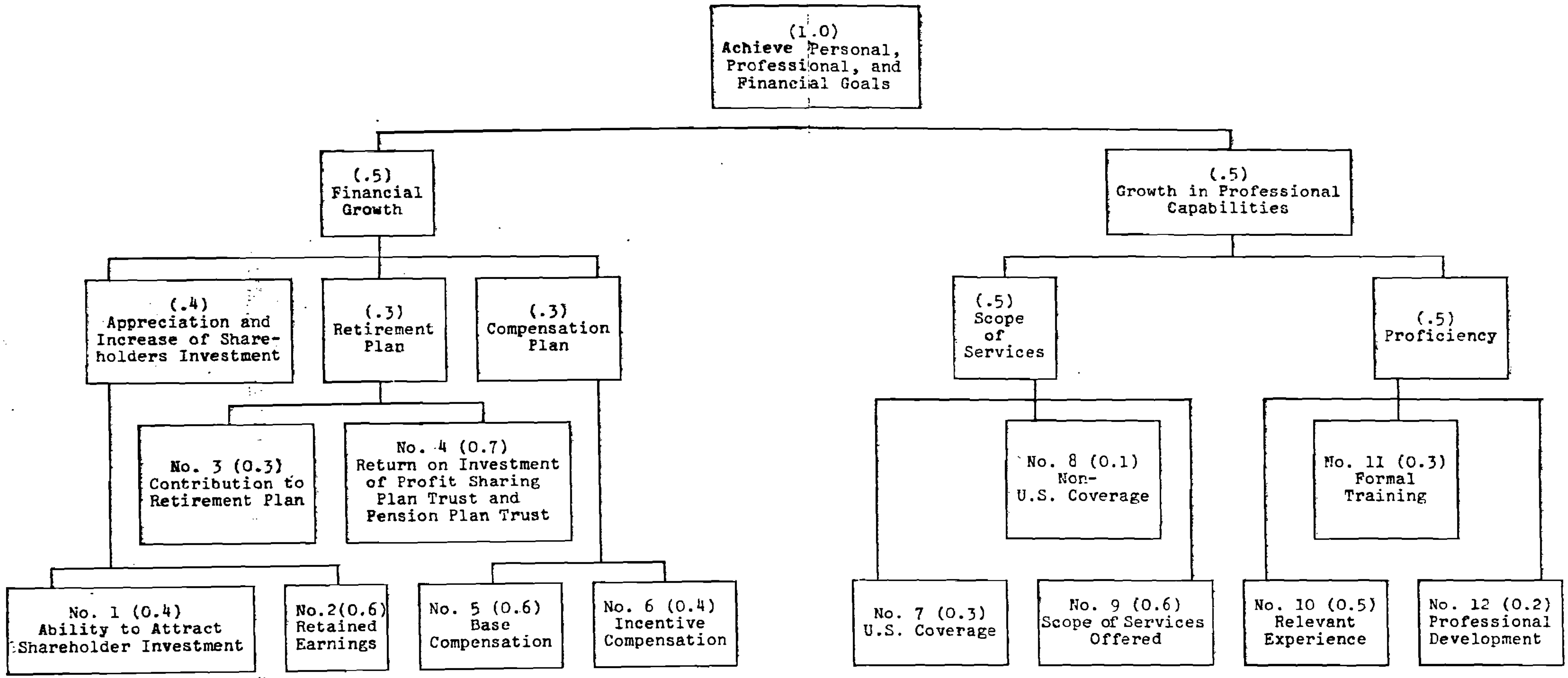


FIGURE 7.9 1972 OBJECTIVES HIERARCHY OF WOODWARD-CLYDE CONSULTANTS

Table 7.5, the weights of each of the attributes associated with the lowest-level objectives and the range of each attribute are identified.

It was implicitly assumed that an additive value function

$$v(x_1, x_2, \ldots, x_{12}) = \sum_{i=1}^{12} k_i v_i(x_i) \quad ,$$

where the x_i's represent levels of the attributes, each v_i is a value function over the i^{th} attribute, v and the v_i's are scaled zero to one, and the weights, that is the k_i's sum to one, was appropriate. For each attribute, component value functions were constructed and present states and desired states, defined as the practical maximum felt to be achievable, were identified. Deficiency on each of these lowest-level objectives was then calculated by multiplying the weight of the objective times the difference in the value of its present and desired states. This indicated "areas" where approvement was needed.

Four shortcomings of the 1972 "quantification of objectives" might be categorized as follows:

(1) the weights were assigned to each objective without explicitly considering the range of the associated attributes,

(2) the component value functions were estimated by a direct value estimation technique independent of each other,

TABLE 7.5

1972 ATTRIBUTES FOR WOODWARD-CLYDE CONSULTANTS

ATTRIBUTE	MEASUREMENT UNIT	RANGE	ATTRIBUTE WEIGHT
Ability to attract shareholders investment	$\frac{\text{Number of shares requested}}{\text{fees}}$ %	0-5	.08
Retained earnings	% of fees	0-8	.12
Contribution to retirement plan	% of fees	0-10	.045
Return on investment for retirement plan	% of investment	0-20	.105
Base compensation	% annual increase	0-20	.09
Incentive compensation	% of fees	0-8	.06
U.S. coverage	$\left(\frac{\text{Geographic centers adequately covered}}{\text{Centers where relevant work can be generated}}\right)$ %	25-100	.075
Non-U.S. coverage	$\left(\frac{\text{Geographic centers adequately covered}}{\text{Centers where relevant work can be generated}}\right)$ %	0-50	.025
Scope of services offered	$\left(\frac{\text{Number of disciplines having threshold capability}}{\text{Number of synergistic disciplines required by society}}\right)$ %	25-100	.15
Relevant experience	$\left(\frac{\text{Existing man-years experience}}{\text{Required man-years experience}}\right)$ %	25-100	.125
Formal training	Number of degrees per professional staff member	1-3	.075
Professional development	% of fees	0-2	.05

(3) the overall objective function, being a value function, was not appropriate for examining policies with uncertain consequences,

(4) the additive value structure did not lend itself to investigating overlap among the objectives.

Even with these weaknesses, the Long Range Planning Committee and the Board of Directors felt this quantification of objectives was a big improvement over informally articulated objectives. This set of objectives and measures has proven to be an excellent basis for modification and improvement, the substance of which we begin to describe in the next subsection.

Before proceeding, let us briefly remark on aspects of the attributes and their measurement units which may not be clear from Table 7.5. For the first attribute, using the number of shares requested divided by fees implicitly assumes the cost of a share is known in order to make the measure readily interpretable. The measure of the scope of services offered is an index meant to indicate breadth in handling the interdisciplinary projects increasingly requested by society. With relevant experience, the idea is to have the staff available to do quality work on those projects which the Woodward-Clyde affiliates would like to do. For formal training, the number of degrees per professional staff member is defined as follows: a doctorate is three, a masters degree two, and a bachelors one. Professional development includes attending management or technical seminars, holding in-house study sessions, etc.

7.4.2 Clarifying the Measures of Effectiveness

One of the first issues Drs. Nair and Keeney jointly considered was whether the measures of effectiveness met the comprehensiveness and measurability criteria discussed in Chapter 2. For each objective, the question "Can a better attribute be found?" was asked. In several cases, the answer was "yes." Let us discuss some examples.

(a) Ability to Attract Shareholders Investment. The measurement unit for this attribute was changed to the dollar value of shares requested divided by the fees. Thus in interpreting trends, and simply in evaluating various levels of the attributes, one does not need to keep the value of the shares implicitly in mind.

(b) Scope of Non-U.S. Coverage. The 1974 Long Range Planning Committee changed this measure to percentage of the United States business in terms of fees received. It was the Committee's viewpoint that the major reason for expanding overseas was to reduce the consequences of a possible recession in the United States and to take advantage of current foreign opportunities. Since Woodward-Clyde will remain primarily a U.S. operation in the foreseeable future, the new measure both is more easily quantifiable than the previous one and also more directly indicates vulnerability to domestic recessions.

(c) Relevant Experience and Professional Development. As demand for Woodward-Clyde services increases, the need to increase their relevant experience grows. The 1972 measure of relevant experience indicated the level at any given time, as opposed to focusing on the increase of relevant experience. Increased relevant experience is funded out of the Professional Development budget and usually consists of opportunities for employees to work on projects under experienced personnel at company expense and to take specialized courses in areas of their practice. Because it is the increase in relevant experience which is currently important at Woodward-Clyde, the measure was changed to percent of fees committed to the relevant experience program.

This change of the relevant experience measure required a redefinition of the components of the professional development measure. In 1972, the latter measure included fees used for obtaining relevant experience. However, with the new relevant experience measure, the professional development measure must explicitly exclude the fees used for acquiring relevant experience.

(d) Formal Training. The measure remained the same for formal training but the desirability of particular levels has greatly changed. The value function in this case is interesting in that it is not monotonic. It is low at a level of 1, since all professionals then only

Table 7.6 An Aid for Evaluating Preferences for the Attribute Formal Training as Measured by the Degrees per Professional Staff Member.

DEGREES PER PROFESSIONAL STAFF MEMBER	5% PhD BS MS PhD	10% PhD BS MS PhD	15% PhD BS MS PhD	20% PhD BS MS PhD	25% PhD BS MS PhD	30% PhD BS MS PhD	35% PhD BS MS PhD	40% PhD BS MS PhD	45% PhD BS MS PhD	50% PhD BS MS [illegible]
1.5	55, 40, 5	60, 30, 10	65, 20, 15	70, 10, 20	75, 0, 25	Not possible	Not possible	Not possible	Not possible	Not possi[illegible]
1.6	45, 50, 5	50, 40, 10	55, 30, 15	60, 20, 20	65, 10, 25	70, 0, 30	"	"	"	"
1.7	35, 60, 5	40, 50, 10	45, 40, 15	50, 30, 20	55, 20, 25	60, 10, 30	65, 0, 35	"	"	"
1.8	25, 70, 5	30, 60, 10	35, 50, 15	40, 40, 20	45, 30, 25	50, 20, 30	55, 10, 35	60, 0, 40	"	"
1.9	15, 80, 5	20, 70, 10	25, 60, 15	30, 50, 20	35, 40, 25	40, 30, 30	45, 20, 35	50, 10, 40	55, 0, 45	"
2.0	5, 90, 5	10, 80, 10	15, 70, 15	20, 60, 20	25, 50, 25	20, 40, 30	35, 30, 35	40, 20, 40	45, 10, 45	50, 0, [illegible]
2.1	Not possible	0, 90, 10	5, 80, 15	10, 70, 20,	15, 60, 25	20, 50, 30	25, 40, 35	20, 30, 40	35, 20, 45	40, 10, [illegible]
2.2	"	Not possible	Not possible	0, 80, 20,	5, 70, 25	10, 40, 30	15, 50, 35	20, 40, 40	25, 30, 45	30, 20, [illegible]
2.3	"	"	"	Not possible	Not possible	70, 30	5, 60, 35	10, 50, 40	15, 40, 45	20, 30, [illegible]
2.4	"	"	"	"	"	Not possible	Not possible	0, 60, 40	5, 50, 45	10, 40, [illegible]
2.5	"	"	"	"	"	"	"	Not possible	Not possible	0, 50, [illegible]

have a bachelors degree, and increases to a peak and then falls rapidly as the level of degrees increases. With a level of 3, the firm would consist entirely of professionals with doctorates. In 1972, the desired state was identified as 2.25, the peak of the value function. On further examination, this level seemed high. If just 25 percent of the professionals of Woodward-Clyde had only a bachelors, a minimum of 50 percent would have to have a doctorate to get the average level to the "desired state" 2.25.

As an aid to thinking about the implications of different levels of "degrees per professional," Table 7.6 was constructed. For evaluating preferences over average degree levels, an individual is meant to select the best distribution of degrees for each average level, and then compare these "best" distributions.

7.4.3 Checking for Independence Conditions

To structure a utility function over the twelve attributes of Table 7.5, modified as indicated in the previous subsection, the process began by examining whether pairs of attributes were preferentially independent of their complements.* In most cases it seemed

---

*Initial assessments were done using Dr. Nair's preferences. Subsequently, Dr. Nair has assessed the preferences of other members of the Long Range Planning Committee.

appropriate to assume preferential independence, but let us indicate three situations where this was not so.

In examining preferential independence assumptions involving the attribute "ability to attract shareholder investment," the Long Range Planning Committee came to the agreement that it was redundant based on present policy. This attribute was meant to indicate the ability and desirability for principals to invest in the corporation. The Committee felt the desirability aspect was adequately captured by retained earnings. On the other hand, the ability to invest was measured by both incentive compensation and base compensation. For these reasons, the "ability to attract shareholder investment" was dropped from the list of attributes.

In another case it at first seemed advantageous to subdivide the objective concerning base compensation into three groups: senior principals, junior principals and associates, and associate candidates. In effect, the current attribute "base compensation" would have been replaced by three attributes, namely base compensation for senior principals, base compensation for junior principals and associates, and base compensation for associate candidates. It was found that one of these attributes taken together with a different attribute, say retained earnings, was not preferentially independent of its complement. The reason was that the rate at which one would substitute retained earnings for base compensation for associate candidates depended on the level of base compensation increases to the principals and associates. If these latter groups received large increases in base compensation, it seemed reasonable to give up more retained earnings to bring increases in base compensation for associate candidates up to some comparable level, than one would give up to make the same increase for associate candidates if in fact the other groups received low increases in base compensation. The concept of equity among the three groups made it inappropriate to assume preferential independence in this case.

There were two other possibilities investigated. Each pair of the three base compensation attributes was found to be conditionally preferentially independent of the third given all other attributes are fixed at an arbitrary level. This would have allowed us to construct an additive component value function over the three attributes. The alternative was to use the original aggregated base compensation attribute. It was felt that members of the Long Range Planning Committee could keep the equity considerations in mind when using the aggregated attribute. Therefore, since it is simpler to use one attribute than the three component attributes, the former was chosen.

Base compensation and incentive compensation do have some overlap in purpose and, because of this, the latter paired with, for instance, retained earnings is not exactly preferentially independent of its complement. However, the overlap is not great since the function of the former is to provide a solid salary for competent work within the "normal" call of duty, whereas the function of the later is to provide motivation and reward for efforts 'beyond' the call of duty. Hence after considerable checking, it was decided that it was a reasonable approximation to assume the preferential independence condition. This "appropriateness" decision was taken in conjunction with the decision to eliminate the attribute "ability to attract shareholder investment" from the list in Table 7.5.

It was decided that the two attributes concerning retirement plan should be aggregated into one called "growth in retirement plan," since in fact both seemed to meet the same fundamental objective. Woodward-Clyde desires that any participant in their retirement plan receive a combined amount from the plan and social security equal to 50 percent of his or her last five years average salary. The new measure for "growth of retirement plan" is the annual increase of assets in the retirement plan. Its range is zero to thirty percent, and it should be clear that this excludes the social security benefits. In effect, this change is simply moving up the objectives hierarchy of Figure 7.9 for a quantitative assessment of retirement plan consequences.

7.4.4 The 1974 Objectives and Measures of Effectiveness

The objectives and attributes updated from the original 1972 list are given in Table 7.7. After considerable examination, Dr. Nair felt that it was appropriate to assume that for the ranges given in the table, each pair of attributes was preferentially independent of its complement. The reasonableness of this assumption has been preliminary accepted by each of the other members on the 1974 Long Range Planning Committee.

7.4.5 Assessing the Utility Function

The preferential independence conditions imply that an additive value function exists over the ten attributes in Table 7.7. From Theorems 6.1 and 6.2, by verifying that just one attribute is utility independent of its complement, either a multiplicative or additive utility function is appropriate to quantify preferences. It was verified that retained earnings was in fact utility independent of its complement, and utility independence was also verified for other attributes to serve as consistency checks. For future reference, it turned out, the final utility function over the attributes in Table 7.7 was multiplicative, and thus expressible in the form

$$1 + ku(\underline{x}) = \prod_{i=1}^{10} [1 + kk_i u_i(x_i)] \quad , \qquad (7.21)$$

where u and the u_i's are scaled zero to one, $0 < k_i < 1$,

TABLE 7.7

1974 ATTRIBUTES FOR WOODWARD-CLYDE CONSULTANTS

ATTRIBUTE	MEASUREMENT UNIT	RANGE
$X_1 \equiv$ Retained earnings	% of fees	0-8
$X_2 \equiv$ Growth in Retirement Plan	% of existing assets	0-30
$X_3 \equiv$ Base Compensation	% annual increase	0-30
$X_4 \equiv$ Incentive Compensation	% of fees	0-8
$X_5 \equiv$ Scope--Geographic (U.S.)	$\left(\dfrac{\text{Geographic centers adequately covered}}{\text{Centers where relevant work can be generated}}\right)$	25-100%
$X_6 \equiv$ Scope--Geographic (Outside U.S.)	% of U. S. business	0-50
$X_7 \equiv$ Scope--Services Offered	$\left(\dfrac{\text{No. of disciplines having threshold capability}}{\text{No. of synergistic disciplines required by society}}\right)$	25-100%
$X_8 \equiv$ Relevant Experience (annual increment)	% of fees	0-1
$X_9 \equiv$ Formal Training	No. of degrees per professional staff member	1.5-2.5
$X_{10} \equiv$ Professional Development (excluding relevant experience)	% of fees	0-1

and k is a non-zero scaling constant greater than minus one which can be evaluated from the k_i's.

The task remaining was to assess the component utility functions, assess their scaling factors, and then evaluate the k-value for the multiplicative form.

Assessing the Component Utility Functions. All the ten utility functions were assessed on a zero to one scale using the techniques discussed in Chapter 4. Let us briefly consider those for retained earnings and formal training, attributes X_1 and X_9 in Table 7.7.

The range of retained earnings is zero to eight percent, so since preferences are monotonically increasing, we set

$$u_1(0) = 0 \quad , \qquad u_1(8) = 1 \quad ,$$

where u_1 is the utility function for retained earnings. Next, by checking certainty equivalents for a number of lotteries, it was verified that Dr. Nair was risk averse in terms of retained earnings. It was found that $2 \sim \langle 0,8 \rangle$. $0.75 \sim \langle 0,2 \rangle$, $4 \sim \langle 2,8 \rangle$, $5.5 \sim \langle 4,8 \rangle$, and for a check, that 4 for certain was indifferent to a 0.75 chance at 8 and a 0.25 chance at zero. The utility function consistent with these assessments is shown in Figure 7.10.

The assessment of the utility function for formal training led to some surprises. What was not a surprise

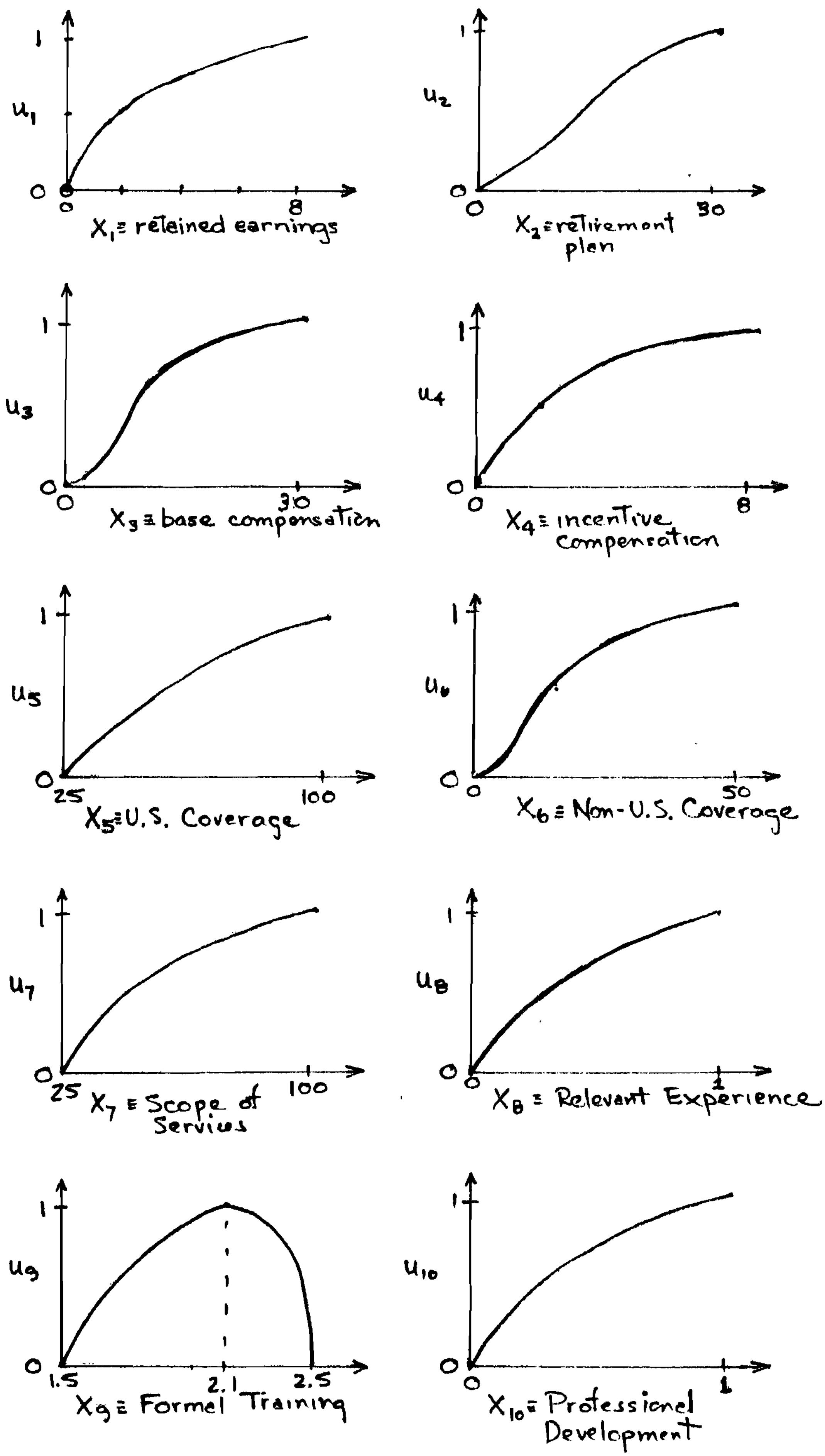


Figure 7.10. Woodward-Clyde's Component Utility Functions

was that preferences for levels of this attribute are not monotonic; they increase up to a maximum point and then decrease. Originally, it was the thought to assess preferences from 1 to 3 degrees per professional staff member. However, once we began this task, it became clear that with levels between 1 and 1.3 and 2.7 and 3, Woodward-Clyde could not exist in a form similar to the present. Hence our viable range was changed from 1.5 to 2.5, which were practical limits for the foreseeable future.

Next, by using the Table 7.6, it became clear that the previously felt optimum level of 2.25 was too high and 2.1 was chosen as an alternative after some consideration. It was also felt that the undesirability of 1.5 or 2.5 degrees per professional was about equally as bad so u_9, the utility function for formal training was scaled by

$$u_9(1.5) = u_9(2.5) = 0 \quad , \quad u_9(2.1) = 1 \quad .$$

Again with the aid of Table 7.6, it was concluded that $1.7 \sim \langle 1.5, 2.1 \rangle$, $1.8 \sim \langle 1.7, 2.1 \rangle$, and $2.3 \sim 1.8$. The resulting utility function is shown in Figure 7.10.

<u>Assessing the Relative Scaling Factors</u>. The ranking of the ten attribute scaling constants of the multiplicative utility function--that is, the k_i's in (7.21)--is given

in Table 7.8. To specify their relative magnitude, Dr. Nair considered the relative desirability of consequences with one attribute at its most preferred level and all other attributes at their worst levels. He decided that the one he would most like to have at its best level was retained earnings. Thus the scaling factor associated with retained earnings is the largest. The attribute he would next prefer to have alone at its most desirable level was formal training so its scaling factor is second largest. Repeating this procedure led to the ranking of the scaling factors indicated in Table 7.8.

To quantitatively establish the relative values of the scaling factors, tradeoffs between pairs of attributes were explicitly assessed. Dr. Nair was asked, for nine pairs of attributes, questions such as:

> "Assume all attributes other than retained earnings and retirement plan are fixed at convenient levels. Now, how high would retained earnings have to be, given the retirement plan is at its lowest level, in order for you to be indifferent between this option and an alternative option with the retirement plan, at its most desirable level of 30 and retained earnings fixed at its lowest level?"

The responses are shown in Table 7.8 in the column labeled "indifference equivalent." Thus if we designate the scaling factor of X_1 as k_1, the scaling factor for

Table 7.8 Evaluating the Scaling Factors in Woodward-Clyde's Utility Function

Attribute	Ranking of Scaling Factor	Range	Indifference Equivalent	Relative Scaling Factor	Scaling Factor
X_1 ≡ retained earnings	1	0-8	—	k_1	.67
X_2 ≡ retirement plan	7	0-30	30 of X_2 ∼ 3 of X_1	$k_2 = .66k_1$	.44
X_3 ≡ base compensation	5	0-30	30 of X_3 ∼ 4 of X_1	$k_3 = .77k_1$	.517
X_4 ≡ incentive compensation	9	0-8	8 of X_4 ∼ 2.5 of X_1	$k_4 = .58k_1$	.391
X_5 ≡ U.S. coverage	6	25-100	100 of X_5 ∼ 3.5 of X_1	$k_5 = .72k_1$	.482
X_6 ≡ non-U.S. coverage	10	0-50	50 of X_6 ∼ 50 of X_5	$k_6 = .5k_5$	.241
X_7 ≡ scope of services	3	25-100	55 of X_1 ∼ 100 of X_5	$k_5 = .75k_7$	.634
X_8 ≡ relevant experience	4	0-1	1 of X_8 ∼ 50 of X_5	$k_8 = .5k_5$	.241
X_9 ≡ formal training	2	1.5-2.5	2.1 of X_9 ∼ 7 of X_1	$k_9 = .97k_1$	.647
X_{10} ≡ professional development	8	0-1	1 of X_{10} ∼ 50 of X_5	$k_{10} = .5k_5$	.241
					4.505

X_2, for instance, must be $.66k_1$ since, using u_1 in Figure 7.10, the utility of a retained earnings of 3 percent is 0.66. This follows since the utility of 3 percent retained earnings, with the growth in retirement plan at its least desirable level, must equal the utility of 30 percent growth in retirement plan, with retained earnings at its minimum level. Because of the preferential independence assumptions, the levels of the attributes other than retained earnings and retirement plan do not matter. The relative values of the scaling constants are also shown in Table 7.8.

Selecting a Utility Function. We felt fairly confident about the relative values of the scaling constants, but to get their absolute magnitudes requires the answer to a difficult question. Dr. Nair was asked:

> "What probability π_1 would you select such that you would be indifferent between option 1 which retained earnings at 8 percent and all other attributes at their least desirable levels and an alternative option 2 consisting of a lottery yielding all attributes at their most desirable level with probability π_1 or otherwise all attributes at their least desirable level?"

Those two options are illustrated in Figure 7.11. Using the "converging method" discussed in Section 4.9, a value of 2/3 for π_1 was selected. This implied that

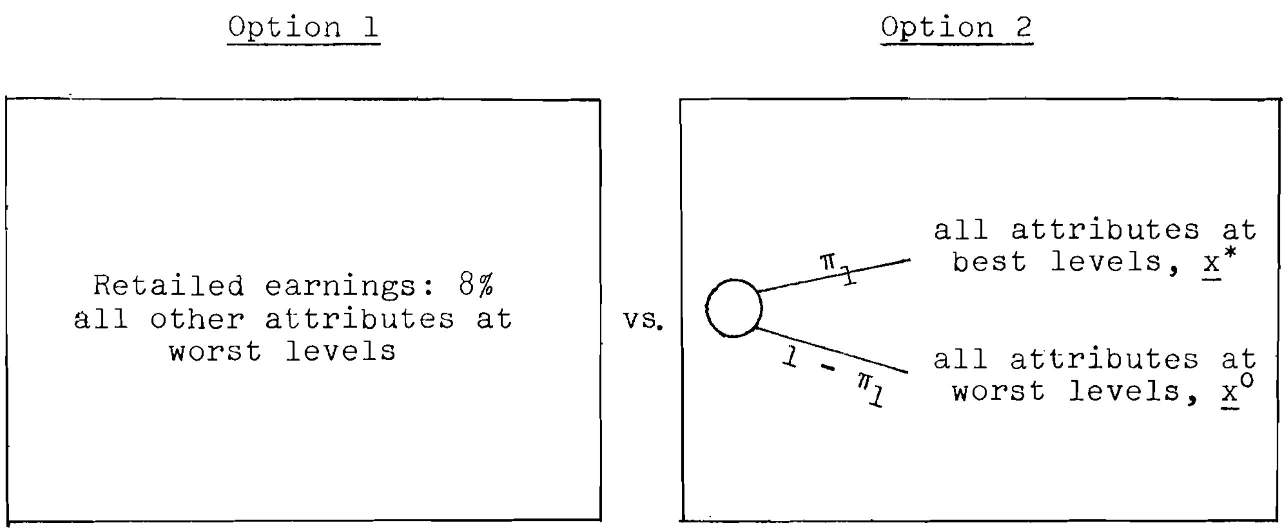


Figure 7.11 Adjust π_1 to get indifference!

the scaling factor k_1 should be 0.67, from which the values of the other scaling factors indicated* in Table 7.8 follow:

Since the sum of the scaling factors is 4.505, we knew the multiplicative utility function (1) was appropriate to express Dr. Nair's preferences. Evaluating (1) for the most desirable consequences one finds

$$1 + k = \prod_{i=1}^{10} (1 + kk_i) \quad , \tag{7.22}$$

which was solved using the routine of Appendix 6B to yield $k = -.998$. Such a low level for k (it must be greater than -1) indicates a high level of complementarity among preferences for the attributes. It is the general feeling of the Long Range Planning Committee that if retained earnings are at a high level, one can "take care of" the other attributes if proper policies are implemented. However, this feeling weakens as the time frame of reference increases. That is if our attributes represent one-year levels, Woodward-Clyde could stand a bad year with most attributes and make it up in the next year. On the other hand, if the attributes of Table 7.8 designate five-year averages, the desirability of waiting five years to "redistribute" high retained earnings to attributes at their lowest levels is understandably much

*The sensitivity of the analysis to π, is discussed shortly.

less. This situation, which became apparent during the assessment process, is clearly important to recognize in discussions of options affecting the future vitality of Woodward-Clyde. The original preference assessments were made using a one-year period. The results reported here are made using annual averages over a three-year period.*

Sensitivity Analysis. Because of the importance of the probability π_1 assessed to specify k_1, a small sensitivity analysis was made of this parameter using the same relative values of the scaling constants in Table 7.8. Recall that $\underline{x}^*$ defines the consequence with all attributes at their best levels and $\underline{x}^0$ the consequence with all attributes at their worst levels. To assist in examining the implications of the various π_1 values, let us make two definitions:

$\pi' \equiv$ the probability such that a lottery with a π' chance at $\underline{x}^*$ and a $(1 - \pi')$ chance at $\underline{x}^0$ is indifferent a consequence with retained earnings and formal training at their best levels and all other attributes at their worst levels,

---

*For reference, the indifference probability π_1 for the options in Figure 7.11 was 0.75 when a one-year period was considered, whereas it was 0.67 for the three-year period.

$\hat{\pi} \equiv$ the probability such that $\langle \underline{x}^*, \hat{\pi}, \underline{x}^o \rangle$ is indifferent to the sure consequence with each attribute at its level of 0.5 utility.

The results, which were calculated using the computer program discussed in Appendix 6C, are shown in Table 7.9, where π_1 is first specified. Then, using the relative scaling factors from Table 7.8, the individual k_i's are fixed. Using these, k, π', and $\hat{\pi}$ were calculated. Further reflection and examination of Table 7.9 led Dr. Nair to stay with his original estimate of $\pi_1 = 0.67$ for the three-year period. Thus, the final scaling constants are those shown in the last column of Table 7.8.

Table 7.9 A Sensitivity Analysis of the Scaling Factor k

π_1	Σk_i	k	π'	$\hat{\pi}$
.87	5.86	-.999	.98	.973
.74	4.96	-.999	.925	.947
.67	4.5	-.998	.884	.928
.60	4.06	-.996	.836	.903
.47	3.15	-.979	.714	.835
.34	2.25	-.900	.561	.733

7.4.6 Uses of Woodward-Clyde's Utility Function

Since the original assessments, Dr. Nair has essentially repeated the assessment procedure just described with each of the members of the 1974 Long Range Planning Committee. These assessments included verification of assumptions, assessing single-attribute utility functions, and specifying scaling constants. This resulted in some minor changes to Dr. Nair's utility function (already integrated into the previous subsections) to achieve what may be referred to as a consensus corporate utility function. This obviously does not mean the Board of Woodward-Clyde will blindly make decisions with this utility function. It is being used to facilitate communication among officers of Woodward-Clyde and to help professional intuition.

The assessment process forced individuals to be a bit more precise in deciding why they felt certain levels of specific attributes were important. As previously mentioned, it also served to indicate how tradeoffs among attributes depended on the time frame of reference. The general feeling of those involved in the utility function assessment may be summed up by the comment of one individual, "I've had to make tradeoff decisions like this all my life, but until now the process has always been somewhat fuzzy and left me with the feeling that I didn't completely comprehend all the implications of my subjective judgements. The use of utility theory and explicit tradeoffs helps considerably." With a better understanding of one's own tradeoffs and preferences, it is a small wonder that it becomes easier to communicate these and discuss the issues with one's colleagues.

The process of assessing a utility function has also led to minor, but important, modifications in the overall evaluation process for long-range plans. Some objectives have been deleted or aggregated, and in other cases, several attributes have been altered to better indicate the concerns of Woodward-Clyde. Changing the attribute measure for relevant experience to reflect the yearly increase in experience is one such example.

Since several of the attributes concern distribution of income available (i.e., percent of fees), it is a simple task to use the utility function to help select the best distribution among salaries, retained earnings, incentive compensation, professional development, relevant experience, and contribution to retirement plan. With any fixed percentage of fees available, the technically feasible surface of fee distribution, as well as the distribution with maximum utility, is easily specified.

As before, the component utility functions can still be used to conduct a deficiency analysis by indicating the difference between the present state and a desired state, representing what is technically feasible in a specified time span. A bit more broadly, by calculating the gradient of the utility function in each attribute for the present state position and combining this with subjectively assessed changes in the state of each attribute for an equivalent amount of effort (time and money), one gets an indicator of policies which may be particularly fruitful to pursue.

The utility function discussed here will no doubt go through additional metamorphosis in the future years, as needs and preferences of individuals at Woodward-Clyde adjust to better reflect their position in society, the external environment, and so on. For example, the Pension Reform Act of 1974, because of certain provisions with regard to the ability of Pension and Profit Sharing Plan Trusts to invest in company stock, is likely to alter the present relative value of the attribute "growth in retirement plan" among the attributes. Woodward-Clyde Consultants is presently examining the effect of this and other external changes on the utility functions for the various individual attributes and the tradeoffs between the attributes. This will be a continuing activity.

The current function does overcome the original shortcomings on the 1972 quantification of objectives outlined in Section 7.4.1. It is being used to examine present decisions which effect the future existence of the company. In addition, the Woodward-Clyde objectives hierarchy partially provides an underlying and unifying basis for evaluating long-range plans and operational activities of the affiliated firms. It is not an overstatement to say that several individuals at Woodward-Clyde find the multiattribute utility concept interesting and helpful. Perhaps more importantly, they are enthusiastic about potential future uses. In this regard, partially as a result of the work discussed here, a special group within Woodward-Clyde Consultants has been set up and funded to begin to transfer the concepts and techniques of decision analysis into their professional practice.

As an interesting anecdote, in 1974, Woodward-Clyde Consultants reorganized its operations from that of a holding company subsidiary relationship to an operating company with five regional divisions, each division having geotechnical and environmental capabilities. The more significant reasons given for this reorganization were to better serve its clients in terms of providing integrated geotechnical and environmental capability, establish a one company image for improved marketing, and increase efficiencies by eliminating various subsidiary management structures. In evaluating the desirability of the organizational changes, many members of the Board of Directors made a subjective determination as to whether the changes would increase the companies ability to improve their level of performance over the various attributes. The explicit statement of attributes made it possible to make this evaluation.

7.5 EVALUATING COMPUTER SYSTEMS

How should management select a computer system? How should the management of a computer facility evaluate the quality of its service? When and how should a time-sharing system be altered to provide better service to its users and to attract additional users? These are representative questions facing various participants, including both managers and users in todays computer industry. It seems that responsible answers to such questions require the consideration of a number of factors: availability of the system, its reliability, response times to different requests, costs, as well as many less tangible aspects. These problems are inherently multidimensional.

In this section we will report on some work of Grochow [1972,1973], which deals with such questions using the concepts and methodology discussed earlier in this book. Grochow assessed a three-attribute utility function for users of time-sharing systems. To illustrate the usefulness of such information for decision making by the management of these systems, we first describe what Grochow did and then discuss its relevance to the questions posed at the beginning of this section.

7.5.1 Preferences of Systems Programmers

To begin, Grochow interviewed a number of users of general time-sharing systems to determine their usage patterns and objectives of importance. His subjects were computer system programmers concerned mainly with the input and editing of programs and the compilation and testing of these. Their ratio of editing sessions to compiling and testing sessions was approximately five to one. Four attributes of the system important to this class of users were

(1) Response time to trivial requests, i.e., editing,
(2) Response time to compute-bound requests, i.e., compiling,
(3) Availability,
(4) Reliability.

Grochow assessed utility functions over the first three of these attributes conditional on reliability being at a high level.

Before beginning the assessment process, Grochow discussed the basic ideas of utility theory with each user and presented a scenario indicating the importance of the three attributes and establishing that reliability was at a high level. For measures of effectiveness he used, for the first two attributes the average number of seconds to satisfy requests, and for the third, the percentage of successful log-ins.

By assessing various conditional utility functions over one attribute at a time given that the other two attributes were held fixed, he established the appropriateness of different utility independence conditions and thus, restricted the form of the utility function. Let us define attributes

$X \equiv$ average response time to trivial requests in seconds,

$Y \equiv$ average response time to compute-bound requests in seconds, and

$Z \equiv$ percentage of successful log-ins.

In terms of this notation, the conditions that Grochow verified as appropriate for the class of users under consideration were

(i) X is conditionally utility independent of Y given Z,

(ii) X is conditionally utility independent of Z given Y,

(iii) Y is conditionally utility independent of Z given X.

It follows directly from Theorem 6.17 in subsection 6.11.4 conditions (i) and (ii) imply that

(iv) X is utility independent of $\{Y,Z\}$.

Using Theorem 5.6, from condition (iv), we know

$$u(x,y,z) = u_X(x,y^o,z^o)\; u(x^*,y,z) + \left[1 - u_X(x,y^o,z^o)\right] u(x^o,y,z) \tag{7.23}$$

where u and u_X are scaled from zero to one with superscripts ^{o} and $*$ indicating respectively the least and most desirable level of an attribute. Then using condition (iii) and the analogous result to Theorem 5.6 for conditional utility functions, we can further break down (7.23) to yield

$$u(x,y,z) = u_X(x,y^O,z^O)\ \big[u_Y^*(x^*,y,z^O)\ u(x^*,y^*,z)$$
$$+ \{1 - u_Y^*(x^*,y,z^O)\}\ u(x^*,y^O,z)\big]$$
$$+ \big[1 - u_X(x,y^O,z^O)\big]\ \big[u_Y^O(x^O,y,z^O)\ u(x^O,y^*,z)$$
$$+ \{1 - u_Y^O(x^O,y,z^O)\}\ u(x^O,y^O,z)\big] \quad , \qquad (7.24)$$

where u_Y^* and u_Y^O are also scaled from zero to one.

One can note that given these scaling conventions,

$$u_X(x,y^O,z^O) \equiv \frac{u(x,y^O,z^O)}{u(x^*,y^O,z^O)} \qquad (7.25a)$$

$$u_Y^O(x^O,y,z^O) \equiv \frac{u(x^O,y,z^O)}{u(x^O,y^*,z^O)} \qquad (7.25b)$$

$$u_Y^*(x^*,y,z^O) \equiv \frac{u(x^*,y,z^O) - u(x^*,y^O,z^O)}{u(x^*,y^*,z^O) - u(x^*,y^O,z^O)} \ . \qquad (7.25c)$$

If one plugs (7.25) into (7.24) we see that (x,y,z) is completely specified by assessing the seven consistently scaled one-attribute conditional utility functions illustrated by heavy lines in Figure 7.12.

The actual verification of conditions (i), (ii), and (iii) was iterative in nature. Each additional conditional utility function contributes to a better understanding of the overall structure of the utility

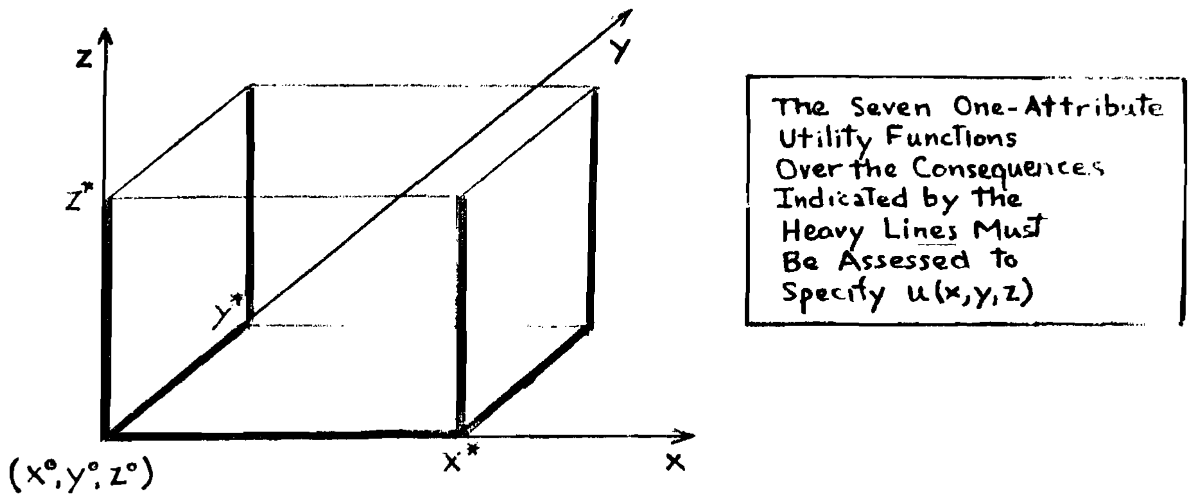


Fig. 7.12 Assessments Required for Evaluating a Utility Function for a Time-Sharing Computer System

function u(x,y,z). The implications of these were discussed with the user throughout the assessment procedure. Whenever there were inconsistencies in the responses of a user, they were pointed out and part of the procedure redone. In all, the utility independence conditions (i), (ii), and (iii) were verified for eight different individuals in the class of users described earlier.

An actual utility function was assessed for only one of these users. The general procedure discussed in Chapters 5 and 6 was used for this purpose. The utility function was assessed over the space $2 \leq x \leq 9$ (seconds), $2 \leq y \leq 120$, and $10 \leq z \leq 100$ (percent).

It turned out that Z was not utility independent of {X,Y} or conditionally utility independent of either X or Y. Grochow states the reason for this: When either response time is at an unfavorable value, for instance, the programmer will be spending most of his or her time contending with the slow response, and consequently will not be as concerned about logging in as when response times are at more desirable levels. The stated reason why Y is not conditionally utility independent of X given Z is that the users may set their relative preferences for response time to compute-bound requests in terms of the response time to trivial requests they are experiencing.

Let us now consider how one might use Grochow's results for making decisions in the computer industry.

Suppose our user was trying to choose among different time sharing facilities which differed not only in terms of X, Y, and Z, but also in terms of their reliability R and their monthly subscription cost S. A proper evaluation here would require a utility function $u'(r,s,x,y,z)$ for the user. However, if $\{X,Y,Z\}$ is utility independent of $\{R,S\}$, then of course, from Theorem 5.6, u' can be expressed as a function of r, s, and u so

$$u'(r,s,x,y,z) = f[r,s,u(x,y,z)] \quad .$$

The original utility function u can be used in a similar fashion if $\{R,S\}$ is utility independent of $\{X,Y,Z\}$ and $\{X,Y,Z\}$ is not utility independent of $\{R,S\}$. Given this assumption, Theorem 5.6 says u' may be expressed as a function of one utility function over $\{R,S\}$ and two utility functions over $\{X,Y,Z\}$ given different levels of $\{R,S\}$. One of these utility functions can be $u(x,y,z)$.

Going one step further, suppose our user (or firm) must decide whether to <u>buy</u> a computer or <u>rent</u> such services. If the choice is made to buy a computer, there may be many options. Clearly, such a decision would involve a time horizon of at least a few years. To remain simple, let us assume that attributes X, Y and Z and a cost attribute are sufficient for the decision. With a five-year horizon, this cost attribute might be $C \equiv \{C_1,C_2,\ldots,C_5\}$ where C_i represents costs in the i^{th} year. Then, as before, with necessary utility independence assumptions between C and {X,Y,Z}, the original utility function u can be used.

Switching gears, suppose the management of a time-sharing service has two objectives: maximize profits and provide the best possible service to customers. A reasonable measure of the quality of service to a user may be its utility function over attributes X, Y, and Z. Hence, given many users, the firm may select a utility function which is a function of annual profits, for instance, and the individual utility functions of its users.

By including potential user's utility functions as arguments of its utility function, the firm may have a tool to help select pricing and service policy. That is, if prices are too high, many users will select competitors and thus reduce the firms profit. If the subscription prices are too low, the firm will also do poorly financially. By maximizing its expected utility, the firm can find the "optimal" price.*

---

*This brief discussion has neglected actions by competitors. The utility functions discussed are applicable in conjunction with game theory, a discipline concerned with these competitive aspects. A basic introduction and survey of game theory is Luce and Raiffa [1957]. A more recent survey is Shakun [1972].

7.6 SITING AND LICENSING OF NUCLEAR POWER FACILITIES

The siting of nuclear power facilities is an extremely complex process. There are many concerned interest groups, each with their own set of multiple objectives, trying to influence the decision making process. The stakes are large, involving hundreds of millions of dollars, possible energy shortages and "blackouts," the possibilities of severe environmental damage, and in some situations, heavy dependency on foreign fuels, to mention a few of the relevant considerations.

In the United States the power company has to prepare its case advocating a particular site or sites and submit these plans for review by governmental regulatory authorites (e.g. power, environmental impact) and by the federal Atomic Energy Commission. These bodies try to reach a decision by weighing the available data, considering the broad tradeoffs, and examining diverse viewpoints: of the power company, of environmentalist groups, of the public as energy consumers, and of local groups, such as the communities near the suggested sites. How can these governmental authorities rationally integrate all the available information in a manner useful for aiding their decision process?

The power companies themselves have difficulties in dealing with the multiple objectives they face. They are, however, mainly concerned with competitive business positions and engineering factors, such as transmission facilities design and network reliability, which directly affect their financial returns. But when a power company is asked, by the regulatory boards, its position on broader questions such as the impacts of its proposals on the environment and local communities, it too must address a broader set of objectives and often their analyses depend on informal and intuitive reasoning. Perhaps with a better understanding and presentation of the fundamental tradeoffs among the conflicting objectives necessitated by each of the alternatives under consideration, the power company might be in better position to select the best alternative in view of its economic objectives, its public responsibility, and the public's requirements. A formal analysis of these considerations may contribute toward reducing the required time of the now lengthy process necessary for approval of nuclear power facilities. The big question is, what are the characteristics of such an analysis and how does one get it done? The literature on this general subject is voluminous but of direct relevance to the techniques and framework introduced in this book, we single out the works by Gros [1974], Papp et al. [1974], Nair et al. [1975], and by Keeney and Nair [1975].

In this section, we do two things:

(1) speculate on the appropriateness of multiattribute utility theory for examining the questions raised in the preceding paragraphs, and

(2) discuss the work of Jacques Gros, who attempts to quantify preferences for nuclear siting problems using results discussed in this book.

The work described below is far from definitive--perhaps it could more appropriately be described as "suggestive research." We do not dwell on important issues, such as: Is the attribute set complete? Whose preferences should be assessed? How does one introduce political relevancies? How does the analysis help (or hinder) conflict resolution? Our purpose is merely to focus on the concepts of the suggestions and to worry little about their pragmatic implementation. Our excuse for speculating on possible uses of a theoretical nature in a so-called "applications" chapter is that we feel the ideas introduced here are important and the framework of analysis may be appropriate to carry out in practice. In this regard, we feel that Gros' accomplishments are encouraging. At the time of this writing, Woodward-Clyde Associates (see Section 7.4) is evaluating the siting and design of nuclear power plants using these same concepts and techniques.

7.6.1 Objectives For Nuclear Power Siting

Each party interested in siting nuclear power facilities will have its own objectives. By and large, however, in each case these objectives might fall under the five categories: environmental, human safety, consumer well-being, economic, and national interest. Let us suppose that the set of objectives listed in Table 7.10 is sufficient for analysis by any of the interested parties, although clearly, there is overlap in this crude list and all of these objectives are not needed by all of the parties. Those objectives of primary interest to the concerned parties are indicated in the table. Also, for future reference the associated attribute--possibly a vector attribute--is designated notationally. No attempt is made to specify specific attributes at this time.

Table 7.10 Some Objectives for Siting Nuclear Power Facilities

Attribute	Category	Objective	Parties Primarily Concerned*
X_1	Environmental	Minimize Pollution	E,L
X_2	Environmental	Provide Aesthetically Pleasing Facilities	E,L
X_3	Human Safety	Minimize Human Health Hazards	E,L,P,S,F
X_4	Consumer Well-Being	Provide Necessary Power	C,E,P,S
X_5	Consumer Well-Being	Minimize Consumer Power Costs	C,S
X_6	Economic	Maximize Economic Benefits to Local Community	L
X_7	Economic	Maximize Utility Company Profits	P
X_8	Economic	Maximize State Revenues	S
X_9	Economic	Improve Balance of Payments	F
X_{10}	National Interest	Reduce Dependency on Foreign Fuels	F

*C ≡ consumers; E ≡ environmentalists; L ≡ local communities; P ≡ power company; S ≡ state agency; F ≡ federal agency.

7.6.2 A Conceptual Framework for Analyses by the Interested Parties

The utility functions briefly discussed here are mainly to suggest a conceptual framework for thinking about crucial preference aspects of the nuclear power siting problem and for communicating these preferences to other interested parties. For brevity, we skip a discussion of the utility functions of the consumers, environmentalists, and local community interests. These are, in theory, more straight forward than the cases we do consider.

The Power Company's Point of View. One might simply say that a power company is concerned only with maximizing its own profits. If such were the case, it would be appropriate to assess the company's utility function $u_P(x_7)$ over attribute X_7 and use this in evaluating the power company's alternatives. However, in this era of broader corporate interest and responsibility, it is more likely the case that the company is also interested in satisfying its consumers preferences for energy, minimizing the detrimental environmental impact of its facilities, and maximizing the net benefits of its facilities on local communities in which facilities are to be built. Let us designate attributes for these three additional objectives as U_C, U_E, and U_L, respectively, and note that they can be measured by the respective utility functions u_C, u_E, and u_L. The power company, at least informally, is concerned with its utility function $u_P(x_7,u_C,u_E,u_L)$ over the four attributes x_7, u_C, u_E, and u_L, in order to analyze which of its possible options is most attractive to pursue. Conceptually, one might define utility u_L to be a function of $u_1,\ldots,u_\ell,\ldots,u_N$, where u_ℓ is the ℓ^{th} community's utility function and N communities are considered as possible sites. The power company must weight its subjective judgments about the relative desirability that community 1 has for proposed plant A against the relative desirability that community 2 has for proposed plant B. Such tradeoffs, although terribly difficult, must be formally or informally addressed by the power company.

The State Agency's Point of View. Let us oversimplify once again and assume there is only one state agency concerned with licensing nuclear power facilities, whose main responsibility is nuclear safety. Thus, the objectives of the agency might be to minimize danger due to nuclear radiation, to provide state revenue, and to satisfy the interested groups. Attributes X_3 and X_8 from Table 7.11 may be useful for measuring the first two objectives, whereas u_C, u_E, u_L, and u_P might do for indicating interest group satisfaction. Thus, the state agencies preferences might be conceptualized by $u_S(x_3,x_8,u_C,u_E,u_L',u_P)$, where u_L' is the state agencies aggregation of the N communities' utility functions.

The Federal Agencies Point of View. The main federal agency concerned with nuclear power plants in the United States is the Atomic Energy Commission. Its problem is quite similar to that of the state agency just outlined. The major difference might be the federal concern for the balance of payments, indicated by attribute x_9, and the national dependency on foreign fuels, measured by attribute x_{10}. It may be useful for the federal agency to conceptualize its preferences with the utility function $u_F(x_3, x_9, x_{10}, u_C, u_E, u_L'', u_P)$, where u_L'' measures the federal agencies concern for the local community impact of nuclear facilities.

7.6.3 Empirical Assessments of Gros

Gros [1974] studies nuclear facility siting from a slightly different viewpoint and in the process has generated evidence that the utility functions postulated in the proceeding section can be meaningfully assessed.* Specifically Gros investigates the usefulness of what he refers to as Paretian environmental analysis in nuclear siting decisions. Generally stated, Paretian analysis attempts to identify the benefits accruing to each of the

---

*Ovi's [1973] results using multiattribute decision analysis for evaluating nuclear versus fossil power plant alternatives, and nuclear siting and decision studies in progress at Woodward-Clyde Associates also lend support to this contention.

various parties involved in a decision making process and to illuminate the tradeoffs among these groups. To illustrate his approach, Gros examines the deployment of 1000 megawatt nuclear baseload units to possible sites along the New England coast.

In the terminology of this book, Gros assessed multiattribute utility functions for four parties involved in nuclear power plant siting in New England: power companies, environmentalists, regulatory agencies, and local groups. These utility functions were each assessed over four attributes:

$Y_1 \equiv$ Capacity at a site, measured by the number of 1000 megawatt units at a coastal site,

$Y_2 \equiv$ Incremental dollar costs, measured by the cost of thermal abatement equipment plus transmission costs expressed as a percent of the minimum cost facility,

$Y_3 \equiv$ Radiation hazard, measured by the population within fifteen miles of the nuclear facility times the number of units at the site, and

$Y_4 \equiv$ Thermal pollution level, measured in degrees Fahrenheit at the outfall of the nuclear facility.

These attributes were generated after interviewing a number of individuals who had previously participated in siting controversies. Rather than focusing on questions of whether or not this set of attributes is appropriate for the problem considered, let us consider the assessment procedure.

For each of the four interest groups, a knowledgeable observer, who had an intimate knowledge of many of the group members preferences, was chosen based on recommendations of group members. The knowledgeable observer's utility function was assessed and his preferences were used as those of the appropriate representative group. The results were verified for reasonableness with other group members. For each of the four utility functions, necessary utility independence conditions were verified to invoke Theorem 6.1 implying the appropriateness of either the multiplicative or additive utility functions.

Gros was also interested in preferences over the forty year design horizon. For each of the knowledgeable observers he verified that preferences for lotteries in any individual year were utility independent of preferences for lotteries over the other years. Also he found preferences in each pair of years to be preferentially independent of preferences in other years. Hence the forty-attribute utility function, representing the forty-year period, was again either multiplicative or additive. Because of the desire on the part of the knowledgeable observers to spread risks over the years, the multiplicative form was selected as appropriate.*

---

*Some interesting assessments of preferences over time indicating some of these issues are found in Chapter 9.

Gros' efforts and empirical assessments help to illustrate something that we firmly believe. Namely it is possible to develop meaningful utility functions, such as those postulated in the preceding subsection, for the various participants in the complex decision processes concerning the siting and licensing of nuclear power facilities. The assessments briefly discussed here are an important first step toward characterizing utility functions directly useful in making nuclear power siting decisions. The task is difficult and the effort required to obtain these preferences is substantial. However, to avoid these problems relegates the crucial tradeoff issues and the preference evaluation of the risks involved to informal analysis.

7.7 OTHER APPLICATIONS

Experience with formal quantification of preferences in multiattribute contexts is growing. Let us briefly mention a number of decision problems, in addition to those in earlier sections in this chapter, where the concepts of Chapters 2 through 6 were utilized.

7.7.1 The Safety of Landing Aircraft

The safety of landing an aircraft depends on many factors: wind, visibility, ceiling, other aircraft in the vicinity, etc. Yntema and Klem [1965] attempted to quantify the safety of various situations which differed in terms of ceiling, visibility, and amount of fuel that would remain at touchdown given a normal landing. Other relevant factors were fixed at a standard value.

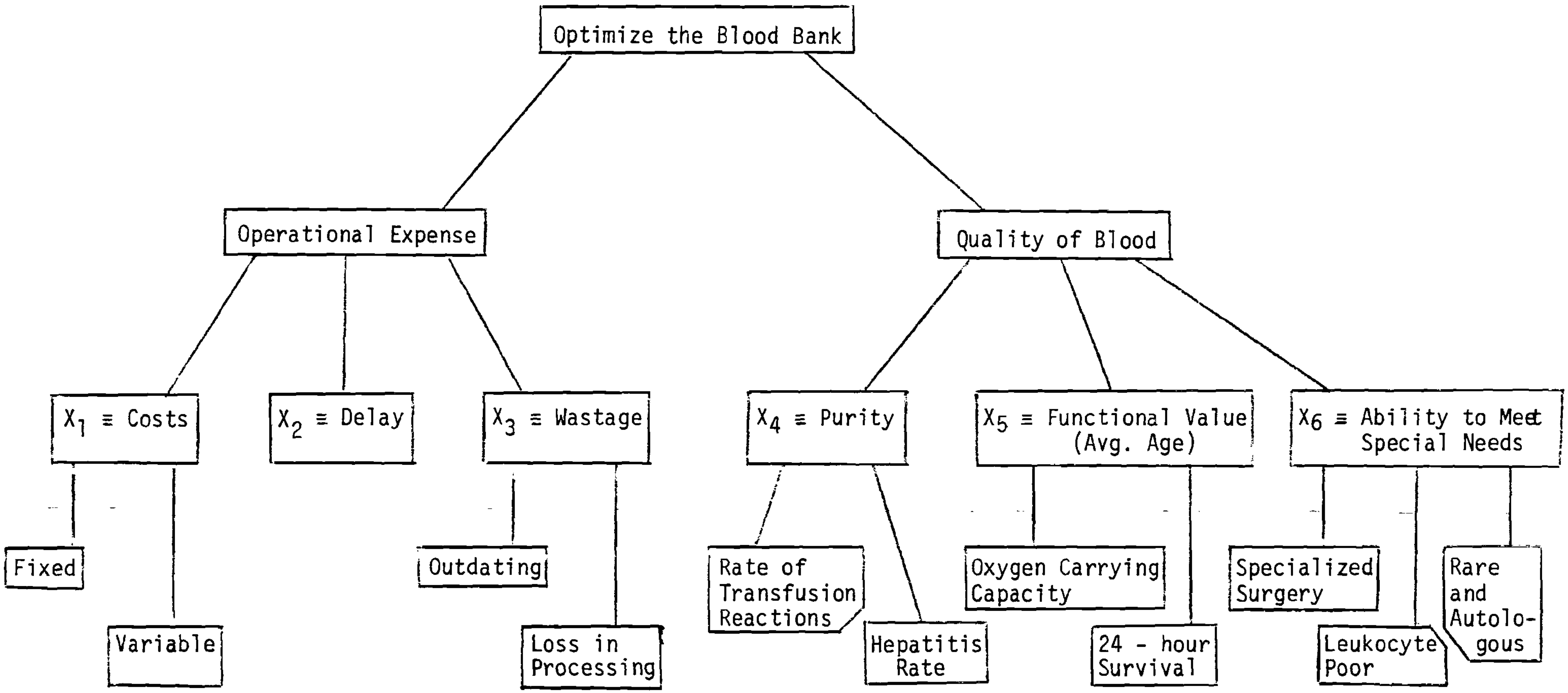
The decision makers for this study consisted of twenty Air Force pilots, each of whom had a good deal of experience in landing aircraft under a wide variety of situations. Using the form of the three-attribute quasi-additive utility function discussed in Result 2 of Section 6.2, utility functions over the attributes ceiling, visibility, and remaining fuel were assessed. In the attribute space, ceiling varied from 100 to 5,000 feet, visibility from 0.25 to 5 miles, and remaining fuel from 15 to 250 gallons. Each decision maker was also presented with forty pairs of consequences and asked to pick the preferable one of each pair. These responses were compared with the implications of each decision maker's utility function. Yntema and Klem concluded "The results were satisfactory."

It should be pointed out that the utility independence assumptions requisite for Result 2 of Section 6.2 were not empirically verified. In fact, the assessments of Yntema and Klem were completed a few years before the formal theory was developed. In spite of this, the resulting utility functions did seem appropriate to represent the preferences of the pilots. Yntema and Klem's pioneering effort gave some support to the contention that it was reasonable and practical to quantify preferences in multiattribute situations.

7.7.2 Strategic and Operational Policy Concerning Frozen Blood

Should a hospital blood bank or system of blood banks invest in expensive blood freezing equipment? And for systems with such capabilities, what are the most desirable proportions of frozen and non-frozen blood? These questions were addressed in a thesis by Bodily [1974]. He also conducted a preliminary investigation of national strategies in blood research and in the usage of frozen blood.

First, after considerable consultation with blood bankers, objectives and measures of effectiveness were specified for evaluating frozen blood issues. The resulting list, given in Table 7.11, indicates the depth at which preferences and probabilities were initially going to be assessed. However, to help the respondent's thinking about the implications of various levels of the attributes, the objectives hierarchy was developed and qualitatively extended as illustrated in Figure 7.13.



A possible configuration for aggregation of attributes for a frozen blood decision problem.

<u>Figure 7.13</u>

Table 7.11 Objectives of a Hospital Blood Bank

Objectives	Measures of Effectiveness
Meet all requests for blood	Average delay or frequency of delay above some acceptable cutoff
Provide high quality blood	Average age at transfusion
Minimize disease	Rate of hepatitis
Minimize cost	Cost/unit
Minimize transfusion reactions	Rate of transfusion reaction
Minimize wastage	Outdating plus processing loss
Provide bloods for special uses	Fraction of special needs that are met with frozen blood or an equivalent unit.

In analyzing the problem of what proportion of frozen blood should be selected for a particular blood bank and the problem of whether or not such a bank should invest in blood freezing equipment, Bodily used a variety of approaches to obtain reasonable probability distributions over the attributes for each alternative. These approaches included utilizing empirical information from blood banks, projections using simple mathematical models of the operations of such blood banks, judgmental estimation of experts, and statistical data available in journal articles, etc.

Bodily tried to assess preferences over the six attributes labelled X_i, $i = 1,2,\ldots,6$ in Figure 7.13 for a number of individuals concerned with blood banking. A first conclusion was that attribute X_6 could be eliminated in considering the question of proportion of blood to freeze. The reasoning was that if a blood bank froze any blood, it would certainly freeze enough to satisfy special needs, and so, the objective "meet special needs" would be equally satisfied with all the viable alternatives. Hence, it could be dropped from the list.

Next attributes X_1, X_2, and X_3 were aggregated since each pair of these was preferentially independent of its complement and substitution rates were constants. Wastage and delay were translated into economic terms using a simple additive value function

$$v(x_1,x_2,x_3) = x_1 + dx_2 + wx_3 \quad ,$$

where d is the equivalent cost per unit of blood delayed and w is the cost per unit wastage. If attribute Y is defined as $X_1 + dX_2 + wX_3$, then what is needed is a utility function $u(y,x_4,x_5)$ over Y, X_4, and X_5.

In the assessment process, it became clear that blood bankers considered the possible range of the average age transfused much less important than the ranges of economic and purity considerations. Hence X_5 was dropped and utility functions $u(y,x_4)$ were completely assessed for one blood banker and one individual with a public health graduate degree and a knowledge of decision analysis. In both cases Y and X_4 were mutually utility independent and so, from Theorem 5.2, the quasi-additive utility function was appropriate. In addition, Bodily ascertained that in a paired comparison of two simple lotteries with identical marginal probability distributions, the blood banker was indifferent. Thus from Theorem 5.4, it follows that the respondent's utility function was additive. Details of these assessments are found in Bodily [1974].

Many of the concepts of Chapters 2 through 6 were explicitly used in the overall assessment process. First, a first-cut hierarchy of objectives was articulated as discussed in Chapter 2, and one objective was then dropped since it was not important enough to influence decisions. Then using preferential independence conditions and the concepts of Chapter 3, a value function over three of the attributes was specified to achieve an aggregation and reduction of dimensionality. Next quantitative considerations led to the exclusion of attribute X_5. Finally, utility independence and the unidimensional assessment techniques surveyed in Chapter 4 were used to specify the final utility functions.

This case illustrates well a typical evolutionary process which starts from a listing of objectives--in this case the specification of subobjectives extended further down the hierarchy than the quantitative analysis--and terminates with the quantification of the final utility function.

7.7.3 Sewage Sludge Disposal in the Metropolitan Boston Area

In Boston, the Metropolitan District Commission (MDC) has responsibility for water and sewage works for the forty-three cities and towns within its jurisdiction. As of 1971, one hundred tons of sewage sludge was being discharged daily into Boston Harbor by the treatment plants of the MDC. Because of increasing public concern and the interest of the U.S. Environmental Protection Agency, the Division of Water Pollution Control of the Massachusetts Department of Natural Resources requested the MDC to make a comprehensive study of new and better alternatives to the present sludge disposal practice. The MDC organized a committee named the Boston Harbor Pollution Task Force (BHPTF) to study the problem and make recommendations. At the suggestion of the Massachusetts Office of Environmental Affairs and with the consent of the BHPTF, Dennis Horgan, then a graduate student at M.I.T., worked with this task force and conducted an independent decision analysis of the sludge disposal alternatives. This subsection briefly surveys Horgan's work.

The viable alternatives for Boston sludge disposal could be categorized as being either marine disposal or land disposal. In the former category one could either barge sludge to a dumping ground ten miles offshore or extend a submerged sludge line approximately seven miles out to sea. In the latter category one could directly spread the sludge on available land and till it into the soil, or alternatively, the sludge could first be incinerated--thus reducing its volume approximately seventy percent and then disposed of at a land site. There are variations of these four basic alternatives, such as different processes of incineration, etc., but, these were felt to be second-order considerations and not explicitly considered in the analysis.

The analysis by Horgan specified four major objectives: minimize costs, minimize water pollution, minimize land pollution, and minimize air pollution. Thus, the classic question concerning tradeoffs of one kind of pollution against another was explicitly addressed. The net present value of costs was used as the measure of effectiveness of the cost objective. Air pollution was measured in tons of particulate matter and gases due to sludge incineration and land pollution was measured in terms of the total area required for sludge disposal sites. To indicate water quality, Horgan defined a subjective index, as discussed in Section 2.3, scaled from zero to ten, based on state water quality standards.

Exploiting probabilistic independence properties where appropriate, probability distributions were specified over the four variables for each of the four basic alternatives.* Concerning preferences, Horgan verified with members of the BHPTF that each of the four attributes was utility independent of its respective complement, and also, that pairs of attributes were not preferentially independent of their complements. Hence, by Theorem 6.3, the multilinear utility function was appropriate. The specific utility function and probability assessments, as well as sensitivity analysis of the results, are found in Horgan [1972].

7.7.4 Selecting a Job or Profession

A critical decision facing each of us from time to time concerns the selection of a job. This problem is different in one important respect from many of the other illustrations in this book in that it is essentially a personal decision. Most of the other problems dealt with a decision maker as representative of his company or as representative of a branch of the government. Here we will briefly summarize two philosophical approaches to job selection, both of which utilize the general ideas discussed in earlier chapters. The works of Miller [1966,1970] and Teweles [1972] will serve as models for our discussion.

---

*For an incineration alternative, air pollution and land pollution, for example, were not probabilistically independent, since they both depend on the volume of sludge, Horgan's model explicitly included such dependencies.

Miller developed and tested a procedure for evaluating the "worth" of various situations described by multiple attributes. One of the problems to which it has been applied involved a graduate student faced with numerous employment offers immediately following graduation. After preliminary analysis, this number was reduced to four viable contenders. The objectives hierarchy and attributes associated with each of the lowest-level objectives which were identified by the graduate student are illustrated in Figure 7.14.

An additive "worth" function,

$$w(x) = \sum_i k_i w_i(x_i) \quad , \tag{7.26}$$

where w_i measures the worth of an amount x_i of attribute X_i, was used to evaluate the alternatives on a zero to one scale. The scaling factors k_i were determined using conditional assessments as described in Section 3.7. For instance, first weights of 0.33, 0.17, 0.17, and 0.33 were assigned to monetary compensation, geographical location, travel requirements, and nature of work, respectively. Then, for instance, of the monetary compensation, a 0.7 weight went to immediate compensation and 0.3 to future compensation. Of the future compensation, 0.65 and 0.35 went to anticipated three-year salary and anticipated five-year salary respectively. Then, the total effective weight assigned to anticipated three-year

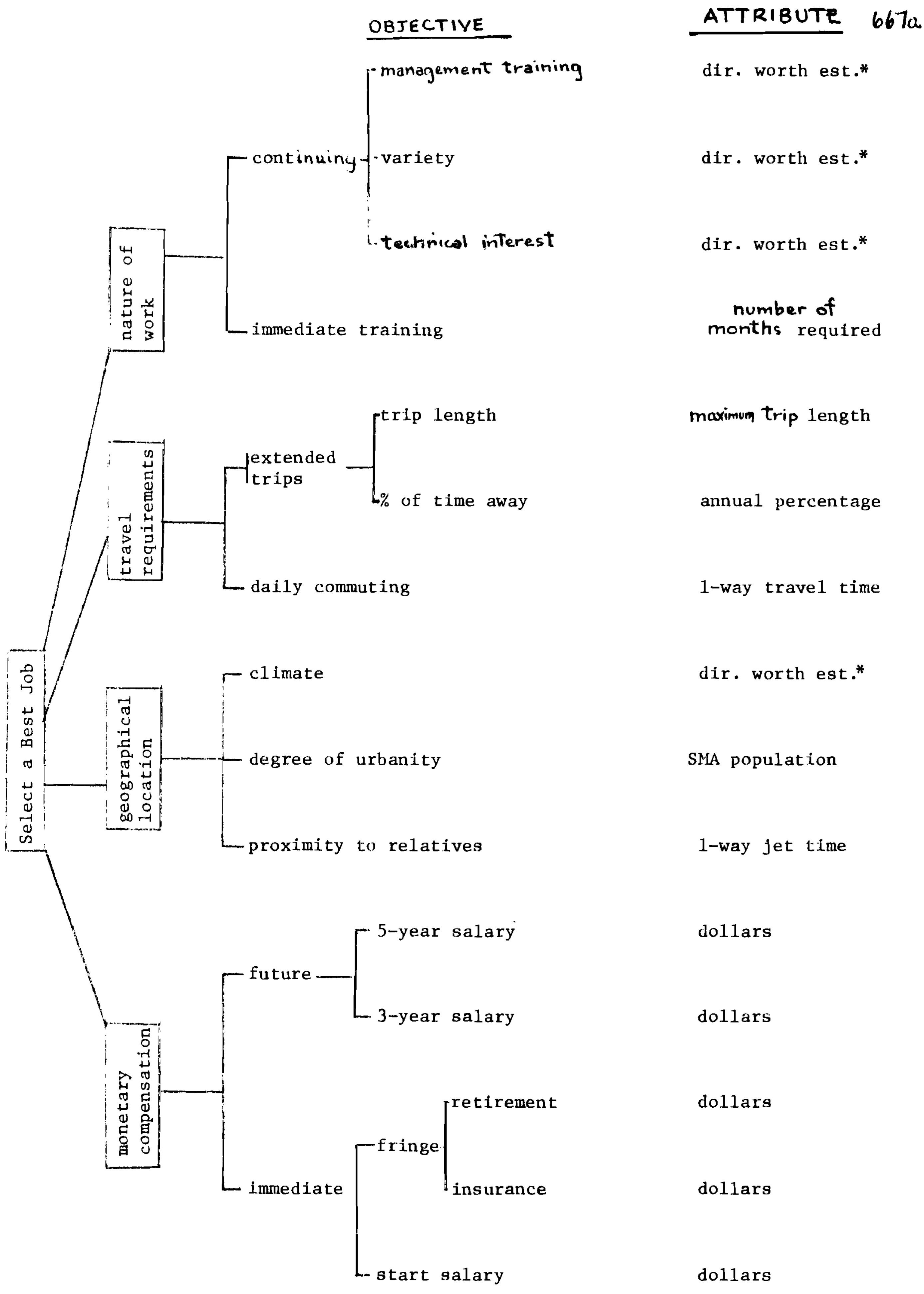


Fig. 7.14. Miller's Objectives for Evaluating Employment Options

*Direct worth estimate: a subjective assessment.

salary was (0.33) (0.3) (0.65) or 0.064. These effective weights were then adjusted to account for the degree to which the attribute actually measured to achievement on an objective. Finally for each of the fifteen attributes, individual worth functions w_i were determined.

The four alternatives were then each represented as a fifteen-attribute vector and the worth of each calculated using (7.26). Uncertainties were not explicitly considered in the problem.

Notice that all the attributes in Miller's problem are in some sense proxy attributes. Presumably, they are proxy for the quality of the decision maker's life. Because of this it was possible to identify many objective measures for these attributes.

Teweles' approach was very different in this respect. He attempted to establish a more direct set of attributes to indicate the desirability of various alternative careers open. Teweles' objectives are given in Table 7.12 along with a short description of the meaning of each.

Table 7.12 Teweles' Objectives for Evaluating Professions

Job Satisfaction--enjoyment derived from doing the type of work you have chosen. Direct benefits of a job such as the opportunity for travel, meeting interesting people, and means of self-expression are included in this factor.

Wealth--the financial remuneration which can be expected from working and the accumulation of capital which can be earned from investment of excess funds. As money is, in a sense, a means of obtaining other goods and services the utility of these products can be substituted for wealth in determining its value.

Security--a condition of relative safety which results from being able to continue your job if you wish to do so. Also included in this factor is the risk to one's health associated with a particularly dangerous occupation.

Family considerations--this factor is an amalgamation of the possible influence a particular career might have on the other members of your family. A wife's attitude, mother's sentiment, child's future, or other considerations should be accounted for in career planning.

Independence--refers to the ability one has of being his own boss and scheduling his own activities. Independence also refers to the short-term flexibility to do what is most important to the individual at a particular time.

Self-esteem--is the self-respect one gains from his own achievements. The self-esteem one could anticipate from a job is very dependent on his ability to be successful at his work.

Prestige--the reputation one acquires within a group as the result of competence, character, power, wealth, etc. The professional respect of one's colleagues may be an important factor to some individuals.

For each of the objectives, except wealth, a subjective index was defined, ranging from zero to one hundred, which was used to indicate the degree to which the corresponding objective was achieved.*

The job alternatives evaluated by Dr. Teweles+ were (1) a private general dentist, (2) a military dentist,

---

*Miller's and Teweles' work illustrates a tendency mentioned in Chapter 2. Namely, as the attributes become more direct indicators of fundamental preferences--as opposed to proxy attributes--it is more difficult to identify suitable objective measures, and one must define subjective indices.

+Dr. Teweles is a dentist, and at the time he wrote his paper, he was completing a Masters of Science in Business Administration and reaching the end of his initial military commitment.

(3) an orthodontic dental specialist, (4) an investment analyst, and (5) a management consultant. These five occupations were evaluated using an additive utility function. Using available data on various professions in addition to personal judgment, Teweles was able to assess probabilities about the degree to which each objective would be met conditional on each alternative. Expected utilities were calculated for each alternative and sensitivity analyses performed.

In Dr. Teweles' report, he states, "The major difficulty in all career planning decisions is for an individual to gain sufficient insight into his own future goals and then learn enough about each alternative to evaluate it objectively." Among Dr. Teweles' conclusions are "As a result of my career analysis, I feel more capable of making the proper career decision at this time. There is no doubt that I understand the factors which motivate me a little better than I did before the analysis." The authors know of many eases where similar personal analyses have been conducted. Some of these resulted in similar conclusions as Dr. Teweles'; other self-analyses, as you might expect, were abortive and useless. We also know of one medical doctor who used this personal self-evaluation technique on a mental patient in a hospital and he reported a surprising success. This doctor took our vernacular phrase, "a framework, for straightening out one's mind," quite literally.

7.7.5 Transporting Hazardous Substances

During the past decade there has been a large growth in the type and the amount of hazardous materials transported within the United States. Shipment of such materials is achieved via all ground modes--rail, highway, water and pipeline. Private citizens, industry, and governmental agencies have become increasingly concerned about the risks associated with transporting these hazardous materials. Aspects of the risk might be divided into two factors:

(1) the likelihoods of various accidents occurring, and

(2) the damage caused by an accident which does occur.

Too often, one has a tendency to assume that "reducing the risks" can always be accomplished by reducing the probability of an accident occurring.* However, one must clearly also include the possible consequences when attempting to reduce risk. Said another way, the risk of the circumstance: "There is one chance in 1,000,000 that a gas leak will lead to a moderate-sized explosion in a populated area next year" seems much greater intuitively than the circumstance: "There are 4 chances in 1,000,000

---

*One can investigate "fail-safe" as well as "safe-fail" techniques.

that a gas leak will lead to a large explosion in the desert next year."

Some pioneering work of Brooks and Kalelkar at Arthur D. Little is currently attempting to measure the relative undesirability of the consequences of various accidents which may result from transporting hazardous materials. In addition, they are investigating which modes of transport are safer for which specific substances.

The aspect of Brooks and Kalelkar's efforts of most interest here concerns their attempts to assess a three attribute utility function over the attributes: human deaths, property damage, and environmental damage. The first attribute ranged from zero to 1200, and the second attribute ranged from zero to ten million dollars. The third attribute was measured by a subjective index scaled from 1 to 13, as defined in Table 7.13.

Table 7.13. Environmental Effects from Hazardous Chemical Spills

Note: This scale applies equally well to water and to land.

1. No effect.
2. Residual surface accumulation of harmless material such as sugar or grain.
3. Aesthetic pollution (odor-vapors).
4. Residual surface accumulation of removable material such as oil.
5. Persistent leaf damage (spotting, discoloration) but foliage remains edible for wildlife.
6. Persistent leaf damage (loss of foliage) but new growth in following year.
7. Foliage remains poisonous to animals (indirect cause of some death upon ingestion).
8. Animals become more susceptible to predators because of direct exposure to chemicals and a resulting physical debilitation.
9. Death to most smaller animals.
10. Short term (one season) loss of foliage with emigration of specific animals that eat the foliage. Eventual reforestation.
11. Death to foliage and emigration of animals.
12. Death to foliage and animals.
13. Sterilization of total environment with no potential for reforestation or immigration of species.

The person whose preferences were assessed by Brooks and Kalelkar was an experienced worker in the field of safety who attempted to take the viewpoint of society as a whole in indicating preferences. It was verified that each of the single attributes were utility independent of the remaining two. Hence, Theorem 6.3 held and the three one-attribute utility functions and the requisite scaling constants necessary for specifying the three-attribute utility function were assessed. The three utility functions are illustrated in Figure 7.15. Details of these assessments are found in Kalelkar et al. [1974].

This analysis raises deep ethical concerns and should be examined critically and constructively by analysts concerned with such problems. At least Kalelkar articulates a utility structure that others can criticize and this is a step forward. Pious, vacuous rhetoric does not help in making such horrendous tradeoffs. We feel that in cases such as the one examined by Kalelkar, implicitly used value and utility structures should be of public concern and should not be suppressed.

7.7.6 Treatment for Cleft Lip and Cleft Palate*

Cleft lip and cleft palate is the second most common congenital deformity in the United States. Treatment for

*Roughly speaking, a cleft lip is a failure of the upper lip to grow together. It usually results in a gap in the lip approximately below one nostril. Cleft palate refers to a split in the palate at birth.

this condition is very involved: it requires many different medical specialists, coordinating from birth to adulthood, not only to correct surgically the physiologic defect, but also to address the child's psychological, social, and mental development. The effects of the treatment of clefts and the effects of the clefts themselves are not completely distinguishable. Both are serious and should be considered in selecting an approach for treatment. With this, a critical issue surfaces, namely: what is the best procedure for treatment in a given situation? Value judgments are essential to answer this question, but because survival of the child is not a factor, various concerned individuals--parents and professionals--often disagree more in their value structures in this situation than in cases where survival is an issue. The best treatment should depend on a number of characteristics, such as the physical features of the child after treatment, the cost, the effects on hearing and speech, etc. Pathbreaking results of Jeffrey Krischer [1974] of Harvard University constitute a very interesting attempt to address some of the critical value issues concerning treatment of cleft lip and cleft palate. Here, we briefly describe his work.

In discussing the importance of cleft lip and cleft palate, Krischer states, " Rarely are there defects so handicapping to the child or so disturbing to the family, yet so amenable to treatment." One major objective of treatment is to correct the physical deformities and provide a normal-looking lip and nose. There are usually uncertainties about the surgical success one will have in this process and there is always the possibility of resulting scars. Defective speech often accompanies those with cleft palate, which can be attributed to both physical and psychological factors. Another complication is the possibility of hearing loss due to a variety of factors. Thus, clearly two other important objectives of treatment are to improve future speech skills and to improve hearing.

Krischer has quantified the preferences of over one hundred people, including surgeons, orthodontists, speech therapists, audiologists, pediatricians, and parents of children with clefts, all of whom are actively involved with individuals having clefts. The four objectives and associated attributes which he explicitly considered are given in Table 7.14 along with the range of these attributes. One unique aspect of these assessments was the attribute evaluating physical effects. Krischer had segments of children's faces showing the nose and mouth area superimposed on a sketched face of a child. These pictures illustrated various degrees of physical deformity after treatment for the cleft. The individuals were asked to assess subjectively their preferences for these pictoral displays. Also note that the hearing attribute only had two values. This, of course, could be generalized. For speech, word intelligibility was measured as the percent

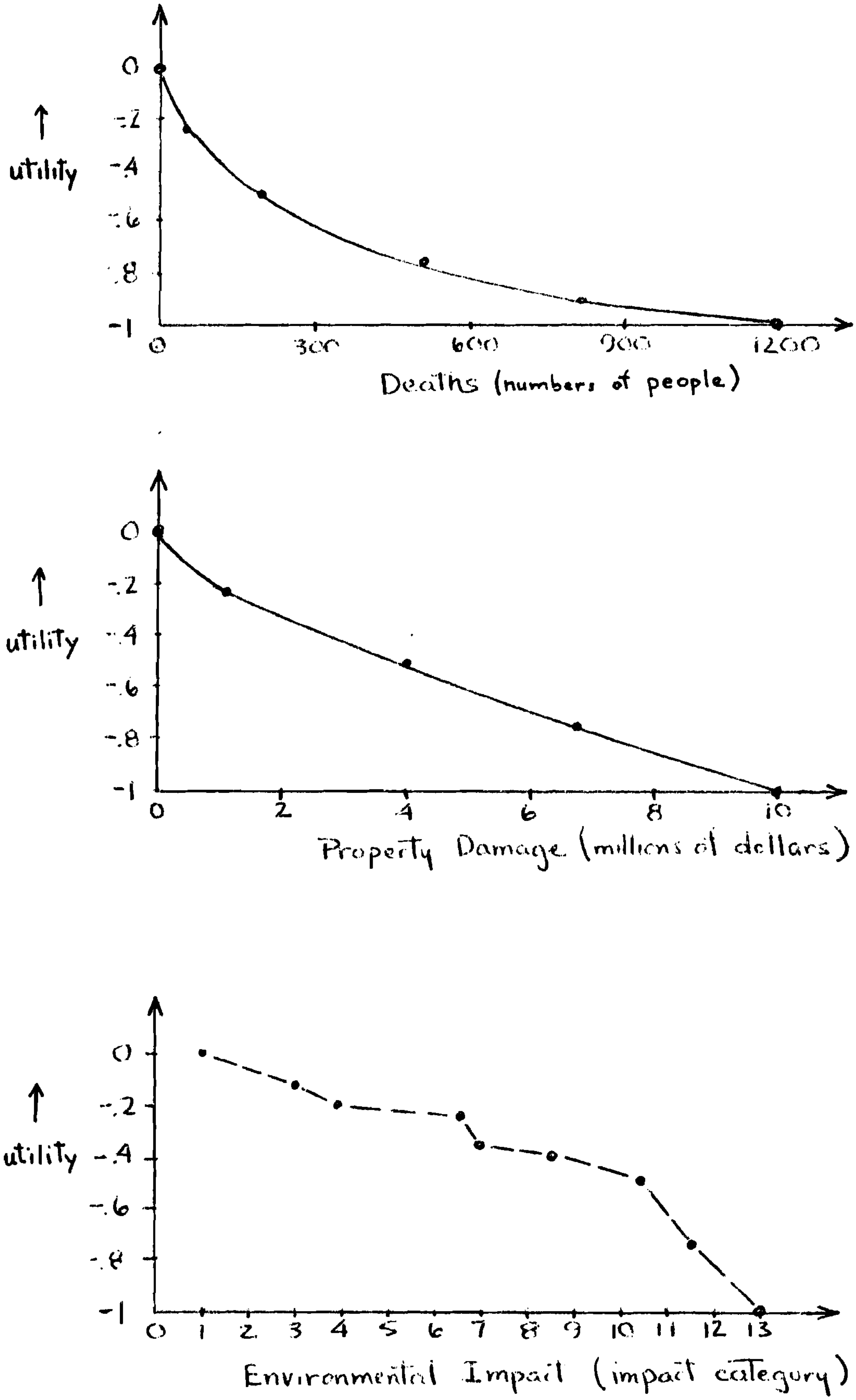


Fig. 7.15. Utility Functions for Evaluating the Transportation of Hazardous Substances

of words accurately identified by a group of listeners with normal hearing. Here 90% is completely adequate, 75% causes mild difficulty in understanding, 50% requires frequent repetition, and 35% is unintelligible.

Once these objectives and attributes were specified, Krischer, working with medical specialists concerned with clefts, developed a questionnaire to assess preferences over the four attributes. This was mailed to medical specialists at numerous cleft-lip and cleft-palate treatment facilities in the United States and through these facilities to some parents of children with clefts. Part of the questionnaire concerned utility independence assumptions and the conditional utility functions for the four attributes and another part concerned preferential independence assumptions and tradeoffs among attributes. Of the first one hundred twenty-five responses, approximately seventy-five percent appear to have accepted requisite assumptions to invoke Theorem 6.1 in formalizing preferences. Details of these assessments, a copy of the questionnaire, and an interesting discussion of individual differences of preferences are found in Krischer [1974].

Table 7.14. Krischer's Objectives for Evaluating Cleft Lip and Palate Treatment

Objective	Attribute	Range
Provide normal looking lip and nose	Pictoral	(see text)
Improve speech	Percent word intelligibility	35 to 90
Improve hearing	Hearing aid required	yes or no
Minimize treatment costs	Dollars	0 - 10,000

7.7.7 Development of Water Quality Indices

Recent work by O'Connor [1973] illustrates some important considerations relevant to specifying and using social indices. O'Connor utilized a modified Delphi procedure (Dalkey [1969]) to combine the judgments of several experts in constructing two separate indices of water quality. One concerned the quality of water to be used as a public water supply and the other described the quality of water for sustaining fish and wildlife populations. Eight experts* were used to (1) specify attributes that should be included in each of the water quality indices and to (2) prescribe a value function over these attributes that would indicate water quality. Since these indices are value functions, they have the property that higher values indicate higher water quality. However, it is not necessarily appropriate to use the expected value of these indices in making decisions when uncertainty is involved.

O'Connor sent questionnaires to and personally visited each of the experts to discuss the attributes which should be explicitly included in some aggregate water quality index and the form of this aggregation function. An additive model was chosen for both the public water supply and fish and wildlife indices. O'Connor emphasizes that an additive model is not appropriate for instance when certain toxic substances enter the water at an unacceptable level or when some of the other attributes, such as pH, reach extreme levels. Thus O'Connor's models are meant to be valid subject to

---

*O'Connor describes the experts as follows: "Eight experts were chosen from an initial set of 20 contacted. Two experts were high-ranking members of The Environmental Protection Agency. Two members were heads of state engineering services departments, and four were university professors in the areas concerned with environmental quality."

the condition that toxic substances are under recommended limits and other attributes are within specified ranges. However, many normal situations probably meet these restrictions. The final attributes used in the public water supply index and in the fish and wildlife index are given in Table 7.15. Details about procedures used and the final value functions are in O'Connor [1973].

Table 7.15. O'Connor's Final Attributes in the Water Quality Indices

Public Water Supply	Fish and Wildlife
Fecal Coliforms	Dissolved Oxygen
Phenols	Temperature
Dissolved Solids	pH
pH	Phenols
Flourides	Turbidity
Hardness	Ammonia
Nitrates	Dissolved Solids
Chlorides	Nitrates
Alkalinity	Phosphates
Turbidity	
Dissolved Oxygen	
Color	
Sulfates	

7.7.8 Examining Foreign Policy

What are the advantages and disadvantages to the U.S. of a Mideast agreement sought to ensure the continued availability of Mideast oil and an increased production to meet the world demand? An exploratory policy analysis done by Decisions and Designs, Inc.* examined how a multiattribute decision analysis might clarify the reasoning and simplify the presentation of conclusions for such a complex problem.

The first phase of the analysis produced a flexible decision model and used it initially to evaluate three sharply different negotiating strategies regarding a possible Mideast agreement. A "base option" involving no change now or later in U.S.-Mideast policies was used primarily as a reference point for purposes of comparison. A maximum option involved an agreement which went most of the way toward what certain Mideast oil-producing countries want. A moderate option was an intermediate strategy reflecting a moderate change in U.S. policy,

---

*Decisions and Designs, Incorporated is an independent research and development company located in McLean, Virginia specializing in decision analysis for the United States Government and industry. Much of their work is devoted to problems involving multiple objectives.

which would be attractive to the Mideast oil-producing countries but not politically difficult for the U.S.

The decision model evaluated the impact of various negotiating postures on Mideastern oil supply and the associated political and economic costs and gains to the U.S. Specifically, the attributes concerned balance of payments, the way Western Europe and Japan would perceive a Mideast agreement, the impact on U.S.-foreign relations, the resulting public sentiment in the U.S., and finally, the effect an agreement would have on other oil producers.

Various sub-models were used to elicit probabilistic judgments and preference assessments at differing levels of complexity and aggregation. The uncertainty side of the analysis was based on judgments elicited from policy makers and substantive experts. Alternative approaches used direct unconditional assessments of oil volume, joint assessments of volume and price, and indirect assessments conditioned on possible political developments. Where different approaches led to inconsistent results, those inconsistencies were resolved by interacting with the respondents.

The preferences used in the problem were solicited from policy analysts charged with making recommendations. For a first analysis, the utility function chosen was additive. The single attribute utility functions for attributes such as "oil volume" were constructed in the manner described in Chapter 4. Tradeoffs were addressed by eliciting statements like, "All other factors held constant, an increase in Mideast oil supply to the U.S. of from .5 to 2.5 million barrels a day at $12 a barrel is indifferent to a gratuitous saving of $4 billion in the federal budget (independent of its level)."

The next phase of the ongoing decision analysis used the model developed, with several variations, to explore a much richer set of realistic options and to update continually the inputs in the light of changing circumstances or perceptions of individual decision makers. More details can be found in Brown and Peterson [1975].

7.7.9 Other Applications

As one can see from the examples described, there is a wide variety of settings in which multiattribute value or utility analysis is being employed. Still our collective experience is not so large that the theory and 'art' of such analyses is anywhere near standardized. Indeed, practically each new analysis contributes to the 'art' of assessing multiattribute preferences, if not to the theory aspects also. For space considerations, we have unfortunately not been able to review many such interesting 'groundbreaking' analyses.

Some of these are Bauer and Wegener's [1975] examination of urban development plans; Gearing, Swart, and Var's [1973,1974] measure of tourist attractiveness and selection of touristic projects for the Turkish Ministry of Tourism; Lorange and Norman's [1973] investigation of risk attitudes of Scandinavian shipowners; Gustafson Feller, Crane, and Holloway's [1971] development of a severity of pain index; Boyd, Howard, Matheson, and North's [1971] decision of whether to seed hurricanes; Dyer, Farrell, and Bradley's [1973] development of curriculum planning information for elementary school principles; and Collins [1975] evaluation of solid waste disposal alternatives in Southeastern Michigan. Huber [1974a] reviews a number of studies that used multi-attribute utility models. Two more general articles which survey the applications of decision analysis in industry are Brown [1970] and Longbottom and Wade [1973].

CHAPTER 8

AIRPORT DEVELOPMENT FOR MEXICO CITY: A CASE STUDY*

This chapter describes the application of decision analysis to a large scale public decision problem--selection of a strategy for developing the major airport facilities of the Mexico City metropolitan area. The purpose of discussing this study here is twofold. First, many of the techniques and procedures developed in earlier chapters of the book are utilized on a very important "typical" problem. Of course, it's typical of those one-of-a-kind strategic decisions which always concern many atypical aspects. Second, although the analysis stresses the value side of the multiattribute problem, it also deals with structuring the problem, aspects of modelling the possible impacts of various alternatives, and the larger framework within which the analysis occurred.

Many people contributed significantly to the study. It was done in the summer of 1971 for the Government of Mexico under the auspices of the Secretaria de Obras Publicas (Ministry of Public Works) and directed by F.J.

---

* This chapter closely follows the development in, and at times takes sections almost verbatim from deNeufville and Keeney [1972] and Keeney [1973a].

Jauffred, Director of the Center for Computation and Statistics, and F. Dovali, Head of the Department of Airports. Richard deNeufville of Massachusetts Institute of Technology and the two of us were consultants assisting SOP on the project. The total time spent by the consultants on the project was fifty man-days.

8.1. THE PROBLEM

Rapid growth in the demand for air travel, combined with increasingly difficult operating conditions at the existing airport facilities compelled the Mexican Government to address the question: "How should the airport facilities of Mexico City be developed to assure adequate service for the region during the period from now to the year 2000?" This was the essential question addressed by the study team.

Our initial problem was not this one however. Two previous studies for developing the airport facilities of Mexico City had recommended very different alternatives. One concluded that the current airport, five miles from the city center should be greatly expanded,* whereas the other suggested moving all aircraft operations to a new airport to be built twenty-five miles north of the city.+

---

* See Ipesa Consultores and the Secretaria de Communicaciones y Transportes [1970].

+ See Secretaria de Obras Publicas [1967] or Wilsey y Ham de Mexico [1967].

Our initial charter was to evaluate the various alternatives, in light of this discrepancy, and to recommend the most effective program for airport development.

For this more limited development decision, one needed to be concerned with the following:

(1) the location of the airport (or airports);

(2) the operational policy defining which services are to be performed and where they will be located; and

(3) the timing for development of different airport facilities.

Because of severe environmental constraints, the two sites previously mentioned are the only ones adequate for a large international airport in the Mexico City metropolitan area. The configurations possible at either site, with respect to the runways for example, were not really significant in this particular problem.

Many different ways of operating the airports -- with substantial differences in the quality of service provided -- were possible, however. In particular, it was necessary to decide what kinds of aircraft activity (international, domestic, military or general) should be operating at each of the two sites.

The question of timing is very important, since failure to act at a given time may preclude future options. For example, land available now may not be available in the future when one might want to develop it. On the other hand, premature action can significantly increase total costs to the nation. The timing issue and operational policies were the most important aspects of this initial airport problem.

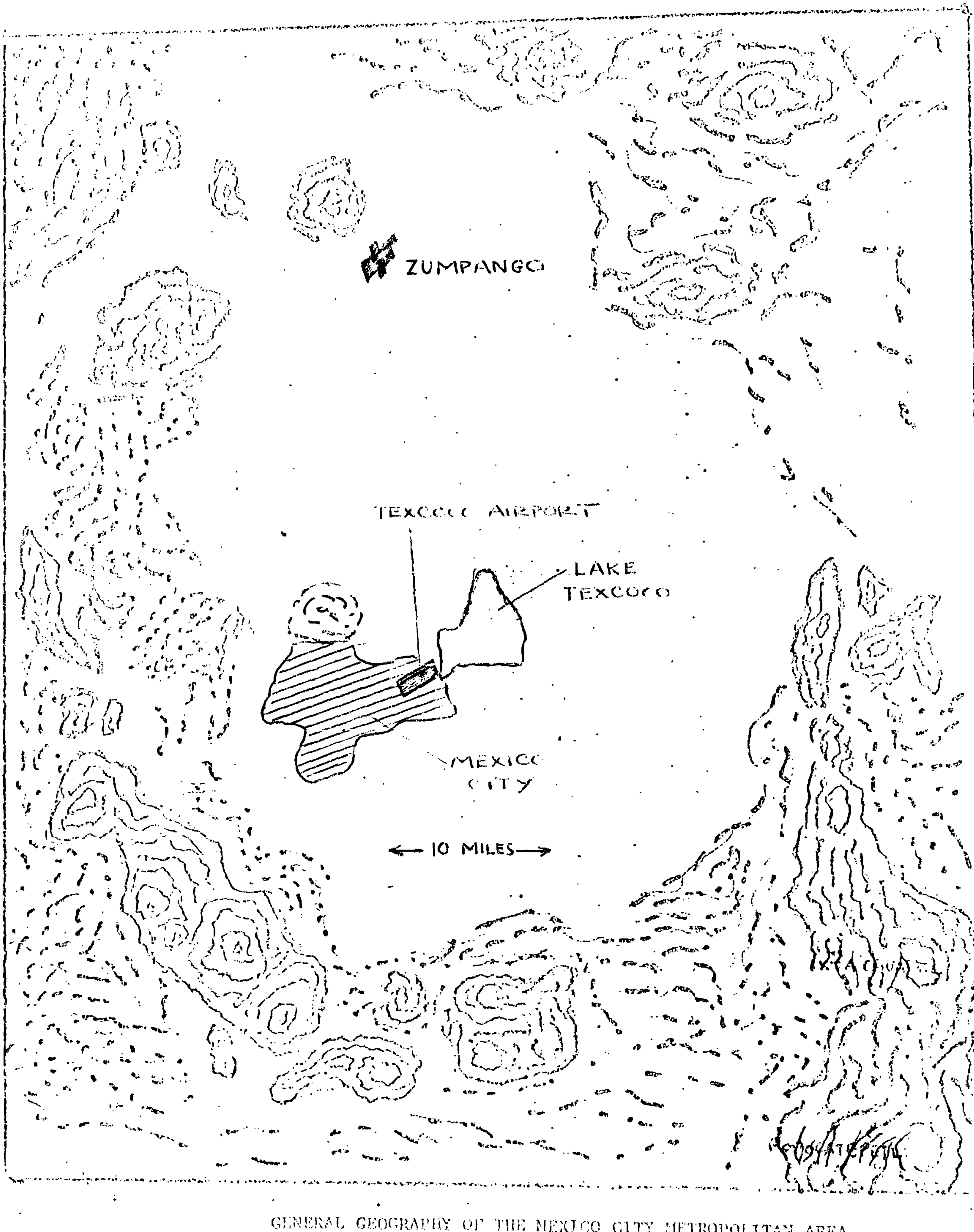
8.2. BACKGROUND INFORMATION

The existing airport is about five miles east of the central part of Mexico City, but still within the city limits on the edge of Lake Texcoco. The other site is 25 miles north of the city in an undeveloped farming area, near the village of Zumpango. The relative location of the two feasible sites is indicated in Figure 8.1.

8.2.1. The Physical Environment in Mexico City

Mexico City is situated at an altitude of about 7,400 feet in a valley ringed with high mountains ranging to over 17,000 feet above sea level. The mountains are very high in all directions except the northeast, where the range lowers to around 10,000 feet. Most flights entering or leaving the Mexico City area fly over these lower mountains to the northeast, although some do proceed through a smaller and higher pass to the south.

The maneuverability of the aircraft at high altitudes is low, especially in hot climates. This requires that the flight patterns over Mexico City be broader than usual and prevents aircraft from safely threading their way through mountainous regions. Thus there are considerable



GENERAL GEOGRAPHY OF THE MEXICO CITY METROPOLITAN AREA
Figure 8.1

restrictions on the usable airspace around Mexico City. This constraint, which principally affects the capacity of the Texcoco site, is serious since Mexico City already handles over 2 million passengers a year and ranks among the the busiest airports on the continent.

When the Texcoco Airport was organized in the 1930's, it was out in the country, but the population of the metropolitan area has grown at the rate of about 5% a year, passing from five million in 1960, to eight million in 1970. During this time, Texcoco has been surrounded on three sides by mixed residential and commercial sections. This has created problems of noise, social disruption, and safety.

Should a major accident occur on landing or takeoff toward the city it would likely cause hundreds of casualties. The area is densely populated and, for example, a large school is located under a flightpath only 500 feet from the end of a runway. Since the approach pattern passes directly over the central parts of the city, high noise levels affect many thousands of people. These noise levels are bound to persist for at least the next 15 years until "quiet" engines are developed and installed on all aircraft. In addition, major expansion at Texcoco could result in displacements of up to 200,000 people. A compensating advantage for the Texcoco site is that major facilities already exist. However they do not meet the standards found in the major airports of other large developed countries.

The location of Mexico City on a former lake bed makes construction especially expensive at Texcoco. Heavy facilities such as runways not only sink rapidly, but at different rates in different locations, depending on their loads. Each of the two major existing runways at Texcoco require levelling and resurfacing every two years. Such repairs closed down half the airport for four months when they were done in 1971. Because the Zumpango site is on higher and firmer ground, it is not expected to have the same kind of difficulties.

Access to the airport by ground transportation appears to be reasonable for both sites. The Texcoco site is near the main peripheral highway which can distribute traffic around the suburbs. It is not, however, especially well connected to the center of the City, to which one has to proceed through congested city streets. The Zumpango site has the clear disadvantage of being further away, but it can be linked directly to the tourist and business areas via an existing north-south expressway.

8.2.2. The Institutional Setting

The government of Mexico has been in the hands of a single party, the Partido Revolucionario Institucional, for almost forty years. Political power tends to be concentrated in the federal government and, for major decisions such as the location of the capital's airport, in the President himself. Any decision about a new airport during 1970-76 will require the approval of President Luis Echeverria. The debate about this decision has been carried on by three major governmental bodies:

(1) The Secretaria de Obras Publicas, SOP, (the Ministry of Public Works);

(2) The Secretaria de Communicaciones y Transportes, SCT, (the Ministry of Communication and Transport), and

(3) The Secretaria de la Presidencia, a body with functions similar to those of the Office of Management and Budget in the United States.

8.2.3. Previous Studies

Both SOP and SCT have commissioned rival large-scale studies of the airport problem within the past few years. The SOP study (SOP [1967], Wilsey y Ham de Mexico [1967]) done for its Department of Airports between 1965 and 1967, recommended that a new airport be built at Zumpango and that all commercial flights be shifted to this facility. The master plan then proposed was not adopted at that time.

The study commissioned by SCT in 1970 (Ipesa Consultores and SCT [1970]) resulted in a master plan for expanding the airport at Texcoco by adding new runway and terminal facilities. Interestingly, this report assumed that aircraft could take off away from the city toward the east, and could land coming into the city from the east in opposing streams of traffic aimed at adjacent parallel runways.

While this proposal "solves" the noise and displacement problems, its implications for safety are extremely serious at any significant level of traffic, and are unlikely to be acceptable for the expected volumes. This report assumed that "quiet" engines would completely eliminate any noise problems outside the airport boundaries by 1990. The SCT study was prepared and submitted during the closing months of the 1964-1970 administration of the previous President. It was not accepted in 1970. The Government of Mexico did, however, wish to resolve the issue. In early 1971 the new administration committed itself to a restudy. As stated by the President in his State of the Union Message of September 1, 1971, "Construction of a new international airport in the metropolitan area (of Mexico City) is also under study at this time." The study referred to is the one presented here.

8.3. EVOLUTION OF THE ANALYSIS

During the short three-month period --the summer of 1971-- in which we analysts were associated with the "airport" problem, it took on many forms. One might say that much of the time was taken defining the problem, but it seemed to be more than this. There wasn't a single problem, but many interrelated problems: What is the best manner to provide acceptable air service for Mexico City? How can one contribute to a reconciliation of differences of judgement, "facts", and opinion of independent government agencies concerned with airport development, in order to improve quality of information available to the decision makers? What strategies for developing the airport facilities are best in light of the financial and political realities facing the government? And so on. The focus of the analysis shifted as the SOP became more sensitized to issues we felt migt be important, as we became more familiar with the total environment in which this analysis was situated, and as segments of the study felt to be important were completed.

Because of the conflicting recommendations of previous reports, the original directive given to our colleagues in SOP was to evaluate various master plans for developing Texcoco and Zumpango. Therefore, this aspect of the problem had to be completed first. Before we entered the scene, SOP had been formulating this problem for a few months. The alternatives were specified and objectives and preliminary measures of effectiveness were defined. Our main effort concerned helping SOP (1) to synthesize the volumes of relevant information in the previous reports, as well as results from additional studies, and to indicate the degree to which various alternatives met objectives; (2) to meaningfully aggregate the effects occurring in different time periods; (3) to quantify a value structure appropriate for the problem; and (4) to develop a system for doing sensitivity analysis and for reporting results.

As this work progressed, the original problem began to be "solved", thus meeting the original directive and freeing the team to address other important issues. Perhaps the most crucial one was to attempt to reconcile the differences of viewpoint held by various parties, especially SOP and SCT, involved in airport development and operation.

8.3.1. An Attempt at Reconciliation through Shared Analysis

It is expected that impartial experts might disagree on many aspects of a complex analysis. It is crucial to know what aspects of the problem they agree or disagree on and why. For instance, there may be agreement on the structuring of the problem, but disagreement on the possible impacts of the various alternatives and disagreement on the value structure. The reasons may simply be that different experts have incomplete information or conflicting information or that traditional viewpoints due to political and professional orientation have been "cast in concrete". The decision analysis model, along with a graphical input-output display developed to assist in the analysis, seemed to offer a useful framework for analyzing these differences of opinion.

Input-output consoles were installed in offices of the study team, the Secretary and Under-Secretary of SOP, the Presidencia, and the President's own office. Our hope was that both SOP and SCT would agree on the basic framework for analyzing the airport problem and that this framework could then help highlight just where fundamental disagreements lay. The Presidencia would then be in a position to better understand the root causes of the different viewpoints, hear the rationalizations of each side, and then commission its own studies if required to clarify critical aspects of the problem. The SOP felt sure that if this reconciliation process were carried out, they would be shown to be right and they were prepared to be quite open -- even about their uncertainties on some inputs. A major problem, of course, lay in the fact that it was SOP who was suggesting the framework (not the Presidencia) and understandably, but regretfully from our point of view, the reconciliation process was never engaged.

Hence SOP had to proceed on a new tack. Clearly their minds were made up about the merits of Zumpango and now their efforts turned to amassing an argument that would convince the President and the Presidencia -- over and above the objections of the SCT. We thus proceeded in the preparation of an advocacy document that was meant to be impressively scientific. Some strange things happened.

8.4. THE STATIC MODEL

Because of the history of the previous studies, the alternatives, objectives, and measures of effectiveness for the static analysis were firmly specified by our clients, the Secretaria de Obras Publicas.

8.4.1. The Alternatives

The alternatives specified what types of aircraft would operate at each of the two possible sites over the rest of the century. In abstracting these, because of similarities in operating characteristics and functions, SOP had categorized aircraft as follows: International (I), domestic (D), general (G), and military (M). It was assumed that at any one time, each category of aircraft could operate at only one of the two sites.

To account for changes in operating arrangements over the thirty-year horizon while keeping the problem manageable, we decided to focus on the three years 1975, 1985, and 1995 as times when changes in the classes of aircraft operating at a site could occur. Thus, an alternative might be "develop the Zumpango site and move general aircraft to it in 1975, shift international to Zumpango in 1985, and operate all classes of aircraft at Zumpango by 1995." Of course, this discretization into three time epochs was done solely to keep the analysis tractable and the actual timing of moves would not be so constained in implementation. We are still discussing a rough-cut level of analysis with presumably more refined tuning coming at a later stage.

Notice that this gives us $(2^4)^3 = 4096$ alternatives. However, many of these were very similar in nature since, for instance, military operations accounted for less than five percent of the aircraft volume. Other alternatives defined as above were unreasonable. One would not move all operations from Texcoco to Zumpango in 1975 and back again in 1985, for example. In the final analysis, the total number of alternatives which were evaluated was approximately one hundred.

8.4.2. Objectives and Measures of Effectiveness

To evaluate the alternatives, one needs to specify some measures of effectiveness which explicitly describe their possible impacts on each of the important groups concerned about the problem. For this problem, the groups might be characterized as (1) the government, as builder and operator of the airports, (2) users of the air facilities, and (3) nonusers. Based on the previous reports of SOP and SCT and lengthy discussions the following six objectives were selected by SOP.

(1) Minimize total construction and maintenance costs;

(2) Provide adequate capacity to meet the air traffic demands;

(3) Minimize the access time to the airport;

(4) Maximize the safety of the system;

(5) Minimize social disruption caused by the provision of new airport facilities; and

(6) Minimize the effects of noise pollution due to air traffic.

Although there is obviously much overlap, the first two objectives account for the government's stake as operator;

objectives two, three, and four for the user's; and the last three objectives for the nonusers. Measures of effectiveness for these objectives were defined as follows:

X_1 ≡ total cost in millions of pesos; with "suitable" discounting;

X_2 ≡ the practical capacity in terms of the number of aircraft operations per hour;

X_3 ≡ access time to and from the airport in minutes, weighted by the number of travelers from each zone in Mexico City;

X_4 ≡ number of people (including non-passengers) seriously injured or killed per aircraft accident;

X_5 ≡ number of people displaced by airport development; and

X_6 ≡ number of people subjected to a high noise level, in this case to 90 CNR or more.*

Clearly, these six measures of effectiveness are not unique or completely comprehensive. For instance, air pollution considerations are absent. However, SOP felt the list did include all the important factors (other than political factors, prestige, etc., which we will discuss later on in

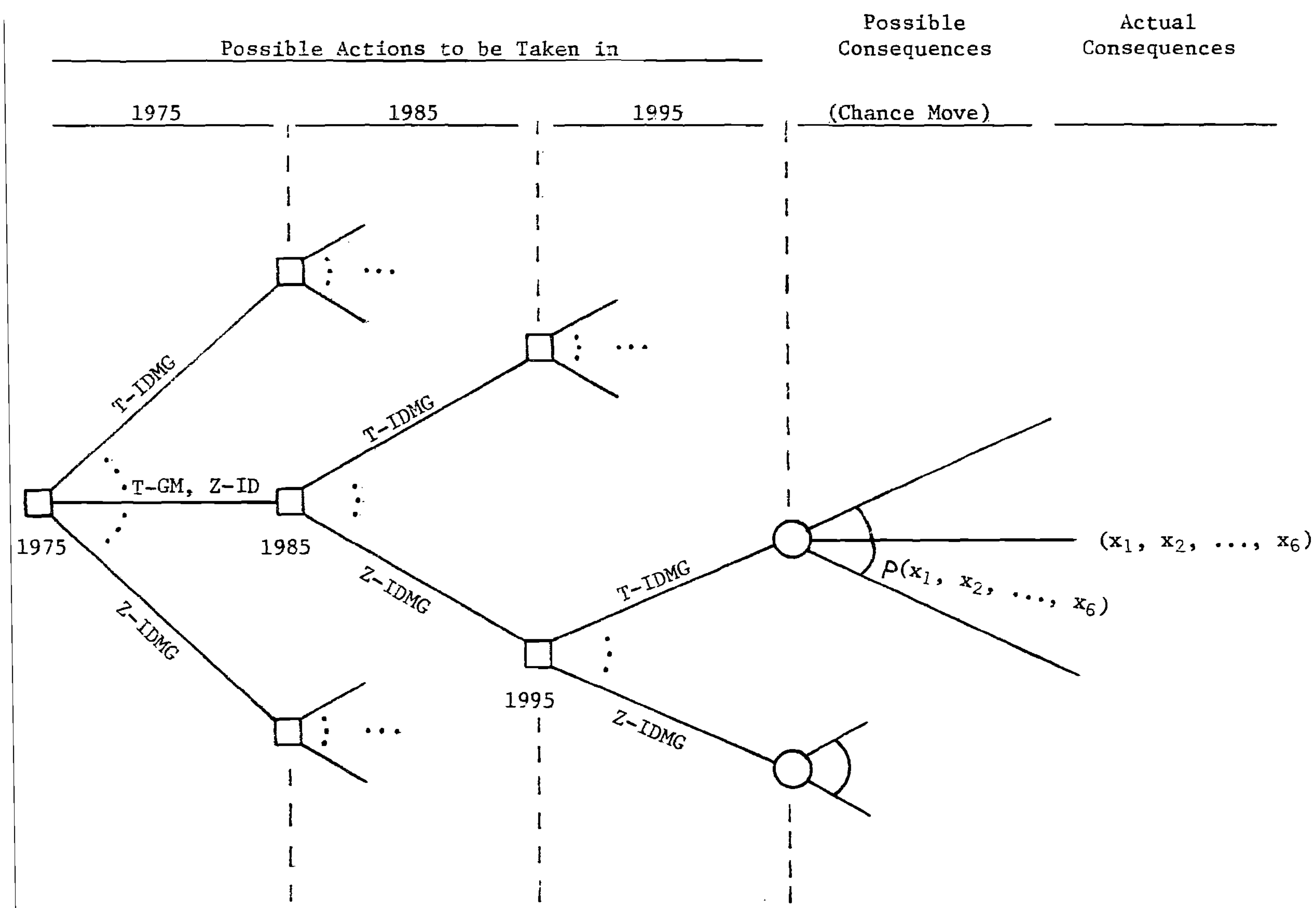
* The Composite Noise Rating, CNR, is a standard index of noise which combines decibel level and frequency of occurence. The 90 level was selected by the SOP Department of Airports.

this chapter) for evaluating effectiveness of the proposed alternatives.

8.4.3. The Basic Decision Model

The basic model is illustrated by the decision tree in Figure 8.2. An alternative is specified by defining what classes of aircraft will operate at which site in each of the three time epochs. As a result of the alternative chosen and events which occur (e.g., demand changes), a consequence $(x_1, x_2, \ldots, x_6)$ will eventually result. However, at the time the decision must be made, uncertainties about this consequence for each possible alternative must be quantified by a probability distribution over the consequences.

The most important point to note about this model is that the alternatives are master plans. They are not designed to adapt to the unfolding of critical events (e.g., demand changes, technological changes, increasing environmental concerns of citizens, etc.) which might occur over the thirty-year period formally considered in the model. Clearly such considerations are essential to any analysis purporting to assist the Government of Mexico in deciding which actions to take in airport development. This was done in the dynamic analysis of options available in 1971 described in Section 8.8. There were two main reasons for first completing a formal analysis of this static problem:



Note: The notation T-GM, Z-ID means operate General and Military aircraft at Texcoco and International and Domestic aircraft at Zumpango.

THE BASIC DECISION MODEL

Figure 8.2

(1) the original request to study the "airport problem" required identifying discrepancies between previous studies, both of which were static analyses, and

(2) without such a study, SOP was very vulnerable to potential criticism of the analysis for excluding the details of such considerations.

The complete description of the probabilistic assessments are given in Section 8.5 , the preference structure is described in Section 8.6 , and the computer input-output along with the results of the anlaysis are given in Section 8.7.

8.5. SPECIFYING THE POSSIBLE IMPACTS OF EACH ALTERNATIVE

The probabilistic assessments were made using the volumes of relevant information from previous studies, the results of parallel studies being conducted by SOP, and the professional judgment of administration within the Mexican Government connected with airport construction, operation, and maintenance. Both reports for SOP (SOP [1967], Wilsey y Ham de Mexico [1967]) and SCT (Ipesa Consultores and SCT [1970]) contain many volumes including detailed demand studies for future air travel, soil mechanics and engineering studies at possible sites, pollution studies considering noise effects, analysis of ground traffic and airport access interaction, cost estimates and projections for various considered airport alterations, etc.. To help in the cost estimates, for each of the sixteen arrangements for aircraft operation at the two possible sites, in each of 1975, 1985, and 1995, general construction plans were outlined indicating where runways, support facilities, and access facilities would have to be built. These plans were used to translate the feasible alternatives specified in Section 8.4 into designs meaningful to airport planners and government officials.

To gain insight on the impacts of alternatives, various experiments were conducted by the SOP. One, designed to gather data on access times, involved dividing Mexico City into ten zones on the basis of residents' pattern of airport patronage, and then studying the driving times to the two airport sites from each zone in different weather conditions at different times of day, etc. This data on travel times and usage characteristics provided the information necessary to assess reasonable distributions for access times for the various alternatives.

In a similar way, detectors were located at various spots in the city to determine the noise levels caused by aircraft. By analyzing current and projected flight paths, superimposed on aerial photos of the city, and the population densities of the affected areas, one acquired a good indication of the noise impacts of various alternatives. These were used in assessing distributions for the number of people subjected to specified noise levels.

By superimposing the various plans for construction on aerial photos of the city, one could easily identify the areas in which people would have to be relocated given that a particular alternative were adopted. The population of those areas was tabulated providing information for assessing the number of people who would be displaced.

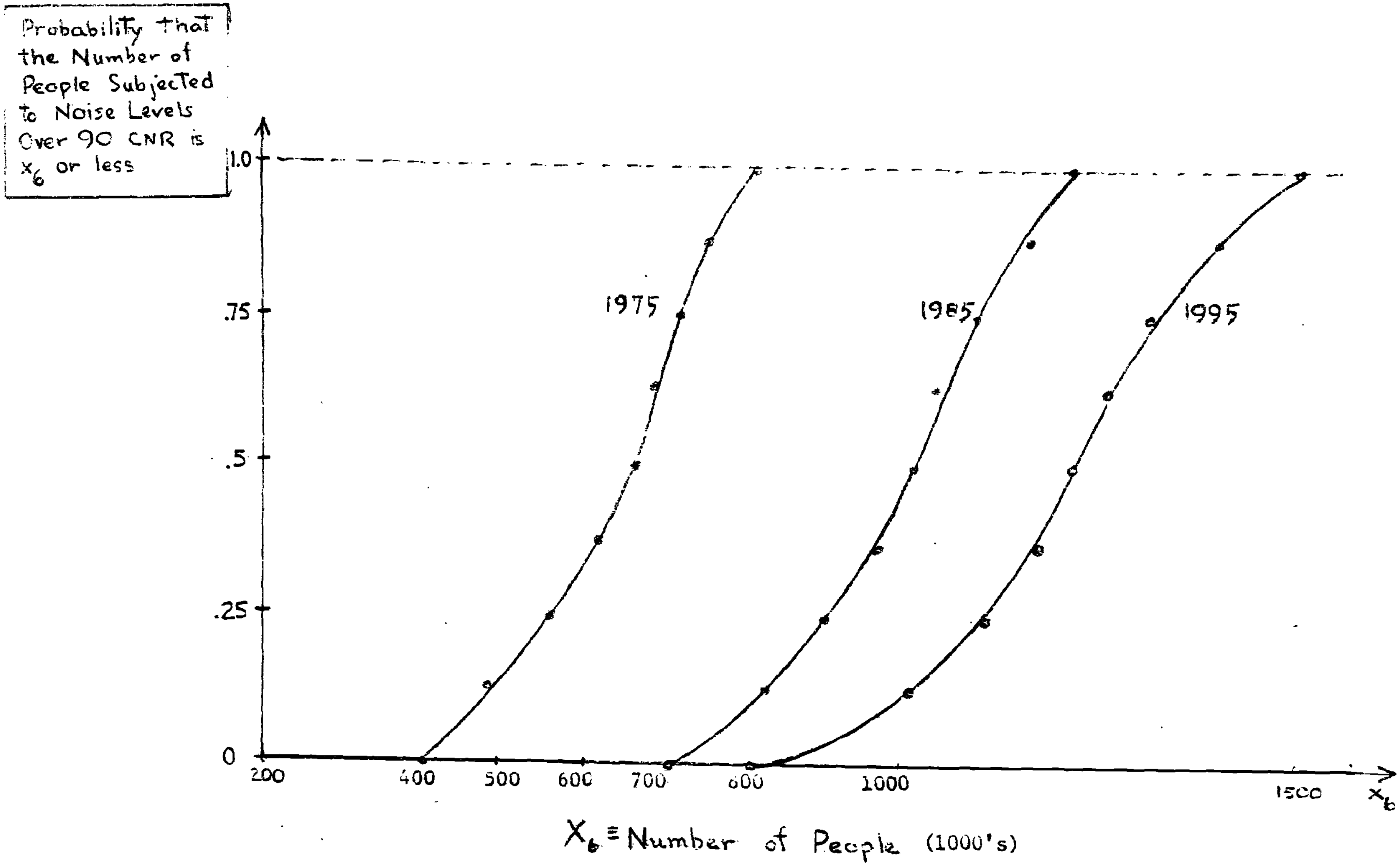
The results of all the previous studies and the data of the concurrent experiments of SOP needed to be integrated to provide meaningful estimates of the impacts of various plans. This integration was done using the professional judgment and experience of members of the Secretaria de Obras Publicas, including the Director of Airports , who is responsible for building and maintaining all the airports in the country of Mexico, the director of the Center for Computation and Statistics, and members of their staffs. The assessments were made in group sessions, where differences in judgments were discussed to arrive at a consensus. The fact that there were no problems in reaching a consensus can probably be attributed to a number of factors: all the professionals had the same information available, all had similar technical training in engineering, they were accustomed to working with each other and knew how each other thought, and the subordinates tended to agree with their superiors.

Having said how in general the probabilistic assessments were conducted, let us get to the specifics. First, the single year assessments will be described, and then the time effects will be accounted for.

8.5.1. One-Year Assessments

The probability density functions were assessed using the fractile method described in Raiffa [1968]. Let us use Figure 8.3 to illustrate the method by example. Consider the possible 1975 noise impact of the operating arrangement "all classes of aircraft at Texcoco." First, the maximum and minimum number of people subjected to 90 CNR or greater was specified as 800,000 and 400,000. Next to 0.5 fractile was evaluated as 640,000. This meant , in the judgment of SOP, the probability that the number of people impacted by 90+ CNR, denoted by X_6^{75}, would be less than 640,000 is one-half. Said another way, it is equally likely that the number of people subjected to the high noise level will be less than or greater than 640,000. The interval between 400,000 and 640,000 was then divided into equally likely parts by choosing the 0.25 fractile as 540,000. The 0.75 fractile was 700,000. Finally, each of the quartiles were divided into equally likely parts in a similar manner.

The fractiles which were assessed are indicated by the dots on Figure 8.3. and the smoothed lines are the cumulative probability distributions describing possible noise impacts for the "all Texcoco" option in years 1975, 1985, and 1995. For any given year, the probability that the impact is between any two adjacent fractile points should be the same, namely 0.125. Thus, to check consistency of the assessments, we asked SOP if in fact their



Assessed Distributions of the Number of People Subjected to Noise Levels Over 90 CNR With the "All Texcoco" Option

Figure 8.3

judgemental probabilities of falling into any of the eight ranges of impact were the same. SOP adjusted their assessments until no more discrepancies could be found. Figure 8.3 indicates the final adjusted curves.

One might ask what are the basic uncertainties which must be considered when assessing the possible noise influence of each airport. First, there is the uncertainty of the population in the flight path area. Current population is known rather accurately, but there is more uncertainty about the population in the future. There is uncercainty about when noise suppressors for jet engines will become operational and incorporated on most jets and about the level of impact of such suppressors. And there is uncertainty about the volume of air traffic in future years. Previous SOP and SCT studies, census figures, SOP experiments, etc., all provided useful information on these basic uncertainties. This information was both formally and informally used by SOP in making their combined assessments for the possible noise impacts.

8.5.2. Incorporating Time Effects

Each of the measures of effectiveness needed to account for the impact over the thirty-year period to the year 2000. Different adjustments seemed appropriate for different measures as indicated:

Costs. The costs that were considered in the model included building and maintenance, but excluded operating costs since it was felt these would be approximately the same for any alternative. As is normal practice for SOP, the present value of the costs was taken as the time dependent attribute of importance. The discount rate used was twelve percent, the standard for the Mexican Government. Sensitivity analysis indicated the choice of a discount rate was not critical for identifying effective strategies.

Noise. For noise, the average number of people annually subjected to aircraft noise levels above 90 CNR was used as the measure of effectiveness. This assumes that it is equally undesirable to have one person subjected to these noise levels for two years or to have two different people subjected in the different years. Furthermore, it assumes the undesirability to an individual of a certain noise level in any year is the same.

Safety. As previously mentioned, safety is measured in terms of the number of people killed or seriously injured per air crash. To adapt this, we chose the average number of people killed or seriously injured per crash averaged over the thirty-year time period. Clearly this measure does not account for the different likelihoods of crashes with various arrangements. SOP was aware of this and of the need to make adjustments to account for this factor. However, they felt it was not prudent to formally include the likelihood of crashes in the model, and chose instead, to make adjustments of the impact per crash in the sensitivity analysis to indicate the effect of differential crash likelihoods.

Access Time. For access time, an average of the possible access times in the various years weighted by the expected number of users in those years was used. This assumes each trip to or from the airport by any individual in any year is as important as any other such trip and that one's preferences for the various access times are stationary over time.

Social Disruption. By reasoning that on the average it would be just as undesirable for a random individual to be moved from his home due to airport development in one year as any other year, we chose the total number of such people displaced to be the measure of social disruption for the analysis.

Capacity. Capacity (maximum possible operations/hour) could not be aggregated in any reasonable way to combine impacts in the different years. This is due mainly to the fact that the relative desirability of various levels of capacity would be very different in different years since demand would probably be larger in later years. Increasing capacity from 80 to 100 in 1975 may be worth very little, since the additional capacity would rarely be needed. However, this same change in 1995 could be extremely important. Thus in the thirty-year model, separate measures of effectiveness for the capacity of 1975, 1985, and 1995 were included.

8.5.3. The Thirty-Year Assessments

By aggregating the three yearly assessed impacts for each measure of effectiveness, except capacity, in the manner just described, we could calculate the probability density functions over the measures to account for impact over time. For instance, with noise, if we difine

$$X_6 \equiv \frac{X_6^{75} + X_6^{85} + X_6^{95}}{3} \qquad (8.1)$$

where X_6^i is the number of people subjected to noise levels over 90 CNR in year i, then by using the probability distributions assessed for the X_6^i for a particular strategy, it is straightforward to derive the probability distribution for X_6. This represents what we've taken to be the overall impact of a particular strategy in terms of noise.

8.5.4. Probabilistic Independence Assumptions

In conducting the assessments over one attribute at a time, we were explicitly assuming that for each alternative, the six attributes were probabilistically independent. For some of the attributes, this assumption seems appropriate. For instance, for any given alternative, noise and access time considerations are probably independent of the other attributes. On the other hand, safety considerations may be dependent on capacity, for instance. The lower the capacity, the more often the airport will be operating under hazardous conditions.

The more important assumption with regard to these assessments was that impacts in separate years were probabilistically independent conditional on the given alternative. This is clearly not true. For instance, for the "all Texcoco" alternative, if we found that 800,000 people were subjected to high noise levels in 1975, we would likely feel that more people will be affected by noise in 1985 than we would have if 400,000 had high noise levels in 1975.

Our analysis was designed in an iterative fashion. First simplifying assumptions (e.g. probabilistic independence) were adopted with the intention at a later stage of recycling back with more realistic assumptions. It turned out, however, that the delicacy taken in modelling the probabilistic part of our analysis was not a critical factor since other considerations dominated, and if we had more time, we would have dressed up the probabilistic analysis to be more credible to the reader. But it would have been mere "window dressing" because the action recommendations we finally suggested could not have been reversed by acknowledging the joint dependence of the random variables involved. It would not have been too difficult to incorporate this complexity -- if not analytically, at least through a simulation mode of analysis. It simply was not worth it in this case.

One could argue that given the oversimplifying probabilistic assumptions and the insensitivities, it might have been just as accurate and simpler to use point estimates of the impacts rather than probability distributions. In retrospect, this seems quite reasonable. However, this does not avoid any of the assumptions made in our analysis, and in addition, no account is made for the possible uncertainty of impact for the single attributes. Our approach forces an explicit recognition of this uncertainty by the decision makers. Also, before our analysis, the lack of sensitivity of the types of effective strategies to the attribute levels was not known. A sensitivity analysis using point estimates could have indicated this, however. The strongest reason for maintaining the detail of using probability distributions was that SOP wanted to avoid potential criticism of the analysis due to exclusion of the uncertainties.

8.6. ASSESSING THE MULTIATTRIBUTE UTILITY FUNCTION

Once we had probability assessments which adequately described the impact of alternate strategies in terms of our six measures of effectiveness, the next step was to assess a utility function $u(x_1, x_2, \ldots, x_6) \equiv u(\underline{x})$ over these measures. Proceeding as suggested in Section 6.6 , we began by exploring the decision maker's preference structure in a qualitative manner. This was to build up SOP's and our own experience in thinking directly about $(x_1, x_2, \ldots, x_6)$ consequences, but more importantly, to ascertain whether any of the preferential independence or utility independence assumptions discussed in earlier chapters were appropriate for this problem. Then we formally verified a sufficient set of such assumptions which allowed us to define for each i, $i=1,2,\ldots,6$, a conditional utility function u_i over X_i and then to construct u as a function of the conditional utility functions. That is,

$$u(\underline{x}) = f[u_1(x_1), u_2(x_2), \ldots, u_6(x_6)], \qquad (8.2)$$

where f is scalar valued. To specify $u(\underline{x})$, the six u_i's and necessary scaling factors were assessed.

The utility assessments incorporated the best professional judgments of both the SOP Director at Airports and the SOP Director of the Center for Computation and Statistics, and members of their staffs. A serious attempt was made to analyze the problem from the point of view of the Government of Mexico.

8.6.1. The Assumptions

To refresh our memory, let us briefly and informally review the concepts of preferential independence and utility independence. Recall that preferential independence concerns only ordinal preferences and no probabilistic elements are involved. Partition the set of attributes into Y and Z. If the rankings of consequences, which differ only in the level of attribute Y, are the same regardless of the fixed level of attribute Z, when Y is preferentially independent of Z.

Utility independence, on the other hand, concerns the cardinal preferences of the decision maker. If the rankings of all lotteries, which differ only in the possible levels of Y which may occur, are the same regardless of the fixed levels of Z then Y is utility independent of Z.

8.6.2. Verifying the Assumptions

Let us illustrate how we verified the preferential independence assumptions used in our work. As an example, consider whether safety X_4 and noise X_6 are preferentially independent of the other attributes. First, we fixed the other attributes at a desirable level and asked what amount of safety x_4 was such that (x_4; 2,500) was indifferent to (1;1,500,000). That is, x_4 people seriously injured or killed given an accident and 2,500 people subjected to high noise levels is indifferent to one person seriously injured or killed and 1,500,000 subjected to a high noise level. After "converging," the amount of x_4 was chosen as 300. The exact number is not important for verifying the assumptions, but our interest is in whether it changes as the other four attributes vary. So we next set these four attributes at undesirable levels and asked the same question and again elicited 300 as response.

Then we asked if this would in general be true for any values of the other four attributes, and the response was "The answer would always be the same given the other attributes were in a static condition." In fact the respondent stated this would be the case concerning any tradeoffs between safety and noise. Hence, we concluded safety and noise were preferentially independent of the other attributes.

By going through identical procedures, we verified that capacity and cost were preferentially independent of the remaining attributes, as was displacement and access time. By this time, the man answering the questions, who was an assistant to the Director of Airports, was in a position to state that ordinal preferences over any two attributes did not depend on the amounts of the other attributes. These conditions were then also verified with other staff members of SOP, including the Director of Airports.

The same general approach was used in verifying the utility independence assumptions -- that X_i was utility independent of its complimentary set $\overline{X}_i$ for all $i = 1,2,\ldots,6$. As an example, consider whether access time X_3 was utility independent of $\overline{X}_3$. The other five attributes were set at desirable levels, and the conditional utility function over access time from 12 to 90 minutes (the range originally specified by SOP) was assessed. We found 62 minutes indifferent to a 50-50 lottery yielding either 12 or 90 minutes. Then we changed the amounts of the $\overline{X}_3$ attributes to less preferred amounts and repeated the question. Again, an access time of 62 minutes was indifferent to a 50-50 lottery yielding either 12 or 90 minutes. A general question indicated this would be true for any fixed amounts of $\overline{X}_3$. We found that relative preferences for any consequences and lotteries involving uncertainties only about access time were indeed independent of the other five attributes.

This condition was verified for all six attributes with both the Director of Airports and members of his staff. In all of these verification procedures, an attempt was made not to lead the respondent to answers he would not have arrived at otherwise. Our opinion is that this was done successfully. Since preferences may vary with time, such questioning of the same people may lead to different conclusions at another point in time. However, the preferences indicated by the individuals questioned appeared to represent their "true" preferences at that time, and hence, the assumptions we made were deemed appropriate for the problem.

8.6.3. Forms of the Utility Function

The main theoretical results used in obtaining the utility function were Theorems 6.1 and 6.2 given in Section 6.3. Informally, these results state that if each pair of attributes is preferentially independent of its complement and if each attribute is utility independent of its complement, then $u(x_1,x_2,\dots,x_6)$ is either an additive or a multiplicative function of the component utility functions $u_1(x_1),u_2(x_2),\dots,u_6(x_6)$. Actually, as indicated in Section 6.3 , this same result is implied by a much weaker set of assumptions - i.e. only one attribute X_i needs to be utility independent of its complement and each pair of attributes including X_i needs to be preferentially independent of its complement. Therefore, many of the assumptions that were verified are redundant, and they can be thought of as consistency checks on the appropriateness of our results.

The exact form of the utility function u, scaled zero to one, is

$$u(x_1, x_2, \ldots, x_6) = \sum_{i=1}^{6} k_i u_i(x_i) + k \sum_{i=1}^{6} \sum_{j>i} k_i k_j u_i(x_i) u_j(x_j)$$

$$+ k^2 \sum_{i=1}^{6} \sum_{j>i} \sum_{n>j} k_i k_j k_n u_i(x_i) u_j(x_j) u_n(x_n)$$

$$+ \cdots + k^5 k_1 k_2 k_3 k_4 k_5 k_6 u_1(x_1) u_2(x_2) \cdots u_6(x_6), \quad (8.3)$$

where u_i is a utility function over X_i scaled from zero to one, k_i is a scaling factor for u_i, and k is another scaling constant. Each k_i must be between zero and one and can be interpreted as the utility u assigned to a consequence with all its attributes except X_i set at their least preferable amount and X_i set at the most preferable amount.

The value of k can be found from the values of the k_i's. When $\Sigma k_i = 1$, then $k = 0$ and (8.3) reduces to the additive form

$$u(x_1, x_2, \ldots, x_6) = \sum_{i=1}^{6} k_i u_i(x_i). \quad (8.4)$$

When $\Sigma k_i \neq 1$, then $k \neq 0$ so we can multiply each side of (8.3) by k, add one to the results, and factor to get the multiplicate form

$$k\ u(x_1, x_2, \ldots, x_6) + 1 = \prod_{i=1}^{6} [k k_i u_i(x_i) + 1]. \qquad (8.5)$$

8.6.4. Assessing the u_i's

Each of the single attribute utility functions was assessed using the techniques discussed in Chapter 4. Let us illustrate this by assessing preferences for access time.

The first step involved obtaining maximum and minimum values for access time. From probabilistic assessments of SOP, we found that the range should go from 12 minutes to 90 minutes, where shorter access times were preferred to longer ones. Thus, to remain consistent with our scaling convention where the utility functions ranged from zero to one, we set

$$u_3(90) = 0 \qquad (8.6)$$

and

$$u_3(12) = 1. \qquad (8.7)$$

From questions to check whether X_3 was utility independent of $\overline{X}_3$, recall that we found 62 minutes for sure was indifferent to a lottery, which we will denote by <12,90> , yielding either 12 or 90 minutes, each with probability 1/2. Hence, the utility assigned to 62 minutes, the certainty equivalent for the lottery, is

$$u_3(62) = 0.5u_3(12) + 0.5u_3(90) = 0.5 \qquad (8.8)$$

Since 62 is greater than the expected access time 51 of the lottery <12,90>, this original assessment indicated that the utility function might exhibit risk aversion. In this context, risk aversion means that the expected amount $\frac{x_3 + x_3'}{2}$ of any lottery $\langle x_3, x_3' \rangle$ would be preferred to that lottery. By asking a couple of questions including specific lotteries and then one concerning the general case, we found that the decision makers were risk averse in the attribute access time. This implied the utility function would be concave as indicated in Fig. 8.4.

By asking more questions to find certainty equivalents of additional lotteries, other points on u_3 were specified. For instance, we found 40 minutes indifferent to <12,62> and 78 minutes indifferent to <62,90>, so

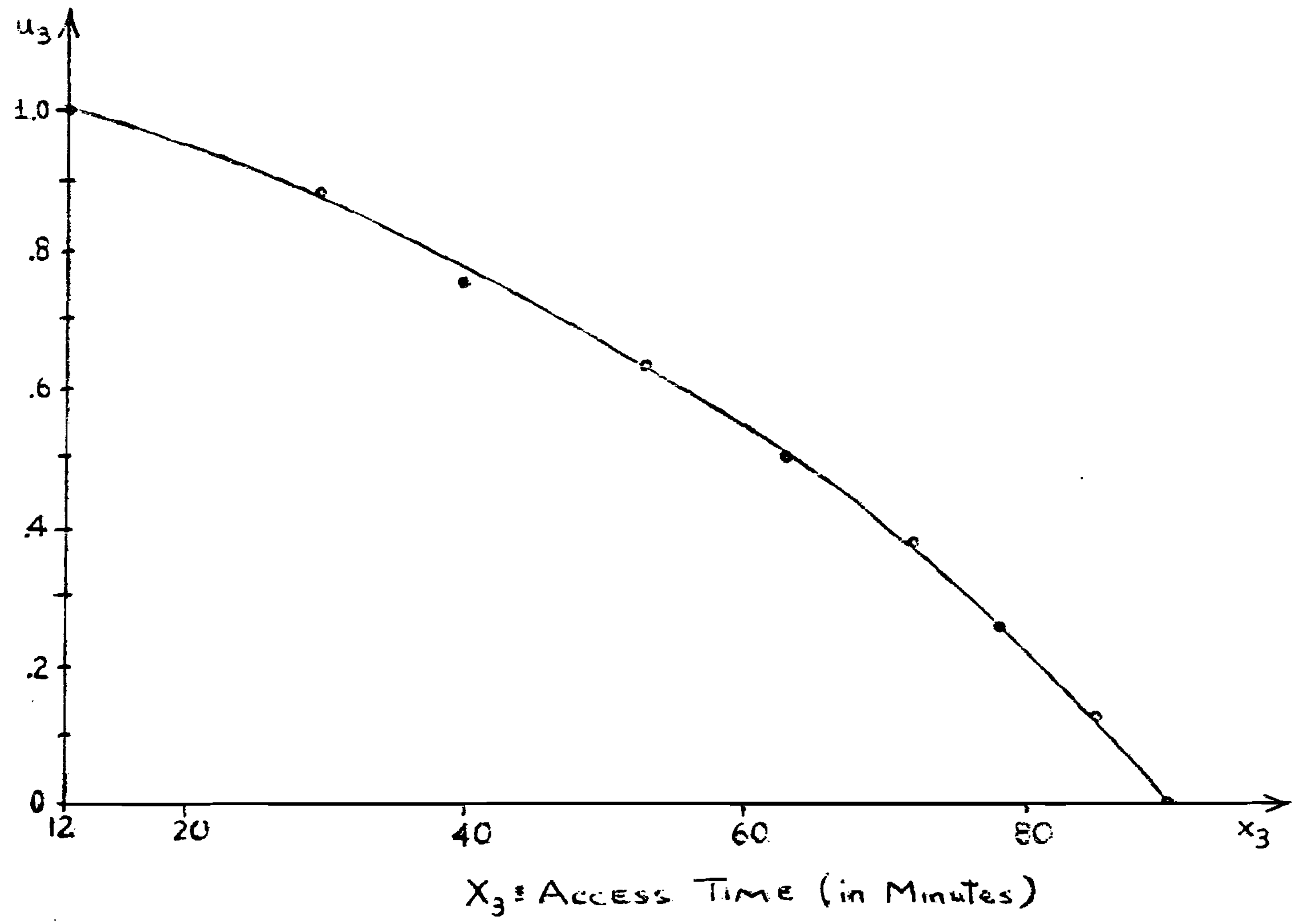
$$u_3(40) = 0.5u_3(12) + 0.5u_3(62) = 0.75, \qquad (8.9)$$

and

$$u_3(78) = 0.5u_3(62) + 0.5u_3(90) = 0.25. \qquad (8.10)$$

Then an exponential utility curve was fitted to the empirically assessed points.

At this stage, we did not immediately try to ascertain and exploit "higher order" risk properties such as decreasing risk aversion. Such properties represent rather fine tunings in a multiattribute utility function relative to the scaling constants "weighting" the levels of the different attributes and more basic properties such as monotonicity and risk aversion of the separate u_i's. If later in the analysis,



The Utility Function for Access Time

Figure 8.4

it had turned out that the precise form of some of the u_i's were important, we would have returned to this aspect and reiterated our evaluation of alternatives. This did not happen to be the case.

Procedures similar to those described above were also used to assess utility functions for cost, safety, displacement, and noise. The results are illustrated in Figure 8.5. However, as mentioned earlier, no single measure was found to combine capacities in different years. Thus, it was necessary to assess the capacity utility function u_2 differently.

Although the general shapes of the utility functions for access time, cost, and noise seem intuitive, the fact that the curves for safety and displacement are linear is not. For instance, concerning safety, one might expect that since governments usually abhor large numbers of deaths resulting from single tragedies the utility function for safety would be risk averse. The reason for this attitude is usually the political impact due to such tragedies. However, our measure of effectiveness in this problem was not meant to capture these political factors. Roughly speaking, if one says each life is equally important, then alternatives with the same expected number of people killed or seriously injured should be equally undesirable in this respect. This was the attitude taken by SOP in the assessments, and so u_4 is linear.

It was important, before proceeding, to do consistency

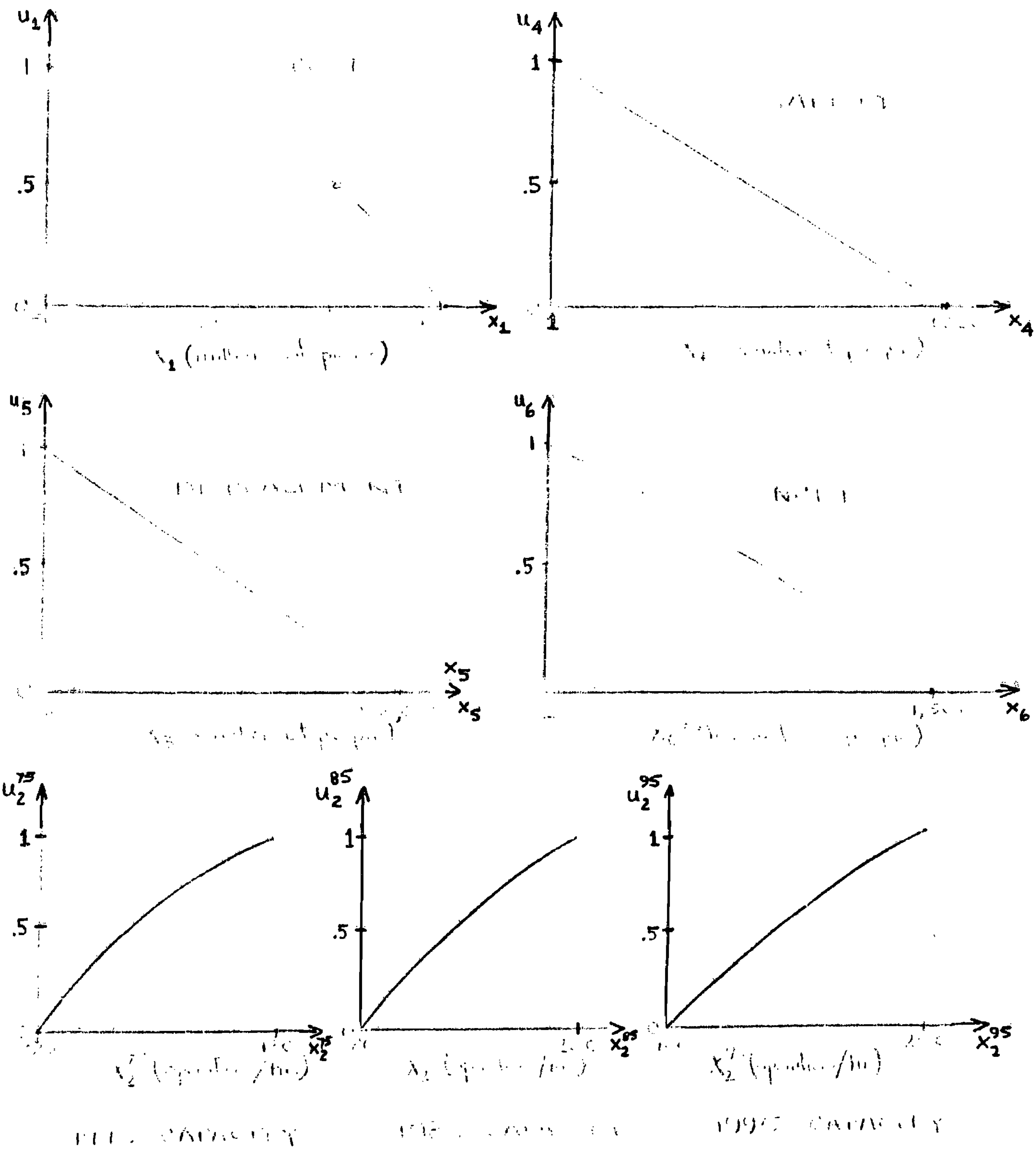


Fig. 8.5 [illegible]
For The Mexico City Airport Study

checks on the reasonableness of the exponential and linear utility functions. This was done by asking additional questions about the decision maker's preferences, and comparing his responses to the implications of the "fit" utility function. When they were consistent with each other, we developed more confidence in the utility function. When they were inconsistent, the inconsistencies were discussed, and part or all of the assessment repeated.

8.6.5. The Capacity Utility Function

Capacity x_2 is a vector $(x_2^{75}, x_2^{85}, x_2^{95})$, where x_2^{75} is the capacity in 1975, etc.

The first step in assessing u_2 was to identify the minimum and maximum possible airport capacities for each year: 1975, 1985, and 1995. There were 50, 80, 100 and 130, 200, 250 operations per hour respectively. Clearly more capacity in any given year was preferred to less capacity, so to scale u_2 from zero to one, we set

$$u_2(50,80,100) = 0 \qquad (8.11)$$

and

$$u_2(130,200,250) = 1. \qquad (8.12)$$

It was verified that each pair of capacity attributes was preferentially independent of the third, and that each attribute was utility independent of the other two.

Thus, we know from Theorems 6.1. and 6.2. that either

$$u_2(x_2^{75},x_2^{85},x_2^{95},) = \sum_{j=75,85,95} c_j u_2^j(x_2^j) \qquad (8.13)$$

or

$$cu_2(x_2^{75},x_2^{85},x_2^{95}) + 1 = \prod_{j=75,85,95} [cc_j u_2^j(x_2^j) + 1], \qquad (8.14)$$

where the u_2^j are the utility functions over X_2^j assessed on a zero to one scale as illustrated in Fig. 8.5 and c and the c_j are scaling constants. Notice that the forms of (8.13) and (8.14) are analogous to the utility functions expressed in (8.4) and (8.5). Since the following discussion concerns how the k_i's and k in (8.5) are assessed, we will not indicate the assessment of the c_j's and c in (8.14) as the procedures are indentical.

8.6.6. Assessing the k_i Scaling Factors

To illustrate the technique for assessing the k_i scaling factors, let us take cost X_1 as an example. We asked the decision makers to compare a consequence with cost at its most preferred amount, and all the attributes at their least preferred amount, to a lottery yielding the consequences with all attributes at their most preferred amount with probability p or the consequence with all attributes at their least preferred amount with probability 1-p. The object is to find the value of p, call it p_1, such that the decision maker is indifferent between the lottery and the consequence. Then, as shown in

Section 6.6, by using $u(\underline{x})$ from either (8.4) or (8.5) and equating expected utilities, k_1 must equal p_1.

Using this procedure involving questions concerning lotteries, we arrived at an initial estimate for the k_i values. Then we used nonprobabilistic questions as consistency checks. For example, we set all attributes at their least desirable level and asked, "Would you prefer to have capacity or cost changed to its most desirable level?" Capacity was the response implying k_2, the coefficient of capacity utility, had to be greater than k_1, the coefficient of cost utility. Then we found a level of capacity, call it $\underline{x}_2^I$, which was indifferent to the best level of cost, denoted by x_1*. Then using either (8.4) or (8.5), we see that $k_2 u_2(\underline{x}_2^I)$ must equal k_1. Since we have u_2 assessed, this gives us a relationship between k_1 and k_2. Pairwise comparison of the k_i's in this manner provided many consistency checks, redundant with others, and forced a readjustment of the k_i values. After several iterations, we ended up using the values of k_i indicated in Table 8.1.

8.6.7. Assessing Parameter k.

Since the sum of the k_i is 1.89, we know the utility function is multiplicative rather than additive; it is additive only if $\Sigma k_i = 1$. Therefore the value of k in (8.5) must be determined by evaluating (8.5) at $(x_1^*, x_2^*, \ldots, x_6^*)$ where x_i^* is the most preferred amount of X_i. This gives us

$$k\ u(x_1^*,x_2^*,\ldots,x_6^*) + 1 = \prod_{i=1}^{6} [kk_i u_i(x_i^*) + 1], \tag{8.15}$$

but from our scaling conventions, we know both $u(x_1^*,x_2^*,\ldots,x_6^*)$ is 1 and the $u_i(x_i^*)$ are all one so

$$k + 1 = (k\ k_1 + 1)\ (k\ k_2 + 1) \cdots (k\ k_6+1). \tag{8.16}$$

Since the k_i are known, parameter k can be evaluated by solving (8.16). As shown in the Appendix 6B, since $\Sigma k_i>1$, the value which k must assume is the solution to (8.16) such that $-1<k<0$. Using the k_i values from Table 8.1, we found $k = -0.877$. Of course, if this were redone from scratch a new k would be found. But it would probably fall closer to -.80 (say) than to .00 or to +.80. In the final analysis, it is important to do sensitivity studies on k and the k_i's.

Table 8.1 SCALING FACTORS FOR THE MEXICO CITY AIRPORT STUDY

Attribute X_i	Scaling Factor k_i
X_1 = Cost	0.48
X_2 = Capacity	0.6
X_3 = Access Time	0.10
X_4 = Safety	0.35
X_5 = Displacement	0.18
X_6 = Noise	0.18

8.6.8. The Utility Function

Procedures identical to those just illustrated were used to evaluate the c_j and c in (8.14). It was found that $c_1 = 0.3$, $c_2 = 0.5$, $c_3 = 0.4$, and $c = -0.46$. These parameters, together with Table 8.1 and $k = -0.877$ and the utility functions illustrated in Figures 8.4 and 8.5 represent the information necessary to specify the utility function $u(x_1, x_2, \ldots, x_6)$. The next section describes how it was used.

8.7. THE ANALYSIS

A computer was programmed to assist in evaluating the alternatives. Computationally, the program was quite simple: given any set of probability distributions and a utility function, it calculated the expected utility for specified alternatives.

To keep the calculations at a reasonable number, as mentioned earlier, many alternatives were eliminated before going through expected utility calculations. For instance, since military aircraft represent a relatively insignificant amount of the total air traffic, most alternatives differing only in terms of the airport for military operations were not considered separately. Secondly, alternatives which shifted certain types of aircraft from the Texcoco site to Zumpango and back again at a later date were excluded.

8.7.1. The Input-Output Display

Graphical input-output consoles were used as an efficient and accessible system for sensitivity analyses and communicating results of the study. This capability was used daily by the SOP, and could also be used by the other interested parties to examine the relative merits of alternative developmental policies. The input-output system allowed any user to use his own probability and utility estimates for evaluating any specified alternatives. There were two options for doing this. Option 1 provided the standard estimates that SOP used in evaluating the alternatives on the console screen. To change these, one just typed in the changes over the SOP estimates. This option was particularly useful for sensitivity analyses. Option 2 allowed the user to enter his own estimates without seeing any others.

The probabilistic estimates of possible impact could be altered by changing the upper and lower bounds on these impacts. For instance, as illustrated in Figure 8.3, SOP's lower and upper bounds on the possible number of people subjected to noise above 90 CNR in 1975 were 400,000 and 800,000, respectively. Merely by typing on the console, one could look at the overall effect on strategy if these were 600,000 and 1,200,000.

To alter the utility function, one changed the scaling factors listed in Table 8.1. Because the meaning of these constants can be easily misunderstood (as discussed in section 5.9) and because of the difficulty in specifying a consistent set of estimates, a short subroutine was developed to assist the user. This routine essentially asked the user on the screen the same questions that we asked SOP in initially assessing the scaling constants. Once a reasonable consistency was achieved among the k_i's, the constant k in (8.3) was calculated. If k=0, the additive form (8.4) was used, and if $k \neq 0$, the multiplicative utility function was used to evaluate strategies. As was the case with the general shape of the probability densities, the individual utility functions u_i could not be changed by graphical input-output. These changes required adjustments in the programs. However, although important, these changes represent fine tunings relative to the options provided for graphically.

Another particularly useful feature of the computer program was a routine which calculated certainty equivalents. Using this routine, the overall possible impact of any alternative could be reduced to an equivalent impact described by a vector of certainty equivalents. Since we assumed probabilistic indepencence* and first-order utility independence (i.e., each X_i is utility independent of its complementary set), from the marginal probability distribution of X_i and the component utility function u_i, it is possible to define the certain equivalent $\hat{x}_i$ by

$$u_i(\hat{x}_i) = E[u_i(\tilde{x}_i)], \ i = 1,2,\ldots,6. \qquad (8.17)$$

Notice that the certainty equivalent $\hat{x}_i$ is independent of the possible impacts on other attributes. Also notice that the certainty equivalent vector $\underline{\hat{x}} = (\hat{x}_1,\ldots,\hat{x}_6)$ does not commit one to any determination of the scaling constants k_i's or k.

If two alternatives A and B are reduced to certainty equivalent vector impacts $\underline{\hat{x}}_A$ and $\underline{\hat{x}}_B$, it is easy to check for dominance. Also, for example, one could investigate exactly how large a change in the impact on attribute X_i of alternative A would be required before it would be less preferred than alternative B.

* If $k=0$ (or close to zero), then u can be taken to be (approximately) additive and only the marginal probability distributions are of relevance. If $k \neq 0$, and joint probabilistic dependence is warranted, then the analysis by certainty equivalents must be considerably modified. One could, however, employ the notion of "conditional certainty equivalence" to some advantage. This was not done.

8.7.2. Effective Strategies

Of the alternatives we did evaluate using expected utility, the top ten according to SOP are indicated in Table 8.2. In the table, the expected utilities are calculated on a scale from zero to one-hundred, where zero utility was assigned to a hypothetical alternative generated by taking the least desirable probability distribution for each attribute from the set of all alternatives. The utility value of one hundred units was assigned to a hypothetical alternative generated by taking the most desirable probability distribution for each attribute from the set of all alternatives. On this scale, the alternative of keeping all aircraft in Texcoco in all three years has an expected utility 5.20.

By looking at Table 8.2, it is clear that two types of strategies are effective. One type might be categorized as the "all Zumpango" alternative and represents building a major new airport at Zumpango as soon as possible. The alternatives in the table involving both International and Domestic aircraft operating at Zumpango in all three years make up this category. The other type of effective strategy is the "phased development at Zumpango" characterized by either International or Domestic aircraft operating in Zumpango in 1975 and then both by 1985 and 1995. All strategies which included keeping a part of the International or Domestic traffic operating out of Texcoco through 1985 did not appear competitive in terms of effectiveness with

Table 8.2 THE BEST TEN ALTERNATIVES

Alternative						Expected Utility	Rank
1975		1985		1995			
Z	T	Z	T	Z	T		
D	IMG	ID	MG	ID	MG	91.23	1
IDMG	-	IDMG	-	IDMG	-	90.90	2
I	DMG	ID	MG	ID	MG	90.79	3
ID	MG	ID	MG	ID	MG	89.30	4
ID	MG	IDMG	-	IDMG	-	88.10	5
ID	DMG	ID	MG	IDMG	-	86.75	6
I	DM	IDMG	-	IDMG	-	86.55	7
IG	IM	IDMG	-	IDMG	-	86.19	8
DG	IMG	IDMG	-	IDMG	-	86.17	9
D		IDMG	-	IDMG	-	85.60	10

To help read the table, the alternative ranked 1 is Domestic aircraft at Zumpango with International, Military, and General aircraft at Texcoco in 1975; and International and Domestic at Zumpango with Military and General at Texcoco in 1985 and 1995.

the two types of strategies outlined above. Of course, these expected utility evaluations depend on two types of judgmental inputs: probability and utility assessments. The ones we used were those of officials of SOP and presumably, if the same analysis were to be made with inputs from officials of the SCT, another ranking of strategic alternatives would result. But more about these reconciliation problems later.

8.7.3. Use of the Analysis

As we indicated earlier in the chapter, the original purpose of the work described here was to identify effective strategies - as measured by our six measures of effectiveness - for developing the airport facilities of Mexico City. It was not to indicate what action should be taken by the Government of Mexico in 1971 to meet its needs. Once the "effective strategies" had been identified, the problem shifted to this second question: What action should be initially implemented?

So far, the formal analysis has included only master plans defining actions for a thirty year period. A more appropriate course would seem to be to make some initial decision and then, based on subsequent events, to revise strategies as necessary. Furthermore, any study which is designed to aid in the selection of an airport development policy for Mexico City must include factors such as political preferences and community priorities. This was the task undertaken in a dynamic analysis of development strategies to be discussed in the next section.

8.8. THE DYNAMIC ANALYSIS

The purpose of the dynamic model was to decide what governmental action should be taken in 1971 which would best serve the overall objective of providing quality air service to Mexico City for the remainder of the century. This model assumed the second step in the decision process could be taken in 1975 or 1976, at the end of the current President's six-year term. The action taken then would depend both on the action taken now and the critical events which might occur in the interim. Our analysis of the dynamic model was much less formal than the one developed for the static model, primarily because of time pressures and the general complexity of the situation.

8.8.1. Alternatives for 1971

We first identified (using common sense) the reasonable alternatives available to the government in 1971. These alternatives differed in the degrees of commitment to immediate construction at the two sites. We chose only four levels of commitment (minimum, low, moderate, and high) giving us the 16 alternatives exhibited in Figure 8.6. Actually, each nominal case in the figure represents a class of specific alternatives. The idea was to do a

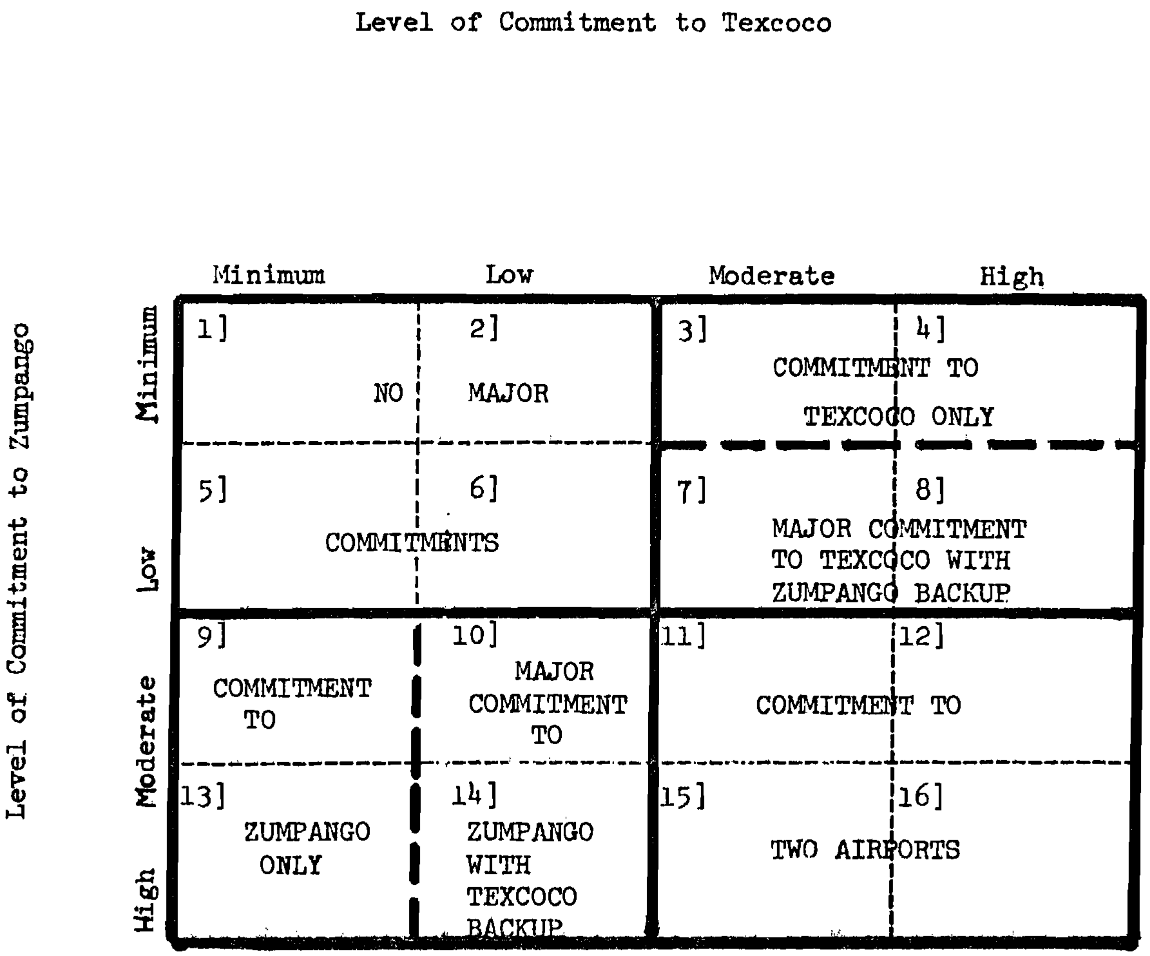


Figure 8.6: THE 16 NOMINAL DYNAMIC ALTERNATIVES FOR 1971

first-cut analysis to decide which classes of alternatives were sufficiently viable to be examined in more detail. It should be noted that the two strategies defined by the static analysis could be compatible with all the nominal dynamic options except 11,12,15, and 16.

The next step involved defining what was meant by the alternatives in some greater detail. Briefly summarized, the alternatives at Texcoco (for the period 1971-1975) were defined as follows:

Minimum - maintenance and introduction of safety equipment only;

Low - extend the runways, upgrade support facilities such as terminals, do all routine maintenance and introduce new safety equipment;

Moderate- in addition to that done with a low strategy, buy and prepare land for building a new runway and expand passenger facilities;

High - build a new runway and passenger facilities, improve the airport access - in short, build a totally new airport at Texcoco.

Similarly, for Zumpango, we defined the commitment levels:

Minimum - at most, buy land at Zumpango;

Low - buy land, build one jet runway and very modest passenger facilities;

Moderate- buy land, build a first jet runway and plan others, build major passenger facilities, and construct an access road connection to the main Mexico City highway;

High - build multiple jet runways, major passenger facilities, and access roads - that is, build a large new airport at Zumpango.

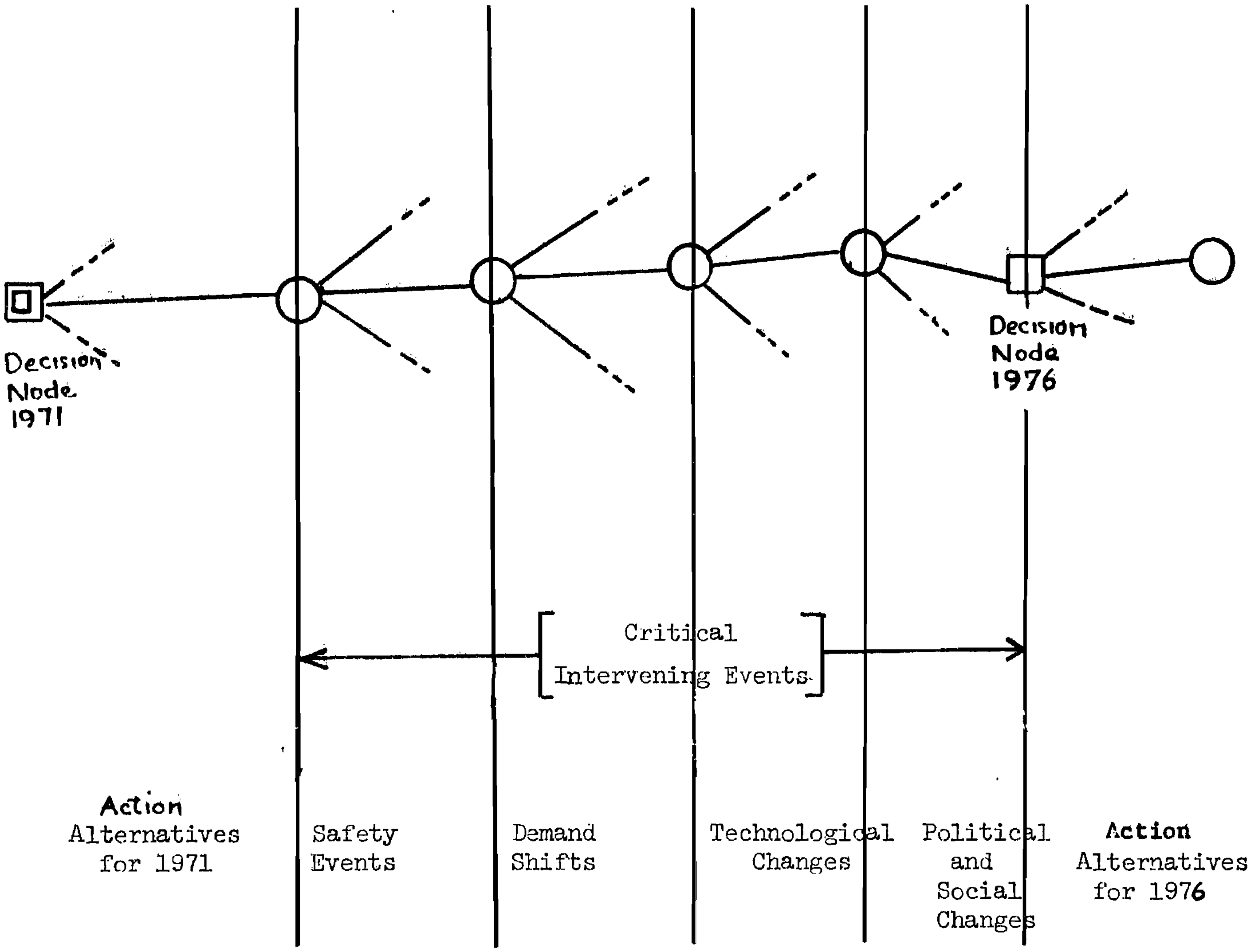
8.8.2. Objectives

We identified four major objectives that were important in choosing a strategy for airport development: effectiveness, political consequences, externalities, and flexibility of the various alternatives. The components of the "effectiveness" attributes are indicated by the six measures of effectiveness covered in the static model. The political consequences were those important to the President - since he was the principal decision maker - involving the political effects which would be felt by SOP, by SCT, and by the Presidencia. Flexibility concerned the range of options open to the President at the second stage of the decision-making process: what freedom would he realistically have at the end of his tenure in modifying his earlier 1971 stance after learning about the intervening uncertain events. Finally, all other important considerations were lumped together as "externalities." These included the amount of access roads needed, the distribution of federal expenditures between the Mexico City region and the rest of the country, the distribution of expenditures for airports and other uses, regional development away from central Mexico City, and the national prestige associated with new airport facilities.

8.8.3. Possible Scenarios

To gain insight into the meanings and implications of each of the classes of alternatives, detailed scenarios were outlined for each. These included: (1) the consideration of important and critical events which could occur in the period 1971-1976, and possibly affect the best strategy in 1976; (2) the likelihood of their occurrences; (3) the strategic reaction to each intervening event-complex; and (4) the possible eventual consequences for each act-event-reaction path. The events involved safety factors and air disasters; shifts in demand in terms of both passengers and aircraft; technological innovations, such as noise suppressors, better runway construction on marshy ground, etc.; changes in citizen attitudes toward the environment; and changes in priorities, such as national willingness to have government funds used for major airport construction. Figure 8.7 depicts a schematic representation of one possible scenario.

In each of the scenarios, the manner in which the 1971 strategy should be altered in 1976 to account for the critical events listed above was defined. For instance, if one originally chose strategy six, then a reasonable response to increased numbers of landings and thus decreased safety, in addition to increased consideration about the impact of noise and air pollution in Mexico City, would be to hasten the building at Zumpango and make it the



SCHEMATIC REPRESENTATION OF A TYPICAL SCENARIO

Figure 8.7

Mexico City International Airport. On the other hand, response to a rather constant demand on the Texcoco facilities and a shift in public priorities toward more medical and educational support from the government, might be to postpone additional construction at Zumpango until a later date.

Suppose that strategy thirteen was initially chosen in 1971, and that air demand greatly increased, environmental concerns of citizens grew, and no technological innovations were developed favorable to Texcoco (e.g., runway technology). Then in 1976, the Government could either easily switch to a two airport option or continue to develop and expand Zumpango. However, if strategy thirteen were chosen and demand didn't increase as predicted, etc., the Government might find by 1976 that it had a "white elephant" in that a new airport existed but was not needed or used. The political effects might be very bad and little flexibility would be available for "correcting" the situation.

The main purpose of these exercises was to indicate better what the overall impact of the 1971 decision might be. This was very important before beginning the evaluation process described next. It should be obvious that certain options in 1971 eliminate the possibility of other options in 1976, regardless of the events which occur in the interim.

8.8.4. First Evaluation of Nominal Alternatives

The sixteen alternatives, defined in Figure 8.6, were evaluated in a series of extensive discussions among the Directors of the Department of Airports and of the Center for Computation and Statistics, other staff members in SOP, and ourselves.

A preliminary evaluation indicated that seven of the sixteen alternatives could be discarded. Alternative 1 did not provide for maintaining the present service levels due to anticipated increases in demand. Alternatives 7, 8, 11, 12, 15, and 16 were undesirable because a high-level of commitment to Texcoco in 1971 would make it the major airport for the near future and remove the need for simultaneous construction at Zumpango. Finally, since the location of the new runway specified by the moderate Texcoco commitment would require new passenger facilities, there was not much difference between the options 3 and 4, so they were coalesced into a single alternative, which we label [3=4].

The next stage of the analysis involved having the members of SOP rank the remaining broadly defined alternatives on the attributes of flexibility, political effects, externalities and effectiveness, as described before. The particular rankings, which were reached by open discussions, represent the consensus judgment. When some alternatives were "indistinguishable" on a particular attribute, they were assigned the same ranking. For the political con-

Table 8.3: PRELIMINARY EVALUATION OF PLAUSIBLE GOVERNMENTAL OPTIONS FOR 1971 BY RANK ORDER

ALTERNATIVE	ATTRIBUTES										
	Flexibility	Political Effects on				Externalities due to					Effectiveness
		Presidencia	SOP	SCT	Overall	Prestige	Reg. Dev.	Bal. of Fed Exp.	Roads	Overall	
2	1	1	8	2	3	4	4	1	1	3	7
3=4*	7	4	5	1	4	1	4	6	3	7	8
5	2	3	6	4	3	3	3	2	1	1	3
6	3	2	7	3	2	3	3	3	1	3	1
9*	4	6	3	6	5	2	2	4	2	2	4
10	5	5	4	5	1	2	2	5	2	4	5
13*	6	8	1	8	7	1	1	6	4	5	2
14*	8	7	2	7	6	1	1	7	4	6	6

*Alternatives dominated by 2, 5, 6, or 10 on overall ranking of four major attributes siderations and externalities, the assessments on the components were first carried out, and then the overall ranking for these attributes was established. The ranking of the alternatives according to effectiveness was provided directly by results of the static model.

The results of the first ranking effort are shown in Table 8.3, where the smaller numbers represent the better rankings. From this table it can be seen that alternatives 3=4, 9, 13, and 14 are each dominated by others on the basis of their overall rankings for the four main measures of effectiveness. Alternative 6, for instance, is better than alternative 14 in terms of all four of the measures. Hence alternative 14 — and likewise alternatives 3=4, 9, and 13 — can be dropped from further consideration. The alternatives which were not dominated are those represented by the nominal cases 2, 5, 6, and 10. It is important to note here, however, that before we actually discarded dominated alternatives we engaged in a devil's advocate procedure: we tried to give the benefit of reasonable doubt to the impending noncontenders to see whether they could be resurrected to a place of contention. They could not.

8.8.5. Final Analysis of Dynamic Options

To refine the analysis of the possible governmental decisions, it was necessary to define the remaining contending alternatives more precisely. This was done as follows:

2 - At Zumpango, do no more than buy land for an airport. At Texcoco, extend the two main runways and the aircraft apron; construct freight and parking facilities, and a new control tower. Do not build any new passenger terminals.

5A - Build one jet runway, some terminal facilities and a minor access road connection at Zumpango. Buy enough land for a major international airport. At Texcoco, perform only routine maintenance and make safety improvements.

5B - Same as alternative 5A, except buy just enough land for the current Zumpango construction.

6 - Extend one runway at Texcoco and make other improvements enumerated in alternative 2. Buy land for a major international airport at Zumpango, and construct one runway with some passenger and access facilities.

10 - Same implications for Texcoco as alternative 6. Build two jet runways with major passenger facilities and access roads to Zumpango.

These five alternatives were ranked in the manner previously described. The results are given in Table 8.4. Proceeding as before, we can quickly see that alternative 6 dominates 10, and alternatives 2, 5A and 6 all dominate 5B. Thus the three remaining viable alternatives are 2, 5A, and 6.

Table 8.4: FINAL EVALUATION OF GOVERNMENTAL OPTIONS FOR 1971

ALTERNATIVE	ATTRIBUTES			
	Flexi-bility	Political Effects	Exter-nalities	Effec-tiveness
2	1	4	4	3
5A	2	3	3	2
5B*	4	5	5	4
6	3	1	1	1
10*	5	2	2	1

*Alternatives dominated by 2, 5A, or 6 on overall ranking of four major attributes.

The relative advantages of these three options were, finally, subjectively weighed by the SOP personnel as follows. Alternative 6 ranks better on effectiveness, externalities, and political considerations than either 2 or 5A. Although it is worse in terms of relative flexibility, it does allow the President to react effectively to all the critical events which might occur between 1971 and 1976, when the second stage of the airport decision could be made. Hence, in the opinion of the members of SOP working on this problem, alternative 6 was chosen as the best strategy.

8.9. IMPACT OF THE RESULTS

Based on the 1965-1967 study by SOP which recommended moving the International Airport to Zumpango as soon as possible, as well as indications early in this study, it was clear that some members of SOP held the opinion that a major move to Zumpango was still the most effective strategy. The static analysis, using SOP's own estimates and preferences, indicated a phased development involving a gradual shift toward Zumpango appeared equally as good. Once political considerations, flexibility of the policy, and externalities were accounted for along with effectiveness in the dynamic analysis of alternatives open to the government in 1971, it was evident that the "phased development at Zumpango" policy was better than an "all Zumpango" policy.

Looking at the implications of <u>their</u> evaluations, the SOP staff was very surprised and bewildered. Using their own preferences over measures of effectiveness they knew were relevant for a realistic set of options, they agreed that the two alternatives, thirteen and fourteen, which were most consistent with their so strongly held position, were completely dominated. Note also that the position of SCT, being most consistent with alternative 3=4 was also dominated.

This glaring inconsistency had a profound impact on many individuals within SOP. They rethought their position, analyzing in their own minds how this "strange" implication came about. As they understood the implication better, they gained some confidence in the result. With the final analysis of non-dominated alternatives and additional group discussions of the dynamic analysis, SOP adopted a new flexible position, exemplified by an initial choice of option six in 1971. Thus a very strange thing happened: an analysis undertaken for unabashedly advocacy purposes (i.e., to justify going all-out to Zumpango) turned out to convince the sponsors of the analysis that perhaps a more flexible stance was really in the best interest of Mexico.

8.9.1. The Ensuing Political Process

SOP recommended a "phased development" strategy to the President in December 1971. Specifically, it was suggested that land be acquired at Zumpango, that a major runway and modest terminal facilities be planned for construction during President Echeverria's term. It was also proposed that he reserve until 1976 a more detailed decision on how the airport facilities for Mexico City should be developed. This recommendation represented a major change in SOP's posture from the 1967 study. The previous recommendations of SOP were for master plans specifying what should be done at various points in time over the next thirty years without regard to the unfolding of relevant uncertain events. Thinking in terms of adaptive dynamic strategies rather than in terms of master plans played a pivotal role in our analysis.

As the last stage of our consulting activities, we, in collaboration with our clients, examined in some detail the steps that had to be taken in order to implement the newly developed stance of SOP. This required developing a strategy for the planning of technical documents, for informal presentations to key government agencies, for private meetings, and for possible announcements. Since we were not certain of the reactions of SCT and the Presidencia we mapped out some contingency plans which themselves were more in the spirit of an adaptive dynamic analysis than of a master plan. We are sure that you will understand that this chapter, however, is not the place to discuss the details of these politically sensitive considerations.

The analyses described in this chapter were completed in early September, 1971. In late 1971, Ing. Jauffred and

Ing. Dovali, together with Secretary Bracamontes of SOP presented the basic ideas of this study to the President of Mexico. Members of SCT and the Presidencia, including the respective secretaries of these ministries, were also present at this meeting.

The meeting, perhaps needless to say, did not eliminate all differences of opinion concerning the two basic points of view--remain at Texcoco or more to Zumpango--positions that had long been established. After the meeting, the President requested that SOP, SCT, and the Presidencia work out philosophical disagreements on the airport issue, as well as technical and financial details of further developing the airport facilities of the Mexico City area. Because of its complexity and importance, the process of 'working out the details' is very time consuming. By mid-1974, no concrete decision had been made. However, the winds seemed to blow a bit differently in 1974 than in 1971. In the earlier year, the basic issue was whether the main Mexico City Airport should be at Texcoco or Zumpango. In 1974, the issue seemed to involve when the Zumpango site would be the main airport--next year, in five years, or twenty. Support for this came from the fact that land for an airport at Zumpango was exappropriated by SOP, who holds this authority, in early 1974. Presumably, whatever decision evolves by the Government of Mexico will be done with greater awareness of the relative influence of the different attributes and of the dynamic issues.

ACKNOWLEDGEMENT

It was a pleasure for us, the authors, to work with F.J. Jauffred, Director of the Center for Computation and Statistics, and F. Dovali, Head of the Department of Airports, and their staffs in the Secretaria de Obras Publicas, and with Richard deNeufville of M.I.T., who added both engineering know-how as well as prodigious language skills to our consulting team. We felt that our colleagues in SOP contributed greatly on the integrity of the study by weighing seriously their answers to vexing value questions. Throughout they demonstrated a healthy skepticism at critical points, demanded meaningful explanations of the process of anlaysis, and were flexible enough to modify previously advocated positions.

DECISION ANALYSIS WITH MULTIPLE CONFLICTING OBJECTIVES PREFERENCES AND VALUE TRADEOFFS

(Chapters 9 & 10, & References)

Ralph L. Keeney and Howard Raiffa

May 1975 WP-75-53

CHAPTER 7

ILLUSTRATIVE APPLICATIONS

Certainly we are dealing in this book with a non-vacuous problem: many difficult, real-world decision problems do involve multiple objectives. Consequently, many of the concepts we have introduced are relevant and must be applied in either a formal or informal analysis of the alternatives. If one chooses to analyze multiple objectives and value or utility tradeoffs in a <u>formal</u> manner, then immodestly, we believe the ideas and procedures discussed in this book can often be of considerable use. The purpose of this chapter and Chapter 8 is to support this contention by illustrating many cases where multiattribute preferences have been formalized. The present chapter, in a variety of settings, focuses exclusively on the preference assessments themselves whereas Chapter 8, which concerns the site selection of an airport for Mexico City, presents a complete case including probability assessments, analysis of alternatives, interactions with the decision makers, and so on, as well as multiattribute preference assessments.

The applications discussed in this chapter cover the range of topics presented in Chapter 2 through 6. Section 7.1 discusses the generation of objectives and the specification of measures of effectiveness for an air-pollution problem. Section 7.2 discusses the allocation of resources for an educational program and the value functions of the members of a local school board and other local education officials are formalized. Next, a five-attribute utility function for response times of various fire trucks is assessed. This problem typically arises in planning operations of emergency services. Section 7.4 addresses the problem of structuring corporate preferences. In sections 7.5 and 7.6, we discuss preliminary work on the quantification of multiattribute preferences concerning decisions involving the selection of computer systems and decisions about the siting and licensing of nuclear power facilities.

The first six sections of this chapter relate in some depth experiences that we and others have had in assessing multiattribute preferences. The last section, 7.7, gives brief surveys of a number of other problems where formal analyses have explicitly considered multiple objectives using concepts discussed in earlier chapters. These include: utilization of frozen blood, sewage sludge disposal, safety of landing aircraft, choice of a job, shipments of hazardous materials, medical and surgical treatment of cleft lip and palate.

Our thesis is that the concepts and procedures introduced in this book are not just of theoretical, but also of operational interest and they can be--and have been--utilized to make contributions in a variety of important contexts. Many analysts are currently applying decision analysis to such crucial problems as those discussed in this chapter and the inventory of case studies is growing rapidly.

7.1 AIR POLLUTION CONTROL*

In New York City, the mayor must decide whether he should approve a proposed major addition to Consolidated Edison's electric power generating station in Astoria, Queens. If this addition is approved, City residents would be reasonably assured of receiving the growing quantity of electricity they will demand over the next several years at reasonable cost. However, approval would result in increased air pollution, particularly in terms of sulfur dioxide, particulates, and nitrogen oxides. Should this addition be approved?

In both Boston and New York City, the respective City Councils must decide whether to pass legislation that would place stringent limits on the sulfur content of fuels burned in the city. If passed, the legislation would lead to a definite improvement in the city's air quality--especially in terms of the air pollutant: sulfur dioxide.

---

*This section draws heavily on the dissertation of Ellis [1970] and adapts material from Ellis and Keeney [1972]. A related dissertation by Mead [1973] goes into more depth on the Astoria problem. Both dissertations were supervised in part by Raiffa.

However, passage of this legislation would require residents to incur added annual costs for heating and electricity to pay for the more expensive fuels with low sulfur contents. Should these City Councils pass such legislation?

In Washington, D.C., the U.S. Congress must decide whether to establish very stringent emission standards for carbon monoxide, hydrocarbons, and nitrogen oxides for all motor vehicles manufactured and sold in the United States. Establishment of these standards would contribute toward improving the air quality. On the other hand, they would require the public to pay significantly more money for new automobiles. Should Congress adopt these stringent standards?

Each of these decision problems is faced presently or has been faced recently by public officials. Moreover, they are representative of a host of similar problems that public officials increasingly confront. The basic question is: Should government adopt a specific, proposed program intended to improve the air quality? With each such investigation there is the additional question: "What should the air quality standard be?"

The major focus of this book has been to describe how a decision maker--in this case a public official--can utilize decision analysis to help make up his mind: how to select a desirable cause of action amongst the myriad of alternatives he confronts. In this section, we focus our attention on the selection of a set of objectives and measures of effectiveness for analyzing governmental programs designed to better control air pollution. We draw heavily upon the concepts discussed in Chapter 2.

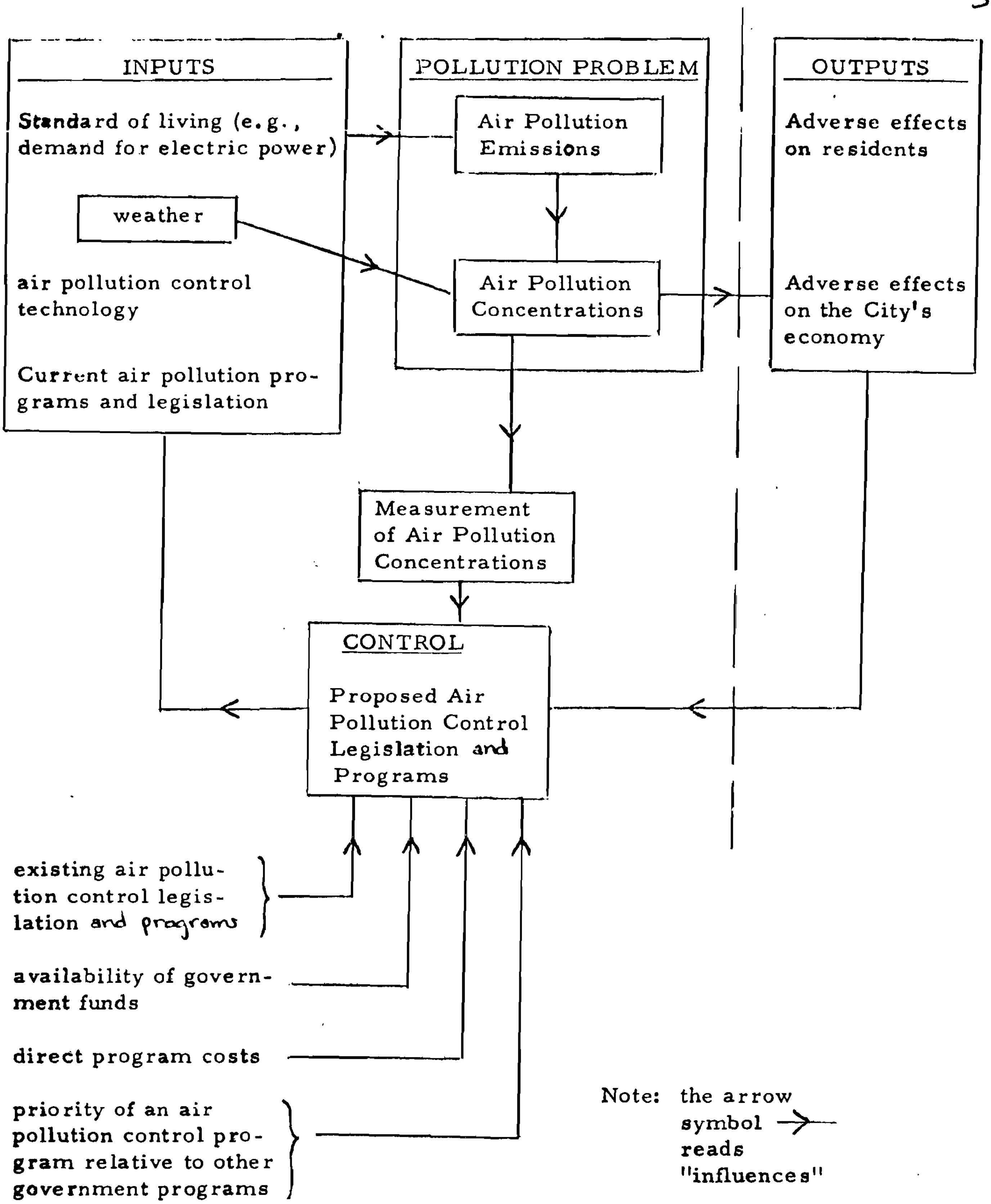
As a vehicle for illustrating or suggestions, focus is placed on one specific problem faced by one particular individual, the Mayor of New York City. Obviously, we would not expect the Mayor of New York to spend his time working on details of the air pollution problem. It would be reasonable, however, to expect members of the Mayor's staff in the Environmental Protection Administration and the Department of Air Resources to work on this problem. These individuals and the Mayor might then review the results and implications of such analyses in formulating and supporting air-pollution control programs for New York City.

In the next subsection, a brief overview of the air pollution control problem in New York City is presented, along with an introduction to the sulfur-dioxide problem. Then, objectives and measures of effectiveness are generated for the analysis of the problem. To avoid leaving the reader in midstream, the final subsection briefly sketches other aspects of this problem that were examined.

7.1.1 The Air Pollution Control Problem of New York City

A general model of the process by which many air pollution control programs are designed and evaluated is shown in Figure 7.1. The main problem with the control process as it is currently practiced in most municipal governments is that the outputs are usually not explicitly considered in choosing air pollution policy. The reason is, of course, understandable. There are simply too many complexities: the difficulties in defining appropriate output measures, in establishing the relationships between pollution concentrations and these measures, and in specifying preferences for the various possible outputs. But since action must be taken in most instances the feed-back loop goes directly from the measured air pollution concentrations to the control mechanism. In a sense, the process can be thought of as being short-circuited at the dashed line in Figure 7.1. Whenever this occurs the decision-making process excludes from formal analysis the most important information necessary for rational control. The suggestions here are meant to eliminate the short-circuit and include the outputs explicitly in the decision-making process. Of course, we do admit that good informal analysis often beats poor formal analysis. But our purpose here is to improve formal analysis.

<u>The Sulfur Decision Problem.</u> A survey of air pollution problems and current air pollution control programs in New York is given in Eisenbud [1970]. In 1970 a major decision still to be made in New York City's air pollution



A General Model for Evaluating Air Pollution Control Programs

Figure 7.1

control program concerned sulfur dioxide. Table 7.1 presents a breakdown of the estimated 1972 emissions of sulfur dioxide from sources within the City (as viewed from 1969). These estimates accounted for all provisions of existing laws enacted through mid-1971.

Table 7.1: Estimated 1972 Emissions of Sulfur Dioxide in New York City [NYC Department of Air Resources, 1969]

Source of Emissions	Emissions of Sulfur Dioxide (tons)	(per cent of total)
Incineration of refuse	2,500	(0.6%)
Motor vehicles	20,400	(5.1%)
Industrial processes	9,900	(2.5%)
Space heating	195,300	(49.2%)
Power generation	169,500	(42.6%)
TOTAL	397,600	(100.0%)

Since over ninety per cent of these emissions arise from the burning of fuels for space heating and power generation and since the only current, practical way to reduce emissions from these sources is to lower the sulfur content of the fuels burned, one important decision faced by the City was: "should the legal limit on the sulfur content of fuels burned in the City (then one per cent) be lowered?"

7.1.2 Identification of Major Objectives

In almost every decision problem faced by the Mayor of New York City, his most fundamental objective is to improve the well-being of his constituents. However, one must spell out in more detail what is meant by this objective as it pertains to air pollution. Precisely what would the Mayor like to accomplish by his actions concerning air pollution? After some serious thought, an evolutionary process led Ellis to divide the overall objective into five major objectives:

1. Decrease the adverse effects of air pollution on the health of New York City residents.
2. Decrease the adverse economic effects of air pollution on the residents.
3. Decrease the adverse effects of air pollution on psychological well-being of the residents.
4. Decrease the net costs of air pollution to the city government.
5. Achieve as desirable a political "solution" as possible.

These objectives require little justification. However, it should be noted that the second objective is meant to include costs of the air pollution control program in addition to costs of pollution itself. The net costs alluded to in the fourth objective include all the direct costs, such as the costs of an air pollution control program, as well as indirect costs such as those due to migration of businesses and industry from the city, less tourism, and tax revenue losses resulting from employee absences, due to sickness caused by air pollution.

Do these five objectives include all the issues of importance to the Mayor? For instance, nothing has been said about the overall consequences of the various alternatives on New York State, on the Federal government, on businesses, or on non-residents of New York City. Should these factors be included in a complete analysis of proposed air pollution control programs? Of course, the Mayor is concerned about these issues. However, note that some aspects of these consequences, such as economic effects due to tourism and businesses moving to the city, are included in the objective "decrease the net costs of air pollution to the city government." Benefits to non-residents from any air pollution program, for example, are probably highly correlated with the benefits to residents, and therefore in a first approximation could be ignored. All in all Ellis felt that explicit consideration of any of these additional objectives would not alter the optimal strategy, and therefore they were initially excluded from his list of objectives. However, after a preliminary analysis, he did reexamine these exclusions --albeit in an informal manner.

7.1.3 Assigning Attributes to Each Objective

The next task is to identify for each of the objectives suitable attributes that unambiguously indicate the degree to which the associated objective is achieved.

Health Effects on Residents. Several possible attributes immediately come to mind for the objective "decrease the adverse effects of air pollution on the health of residents." These include the annual number of deaths attributable to air pollution, the annual number of man-days of morbidity attributable to air pollution, and some subjectively assessed health index that includes consideration of both morbidity and mortality.

Important objections can be raised against each of these. The annual number of deaths attributable to air pollution is not comprehensive in that it does not account at all for what is believed to be the more prevalent effect of air pollution on health--namely, its effect on morbidity. Similarly, the annual number of man-days of morbidity does not account at all for the extremely serious effect of air pollution on health in terms of mortality.

Thus, it seems clear that no single measure of effectiveness, aside from possibly a subjective health index, can be identified for this objective. However, because such an index lacks a physical interpretation, it is not particularly desirable in terms of the measurability criterion discussed in Chapter 2. Hence, the alternative of specifying the major objective in more detail was chosen. Health considerations were divided into two detailed objectives, "decrease mortality" and "decrease morbidity."

For the first of these, two of the possible measures of effectiveness are the "annual number of deaths attributable to air pollution" and the inversely oriented scale, "per capita increase in the number of days of remaining lifetime due to improved air quality." The first equally weighs the death of an old person and the death of a child, whereas the second measure weighs the death of a young person more heavily. The latter measure was chosen since it was felt that in this case, it more adequately describes the impact of a program alternative with respect to "decrease mortality."

For the objective "decrease morbidity," the "per capita decrease in the number of days of bed disability per year due to improved air quality" was chosen as the attribute. Obviously, this does not include such effects as sore eyes which would not force one to a bed. Part of the consequence of sore eyes is psychological, which can be accounted for by the third major objective. However, the physical aspects of sore eyes intuitively seem important enough to be formally included in the analysis. To do this we would suggest calibrating a number of days of bed disability per year which one would feel is equivalent to having sore eyes of different levels of severity during the year. Then for each program alternative, the effects due to sore eyes would be included in the analysis by adding an "equivalent number of bed-days disability" to our measure of the degree to which "decrease morbidity" is met.

Economic Effects on Residents. No single attribute could be identified for the objective "decrease the adverse economic effects of air pollution on residents of New York City," because the Mayor would want to consider the economic impact on residents at various income-levels. As a compromise Ellis chose to consider a dichotomy: the economic effects on low-income and on other residents. Per capita annual net cost to residents was used as the measure of effectiveness for each group.

Psychological Effects on Residents. There seems to be no direct measure of effectiveness for the objective "decrease the adverse effects of air pollution on the psychological well-being of the residents." One could, however, define some subjective index and perhaps interview residents about their feelings for various levels of air quality. But Ellis chose a simpler approach, which used the daily concentration of sulfur dioxide as a proxy attribute for "psychological well-being."* Since this pollutant can easily be detected both visually and by

*It is important to emphasize that this concentration level is to be viewed as a proxy for psychological well-being only and not for the other objectives.

breathing, it seems reasonable to assume "psychological well-being" is closely related to the concentration levels.

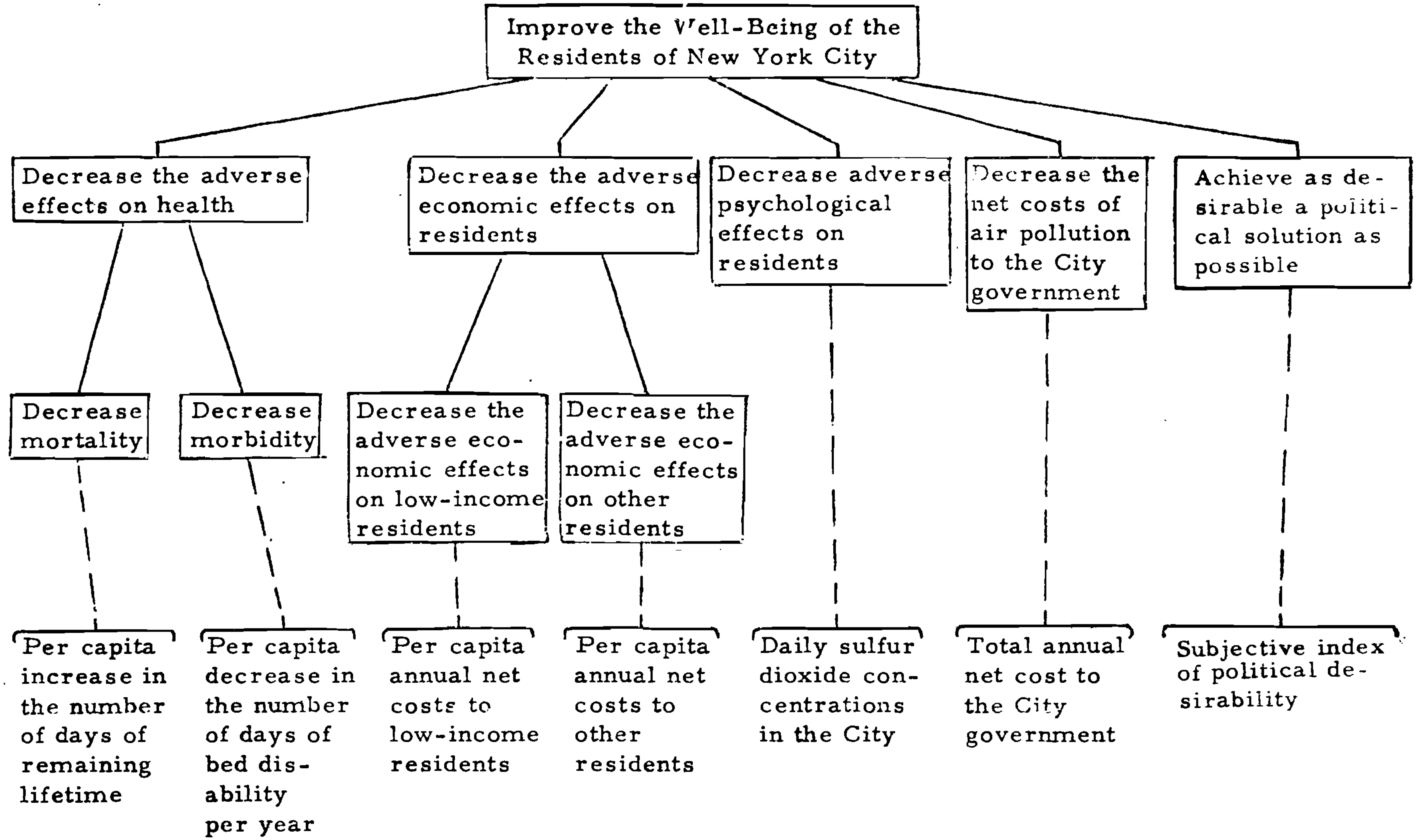
Economic Effects to the City. As a measure of effectiveness for the fourth objective, "decrease the net costs of air pollution to the City government" Ellis used "annual net costs." As mentioned previously, this includes both direct and indirect costs.

Political Implications. The fifth objective, "achieve the best political solution to the air pollution problem," has no nice objective measure of effectiveness and a subjective index was used. Many considerations must be included in measuring the index, such as the possibility of court suits brought by landlords or home-owners who are forced to pay higher fuel prices for heating, the Mayor's relations with the City Council and with Con Edison and with any of the political groups in the city, and the support of the general public for various program alternatives. All of these have a potential effect on the Mayor's political future which also should be taken into account.

7.1.4 The Final Set of Objectives and Attributes

Figure 7.2 exhibits the hierarchy of objectives and their associated measures of effectiveness used by Ellis in his study of air pollution control in New York City.

Of course, there may be important objectives which



The Complete Set of Objectives and Attributes
for Choosing an Air Pollution Control Program
for New York City

Figure 7.2

Ellis did not think about that are consequently not included in his analysis. However, if one cannot identify such omissions before utilizing the implications of the analysis, the same omissions might have occurred if any less formal procedure for guiding the decision making process were followed. And in this case we would be no worse off using formal analysis than not. Admittedly, with an informal analysis one might think intuitively or subconsciously about objectives that one might not be able to articulate. And also admittedly, a formal analysis may inhibit this mysteriously creative, gestalt way of thinking. But on the other hand, this type of unstructured introspective analysis is so very private that others cannot share in the process and suggest additions or modifications they deem appropriate.

7.1.5 Decision Analysis of the Sulfur-Dioxide Decision Problem

Since the purpose of this section was to develop objectives and attributes for the sulfur-dioxide decision problem, the ensuing analysis will only be briefly mentioned. The interested reader may refer to Ellis [1970] for details of the assessments or to Ellis and Keeney [1972] for an overview.

Since Ellis' work was done as a doctoral thesis designed to illustrate the methodology, only two alternatives were explicitly evaluated. These were the status quo, which entailed maintaining a one percent legal limit on the sulfur content of oil and coal used in New York City, and an alternative which lowered the legal limit to 0.37 percent for oil and 0.7 percent for coal. To analyze the full range of alternatives would require a team of researchers rather than one individual.

The alternatives were evaluated in terms of the seven attributes defined as follows:

$X_1 \equiv$ per capita increase in the number of days of remaining lifetime,

$X_2 \equiv$ per capita decrease in the number of days of bed disability per year,

$X_3 \equiv$ per capita annual net costs to low-income residents,

$X_4 \equiv$ per capita annual net costs to other residents,

$X_5 \equiv$ daily sulfur-dioxide concentrations in parts per million,

$X_6 \equiv$ total annual net cost to the City government,

$X_7 \equiv$ subjective index of political desirability.

Joint probability functions describing the possible impact of the two alternatives were assessed exploiting probabilistic independence, conditional probability assessments, and a small simulation model. Exploiting some utility independence assumptions which were deemed to be appropriate on the basis of discussions with staff members in the Department of Air Resources, a seven-attribute utility function felt to parameterize the Mayor of New York's preferences was structured.

It is interesting to note that Ellis' did not conclude that the Mayor of New York would view each attribute as utility independent of its complement. The main reason for this was the feeling that the Mayor would likely be more risk averse in terms of attributes X_1, X_2, X_4, and X_5 if the political effects were at an undesirable level than he would given desirable political effects. From his interaction with the Department of Air Resources, Ellis did conclude that for the ranges of the possible consequences, the attributes X_3, X_6, and X_7 were each individually utility independent of their respective complement. Also, he felt that given any fixed level of attributes X_7, the attributes X_1, X_2, X_4, and X_5 would each be conditionally utility independent of the remaining attributes. With these assumptions, the assessment of the complete utility function required (1) assessing seven one-attribute utility functions, one over each effectiveness measure, and (2) assessing eighteen scaling constants to insure the seven utility functions were properly scaled. No assessments of the utility function were completed, although details about the functional form of the utility function and the reasonableness of the utility independence are given in Ellis [1970]. Appropriate techniques for performing each of the necessary assessments are found in earlier chapters of this book.

7.1.6 Impact of This Work

The ideas and results expressed in this section may have had some influence on the thinking of individuals responsible for air pollution control programs in New York City. Although no claim can be made concerning causality, the following events have occurred:

The results of this work, concerning the range of possible effects of a program which lowered the legal limits of oil and coal used in the city from the present one percent to 0.37 and 0.7 percent respectively, were made available to the New York City Environmental Protection Administration, which was in the process of preparing a new air pollution control code for the City. This group included, as one of the key provisions in its recommended code to the City Council, a program which was essentially the same program as the one Ellis analyzed.

These same results, as well as the methods of analysis upon which these results are based, were presented by Howard M. Ellis in testimony before the New York City Council in its legislative hearings on the proposed new air pollution control code. The code was approved by the City Council and became law in 1971. Ellis continued to consult with the City after his thesis was completed.

The present authors suspect that, as is the case with many analyses of this type, the detailed quantitative work involved in doing the full-scale study probably helped the investigator to better understand the qualitative implications of the problem, and it was this qualitative understanding which helped him influence the governmental officials. Perhaps this level of sophistication could have come about through other means, but one should not underestimate the important intellectual and emotional impact that arises when one is forced to express vexing tradeoffs in unambiguous quantitative terms. It forces one to think harder than one is ordinarily accustomed to...especially if one then has to defend his assessments in front of other experts.

7.2 PREFERENCE TRADEOFFS AMONG INSTRUCTIONAL PROGRAMS

Roche* considers the problem faced by a decision maker who has to choose among alternate budget allocations to diverse activities which compete for the same scarce resource. He is concerned about the role played by the decision maker caught "in the middle." That is a decision maker who is in the position where he must, on one hand, obtain funds from some approving authority and, on the other hand, approve the budgets for programs directed by professionals in his employ. With a constrained budget he can increase the budget of one program only at the

---

*In this section, we summarize and review the work of Roche [1971]. His doctoral thesis, which was supervised by Raiffa, makes extensive use of the material in Chapter 3 on tradeoffs under certainty.

expense of other programs. He must take from Peter to pay Paul and do it in such a manner to convince his overseers of the reasonableness of it all. Roche was motivated to see if formal preference analysis of the type we are disucssing in this book could help such a man-in-the-middle both to crystallize his own tradeoffs and to communicate this process to the body that controls the dispensation of funds.

Roche chose to study the budget-allocation problem in the context of a small school district. The school superintendent was the decision maker "in the middle," Roche's principal client; the people below the superintendent were the school principal and the coordinators of various educational programs; the people above the superintendent were the school board which acted as the funding agency for the town. School boards in New England have a great deal of fiscal autonomy and can impose financial obligations on the town. But, of course, these school-board members are themselves elected officials so that the ultimate responsibility does reside in the collectivity of town citizens.

Roche was indeed fortunate--but it was far from all luck--to find a chairman of a school board and a superintendent who were initially interested in pursuing a pilot test of Roche's ideas. It is a credit to Roche that the initial curiosity of these cooperating individuals bloomed into full-scale enthusiastic cooperation and, as we shall see, he was skillful enough in his personal relations to involve other individuals in the measurement exercise. Roche, in his thesis, disguises the name of the town, which he fictiously calls "Somerstown," and he disguises as well the names of the characters that participated in the exercise. However, we assure you that many of the dialogues recorded in the thesis are verbatim reports of actual measurement sessions.

7.2.1 Refining the Problem

Somerstown began a program budgeting effort in September 1969, a couple of years before Roche entered the scene. One of the school board members was a business school professor, and it was through his intervention that the superintendent recast the traditional line-item budget into a program format. At the junior high-school level, the basic program format was segregated according to subject matter. The superintendent and the business school professor alluded to above, admitted however, that to their disappointment the program budgeting effort had practically no effect whatsoever on the reallocation of funds to different school-subject programs. Each year, the funds were allocated like the year before except perhaps for a uniform percentage increase. This background may partially explain the receptive audience that Roche received when he approached Somerstown authorities with the idea of examining fundamental tradeoffs among the funding of different subjects. We also point out, in the way of background material, that Somerstown is a small homogeneous community whose educational program was deemed comparatively stable and free of the many frictions that plagued other educational systems at that tumultuous time.

Roche concentrated on the allocation process for four subject programs in the junior-high program:

i. English/Language Arts

ii. Science

iii. Mathematics

iv. Social Studies

The Somerstown Schools have a coordinator for each of these programs and the coordinators prepare an annual budget for their respective domains of responsibility. Each feels a responsibility to do better each year than the year before and each tries to get increases in funding for his or her program--the usual advocacy procedure. When Mr. A asks for an increase, he seldom feels obliged --nor would it be considered good form--to argue that the money he is seeking should come from Mr. B's program. It is the task of the superintendent to juggle these requests and to suggest a compromise among them in a fashion that maintain the loyalty of his staff and at the same time gains the acceptance of the school committee.

The first half of the thesis is concerned with the creation of a suitable production function: the transformation of financial and personnel inputs to educational outputs--no mean task! For a long time before Roche started his probing, the Somerstown school authorities worried about educational indices. Several indicators could be chosen but many are highly correlated and for convenience of the exercise, Roche and his collaborators chose for each of the four subjects the index, "Percentage of students achieving at or above grade level on the standardized achievement test."

In a later chapter of his thesis Roche does discuss the inadequacy of this output measure. He defends his use of it, however, on pragmatic grounds and he does discuss what other researchers might do if they were to choose other output indices. We feel that the chosen index is far from a good surrogate for educational performance and we feel that it is not an elementary task to suggest how Roche's analysis could proceed using a more sensitive set of output indices. But for the time being we are stuck with the index used and let us get on with the story even though it is marred by the exclusive use of this over-simplified output index.

7.2.2 Relating Program Costs to Output

Let us look at the process Roche followed in confronting the science coordinator. The science budget for the existing year was $81,000 and 59% of the students performed at, or better than their specified grade level. Roche first inquired about the effects of dropping the science program altogether. The coordinator did concede that many of the students would continue to perform at or above the hurdle level. He then inquired about the effect of a 10% increase, (i.e., an increase of $8,100). "What would I be allowed to do with the money?", Dave Flaherty queried.

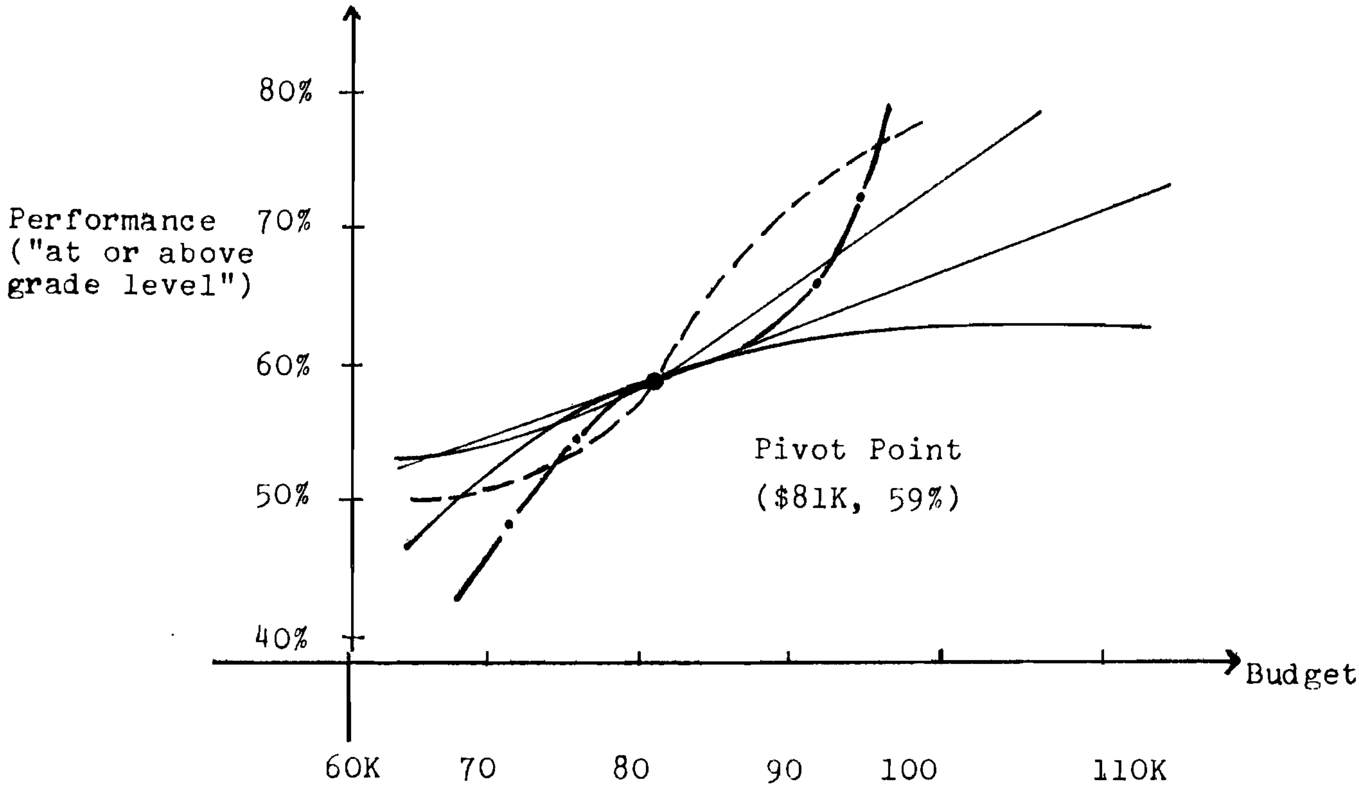
"It's up to you," responded Roche. "The essential point is, Dave, that none of us knows how to use an additional $8,100 in science better than you do. Once you decide what you would do with the $8,100 I will ask you to assess what impact those additional funds would have on the students in the same way we did before. That is, we shall ask in turn: What would you do with the increased funds? Which levels or sections in which grades would be effected? What would you expect the effect would be along the dimension of number of students achieving at or above grade level in science?"

Roche coached Dave Flaherty to think hard about the questions posed. He encouraged the science coordinator to steep himself in the past data, to think about the increased money not in the abstract but in terms of what it would buy in the form of additional teaching help or additional audio-visual facilities, and so on, and to think about the effect on individual students. He posed such questions as: "If you do so-and-so, would this really help Mary Jane over the hurdle?"

The production function ideally should have been probabilistically assessed but all Roche had the courage and time to do was to elicit in each case a median value, i.e. a value for which the assessor thought the true value would be equally likely to fall above or below the estimated value. He formalized the assessment procedure in terms of a written protocol with several pages of work sheets that the coordinator took many hours over a period of days to answer.

The end product that Roche sought from the science coordinator was a curve that plotted estimated performance (% at or above grade level) on the vertical axis against budgetary values on the horizontal axis. This curve, the assessed production function, was meant to go through a pivot point at the status quo level--i.e., a budget of $81,000 produces a performance of 59%.

After Flaherty completed the work sheets prepared by Roche, he was presented with the following task: "Now that you (Flaherty) have completed the assessment questions, we would like to probe your qualitative judgement about the possible shapes of a performance function for the Somerstown Junior High Science Program." Roche then showed Flaherty several shaped curves as shown in Figure 7.3 and they discussed the qualitative meaning of each. After Flaherty seemed to understand the implications of each shape he was asked to select one of the shapes presented or to invent a shape that reflected his true feelings.



Qualitative Shapes of Performance Functions

Figure 7.3

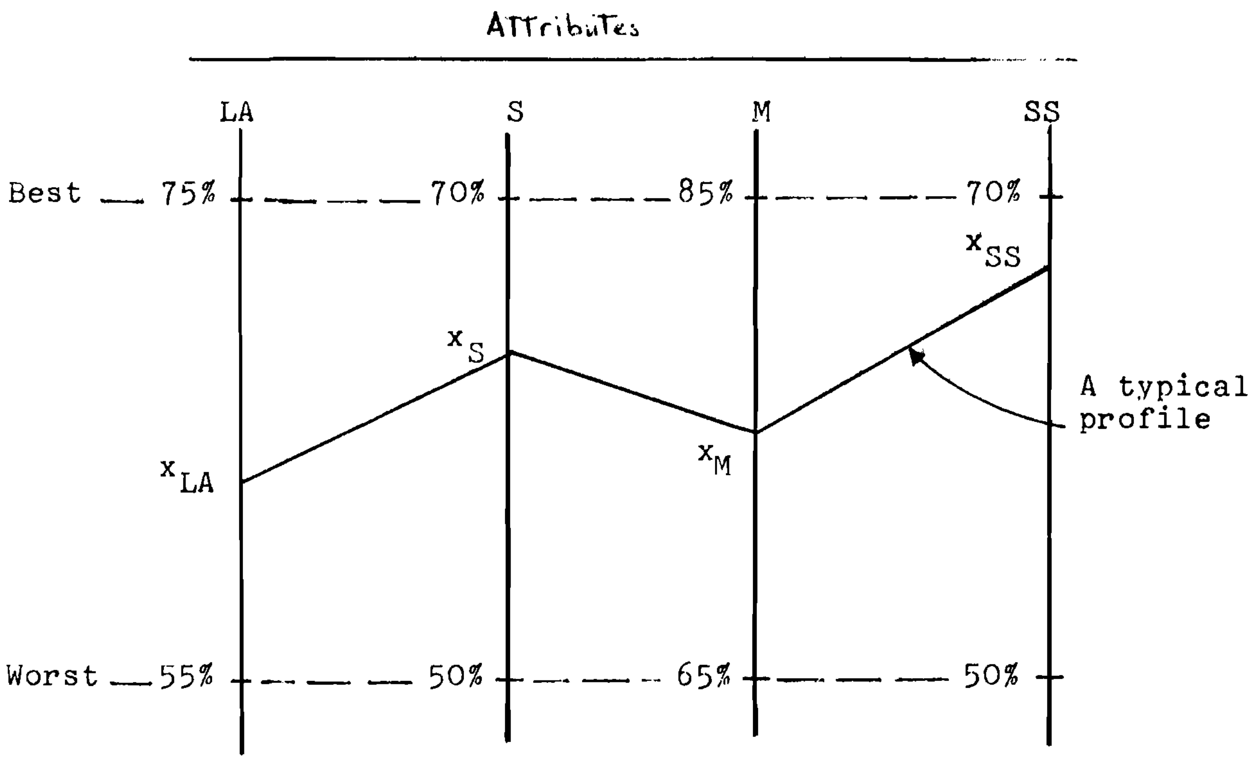
In a gentle manner Roche discussed with Flaherty some of his responses and indicated some inconsistencies amongst the answers he recorded--but he did this corrective procedure with the supportive advice that Flaherty should not be embarrassed at these inconsistencies, since anyone put into his position would be equally inconsistent. The important thing was to have Flaherty reflect and ponder about these inconsistencies and then try to modify some of his earlier assessments so that the revised set of responses would be internally consistent. And what is perhaps more important, the revised answers should be felt to accurately portray the current best assessments Flaherty could make in light of his new level of understanding.

All we can hope to do here in this summary is to give the reader a flavor of the care that Roche took to generate a performance function from each of the four coordinators. The superintendent, Dr. Nelson, had his own views about these performance assessments and felt compelled in some circumstances to modify the assessments of his subordinates. Dr. Nelson remarked, however, that if this assessment process were to be repeated year after year then he would be able calibrate his coordinators on the basis of a track record. The school committee, which monitored the entire exercise felt that it was most appropriate for the superintendent to modify these performance functions in collaboration with his coordinators, since the school committee superintendent had to take full responsibility for the finally recorded performance functions. The committee explicitly stated that their deliberations would be based primarily on the superintendent's own assessments, which, in turn, would be based in part on the inputs he received from his coordinators.

7.2.3 Assessing a Value Function

Now let us turn to the second part of the thesis dealing with preference structures. Roche investigated the preference structures of several concerned indivuduals for different performance profiles. A typical profile is a four-tuple $(x_{LA}, x_S, x_M, x_{SS})$ which refers to performance scores on language-arts, science, mathematics, and social sciences respectively, and where x_{LA}, for example, represents the percentage of students at or above grade level in language/arts.

As is evident in Figure 7.4 each of the performance ranges was restricted to a subinterval of the theoretically feasible range from 0% to 100%. For example, mathematics performance was restricted from the worst case of 65% to the best case of 85%. These restricted ranges were ample to accommodate budgetary changes that could realistically be recommended. It was critical to restrict these ranges so that one could adopt various preferential independence assumptions. We shall expand



Performance Ranges and a Typical Profile

Figure 7.4

on this point shortly.

Due to the considerable support Roche received from Dr. Nelson, the superintendent, and Mrs. Humphrey the chair-woman of the school committee, Roche was able to field test preference assessments with every single administrator and policy maker involved in the decision-making process of the junior high school. These involved the principal and assistant principal of the junior high school, the superintendent and assistant superintendent, and all five members of the Somerstown school committee. Absent from this listing are the citizens and the parents of school children. In addition, the preference procedure was also field tested on a group of 18 doctoral students in educational administration.

It was surprisingly easy to verify the reasonableness of pairwise preferential independence. For example, Roche set x_M and x_{SS} at low levels of 70% and 55% respectively, and then probed conditional preference tradeoffs between X_{LA} and X_S. After he thoroughly engaged his subjects in this problem he asked parenthetically whether any of the tradeoff responses between X_{LA} and X_S would be altered if x_M and x_{SS} were not set at 70% and 55% respectively. Practically all of his subjects felt that these tradeoffs would certainly not be influenced by such modifications of the fixed levels of x_M and x_{SS}. Some subjects, including the superintendent emphasized the point that the tradeoffs would not depend on the fixed levels of X_M and

X_{SS} provided that these levels were within the specified bounds. He felt for example, that if x_M were set at 30% this would be such a shock to the system that his trade-offs between X_{LA} and X_S would be affected.

For all subjects, Roche felt that the necessary pairwise preferential independence assumptions were satisfied to legitimatize adopting a value function of the form

$$v(x_{LA}, x_S, x_M, x_{SS}) = k_{LA}v_{LA}(x_{LA}) + k_Sv_S(x_S) + k_Mv_M(x_M) + k_{SS}v_{SS}(x_{SS}) \quad (7.1)$$

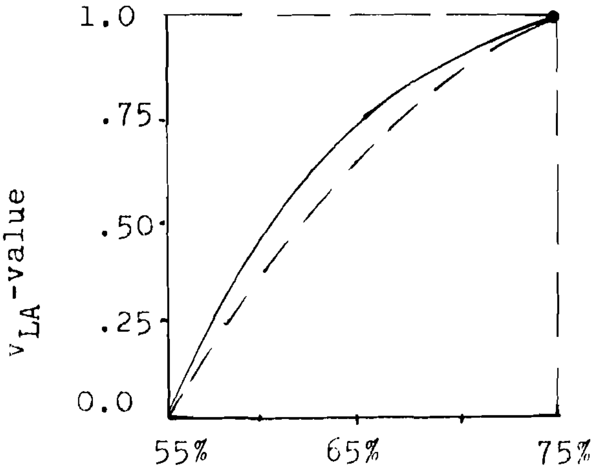
where the component v's were normalized respectively at 0 and 1 for the worst and best alternatives (e.g. $v_{LA}(55) = 0, v_{LA}(75) = 1$, etc.), where the k's were non-negative, and where

$$k_{LA} + k_S + k_M + k_{SS} = 1 \quad . \quad (7.2)$$

Roche followed the assessment procedure described in Section 3.7. He assessed for each subject the component value functions by the mid-value technique: for each component function he first found the .50-value point, next the .25 and the .75 points, then he checked the .50-point against the .25 and .75 points, and finally he discussed the general shape of the v-component functions. Next he sought the k-weights. He asked such questions as: "Suppose we consider a disastrous profile such as (55, 50, 65, 50) where all performance measures are at their worst levels. Now suppose you could push one of these worst scores up from the worst level to the best, which would you choose? Would you prefer to push language/arts up from 55 to 75, or science from 50 to 70, or mathematics from 65 to 85, or social science from 50 to 70?" He thus probed each respondent for rankings of the k's. Next, he followed the technique discussed in Section 3.7 and determined precise numerical values for the k-weights. Figure 7.5 depicts the assessments of Superintendent Nelson and his assistant, Mr. Elliot. Table 7.2 summarizes some salient data collected from the nine principal actors involved in the exercise. Roche not only obtained Nelson's assessments but he had Nelson guess at what some of his associates would record. It's fascinating to read how Nelson rationalized some of the recorded assessments of members of his staff and the school committee members. There are striking differences of opinion!

As regards the 18 students in the doctoral seminar in educational administration, all of whom were subjected to the same assessment procedure, we quote from Roche:

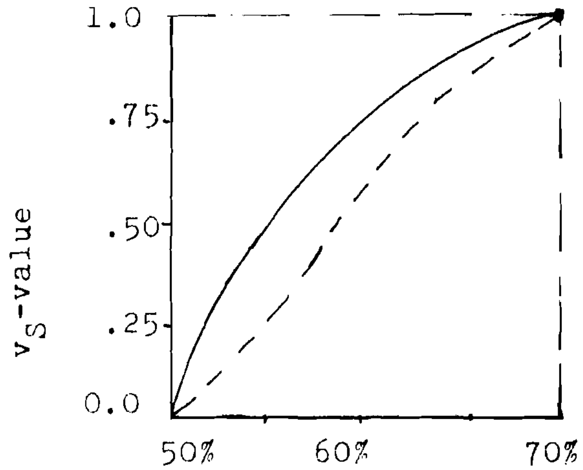
> "There is little to be gained at this point in the study from exhibiting the eighteen structures. However, the following summary information might be of interest.
>
> 1. With respect to the Language Arts program, 11 of the curves were concave, 6 were linear, and 1 was S-shaped about the current



Language Arts Performance

Nelson: k_{LA} = .30

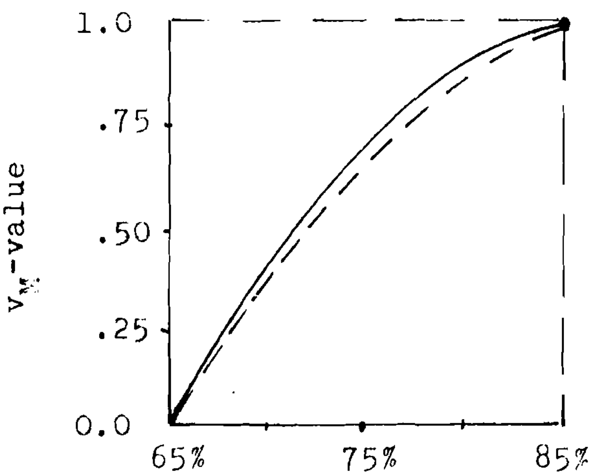
Elliot: k_{LA} = .33



Science Performance

Nelson: k_S = .21

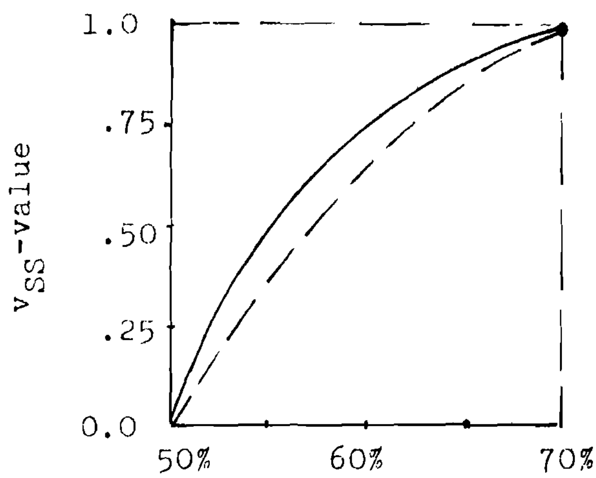
Elliot: k_S = .20



Mathematics Performance

Nelson: k_M = .26

Elliot: k_M = .27



Social Studies Performance

Nelson: k_{SS} = .23

Elliot: k_{SS} = .20

Performance ranges use the evaluator: "percentage of students at or above grade level."

Dr. Nelson, Superintendent: ————

Mr. Elliot, Assistant Super.: ---------

Figure 7.5

Value Functions Assessed by The Superintendent and Assistant Superintendent of The Somerstown School System

TABLE 7.2

Assessed k-Values and .50 Mid-Value Points of Principal Subjects

Name	k-Values				(.50)Mid-Value Points			
	LA	S	M	SS	LA	S	M	SS
Administration:								
(1) Mrs. Carter (Principal)	.20	.25	.22	.33	60.5	54.5	70	54
(2) Mrs. MacGregor (Asst. Principal)	.21	.24	.23	.32	61	54.5	68	53.5
(3) Dr. Nelson (Superintendent)	.30	.21	.26	.23	60.5	55	71.5	55
(4) Mr. Elliot (Asst. Superintendent)	.33	.20	.27	.20	62	59	72	57
School Committee:								
(1) Mrs. Humphrey (Chairwoman)	.36	.13	.30	.21	62	63	69	57.5
(2) Mrs. Clark	.22	.26	.23	.29	65	59	67.5	57
(3) Mr. Cowles	.53	.10	.27	.10	65	62	70	63
(4) Mrs. Oscar	.47	.11	.35	.07	65	62	69	60
(5) Mr. MacMillan	.29	.23	.28	.20	59	55	68	54.5
lowest mid-value point assessed:					59	54.5	67.5	53.5
highest mid-value point assessed:					65	63	72	63

Key: Each row contains (1) the scale factor for each program, which indicates the subject's tradeoffs among programs; and (2) the "global" mid-value point for each program, which gives an indication of the subject's tradeoffs within a program. A low mid-value point indicates a strong aversion to poor performance.

performance level.

2. In the Science program, 8 were concave, 2 were linear, 4 were S-shaped about current performance, and 4 were convex. Thus, there was much less concern with poor performance in Science than in Language Arts.
3. Interestingly enough, all 18 of the curves were concave in the Mathematics program. That is, there was unanimous concern with poor performance in Mathematics.
4. In the Social Studies program, 11 of the curves were concave, 3 were linear, 2 were S-shaped about current performance, and 2 were convex.

It is of interest to note that the doctoral students, like the subjects in Somerstown, basically fell into two groups: (1) the "educators" who were essentially concave in all programs (eight of the students fell into this group); and, (2) the "policy makers" who were either concave or linear in Language Arts, concave in Mathematics, and S-shaped or convex in either/or Science and Social Studies (eight of the students fell into this group). Only two of the students did not fall into either of these groups. This was because these two students were S-shaped about current performance in Language Arts. It may be coincidental, but one of the students whose structure very closely approximated the typical school committee member's structure in Somerstown, had just recently run for election to the Boston School Committee.

Of even more interest to the analyst was the fact that no student was linear in all the programs. Therefore, without knowing it, the students demonstrated that the typical "priority list" approach, i.e., the constant linear form, would be inappropriate for analyses of their preferences among programs. When this evidence is added to the data generated in Somerstown, it suggests that the analyst should be extremely careful about using the constant linear form.

With respect to the determination of scale factors during the second part of the assessment procedure, the vast majority of the students behaved as did the Somerstown superintendent and a majority of the Somerstown School Committee. That is, 15 out of the 18 students chose Language Arts as that program they would want to "push-up" first. Science was picked by 2 students, and one chose the Social Studies program. Although none of the students picked Mathematics as the "base" program, 9 of them chose this program as the second program they would like to see "pushed-up." The remaining 9 students all chose Social Studies as the second program."

After Roche obtained the full assessments from his subjects he asked each of four of the School Board members plus the assistant superintendent to suggest budgetary alternatives that would either be most appealing to themselves and would have some chance of being accepted by the group or be of a type that they would welcome seeing evaluated. Five alternatives besides the no-change position were thus generated. Again we quote from Roche:

"The "no-change" alternative for the Junior High School Core Program was as follows: allocate $92,000 to the Language Arts program, $81,000 to the Science program, $76,000 to the Mathematics program, and $75,000 to the Social Studies program. The alternative allocations (expressed as changed to the "no-change" case) are listed below with the names of the individuals who suggested them.

1. The Humphrey alternative. Take $6,000 from Science, and $6,000 from Social Studies. Increase Language Arts by $10,000 and increase Mathematics by $2,000.
2. The Oscar alternative. Take $7,000 from Science, and $2,000 from Social Studies. Increase Language Arts by $6,000, and increase Mathematics by $3,000.
3. The Elliot alternative. Take $7,000 from Science, and $1,000 from Social Studies. Increase Language Arts by $3,000, and

increase Mathematics by $5,000.

4. The Cowles alternative. Take $3,000 from Language Arts, and $6,000 from Mathematics. Increase Science by $4,000, and increase Social Studies by $5,000.

5. The Clark alternative. Take $2,000 from Language Arts, $2,000 from Science, and $1,000 from Mathematics. Apply all $5,000 to Social Studies."

Using the performance functions as generated by the program coordinators and modified by Superintendent Nelson, and using the preferences of each of the four administrators and five Board members, it was possible to evaluate the six suggested proposals. These preferences are shown in Table 7.3 and comparative rankings are shown in Table 7.4. We can see readily that the Humphrey alternative strictly dominates the "no-change" alternative and the Elliot alternative. Furthermore when power realities are also considered, the Humphrey alternative essentially overpowers the Clark and Oscar alternatives as well. This leaves a contest between the Humphrey and the Cowles proposals. Again, however, looking at the personalities and the strengths of preferences one would be tempted to single out the Humphrey proposal as the obvious winner.

Roche raises the question whether the above described procedure could seriously be implemented for group decision making. He writes:

TABLE 7.3

Preference of Each Individual for Each Budget-Alternative
Generated by the Educational Value Function

		Alternative Budget Allocations					
		"No change"	Humphrey	Oscar	Elliot	Cowles	Clark
Administrators and Board Members	Dr. Nelson	.730	.743	.737	.743	.730	.727
	Mr. Elliot	.642	.650	.643	.646	.646	.637
	Mr. Carter	.771	.778	.778	.777	.793	.785
	Mr. MacGregor	.765	.771	.771	.771	.784	.775
	Mrs. Humphrey	.667	.697	.686	.686	.668	.667
	Mrs. Clark	.638	.647	.629	.628	.676	.632
	Mr. Cowles	.584	.647	.624	.608	.563	.570
	Mrs. Oscar	.608	.647	.650	.631	.588	.597
	Mr. MacMillan	.813	.816	.807	.809	.816	.809

Key: In each row, the preference or value number as determined by each subject's preference structure is presented for each alternative budget allocation.

TABLE 7.4

Rankings of the Alternatives for Each Individual
Developed from the Preference Numbers in Table 7.4

		Alternative Budget Allocations					
		"No change"	Humphrey	Oscar	Elliot	Cowles	Clark
Administrators and Board Members	Dr. Nelson	3	1	2	1	3	4
	Mr. Elliot	4	1	3	2	2	5
	Mr. Carter	5	3	3	4	1	2
	Mr. MacGregor	4	3	3	3	1	2
	Mrs. Humphrey	4	1	2	2	3	4
	Mrs. Clark	3	2	5	6	1	4
	Mr. Cowles	4	1	2	3	6	5
	Mrs. Oscar	4	2	1	3	6	5
	Mr. MacMillan	2	1	4	3	1	3

"Under normal conditions, I don't believe it would be reasonable to expect that policy makers would allow their own preference structures to be communicated. Recall that Dr. Nelson said that he would usually want to keep his own preference structure private. The administrators and policy makers in Somerstown are rather unusual. They willingly cooperated in this effort in order to further research on decision making. Additionally, there are no major educational problems in Somerstown. That is, there are no sensitive issues at stake. Therefore, no individual felt threatened by having his or her preference structure recorded. In such a case, decisions would be of the fine tuning variety, rather than the sensitive policy decisions."

Roche developed a computer program that takes the performance functions and the preference structure of a single decision maker--he used Nelson's as an example--and generates the optimum allocation for a given overall budget level. It is essentially a resource-allocation type of dynamic program. Given this program it is easy to generate the program implications of various overall budget levels. Roche however, did not choose to formalize tradeoffs between money and the four indices of scholastic performance. If he had chosen to do this, undoubtedly the set of four scholastic attributes would have been preferential independent of the monetary attribute so that all of Roche's work would also be relevant and appropriate in the extended framework. The computer program also makes it relatively easy to investigate various sensitivity studies: for example, dependence on the k-weights or on changes in performance functions.

We conclude this section with a quote from Roche:

> "Although this research demonstrates that these new techniques could be used to examine budgetary alternatives among programs, the demonstration was within a very narrow context. There may be problems in attempting to use these formal techniques elsewhere. The local educational setting served as a 'laboratory' for the investigation of these techniques. I believe that this setting is representative of numerous non-profit organizations. However, on the basis of this research we cannot say that these formal techniques should be used <u>everywhere</u>, but, rather, that they could be used <u>somewhere</u>."

7.3 FIRE DEPARTMENT OPERATIONS[+]

In any analysis of fire-department policy a classical question is: "How much is a minute of response time[‡] worth?" Clearly the value for any particular fire depends on the detailed circumstances of that fire. It is not feasible to treat individually each of the several thousand serious fires which the New York Fire Department extinguishes each year. Instead, we will focus on what will be referred to as the "typical New York structural fire."

[+]The work discussed in this section was done for the New York City-Rand Institute by Keeney employed as a consultant. He wishes to thank Deputy Chief Francis J. Ronan of the New York Fire Department and Edward H. Blum of the New York City-Rand Institute for their important contributions to this work. The effort represents part of the joint work by the New York Fire Department and the New York City-Rand Institute to understand and improve the bases for deploying fire department resources. This section utilizes material originally published in Keeney [1973c].

[‡]The response time for a particular piece of equipment is defined as the time elapsed between that apparatus's leaving the fire house and its arrival at the location of the incident.

A general formulation is developed which specifies the value of response time to this "typical fire" as a function of the particular piece of equipment, the response times of the other fire-fighting vehicles, and whether, for example, it is the difference between a 2 and a 3 minute response or the difference between a 7 and an 8 minute response.

An approach to this inherently difficult problem might include (1) engineering research on fire development (e.g., how fast do different materials burn); (2) analyses of data relating losses, damage, etc. to fire department performance; and (3) analysis and distillation of accumulated fire-fighting experience.

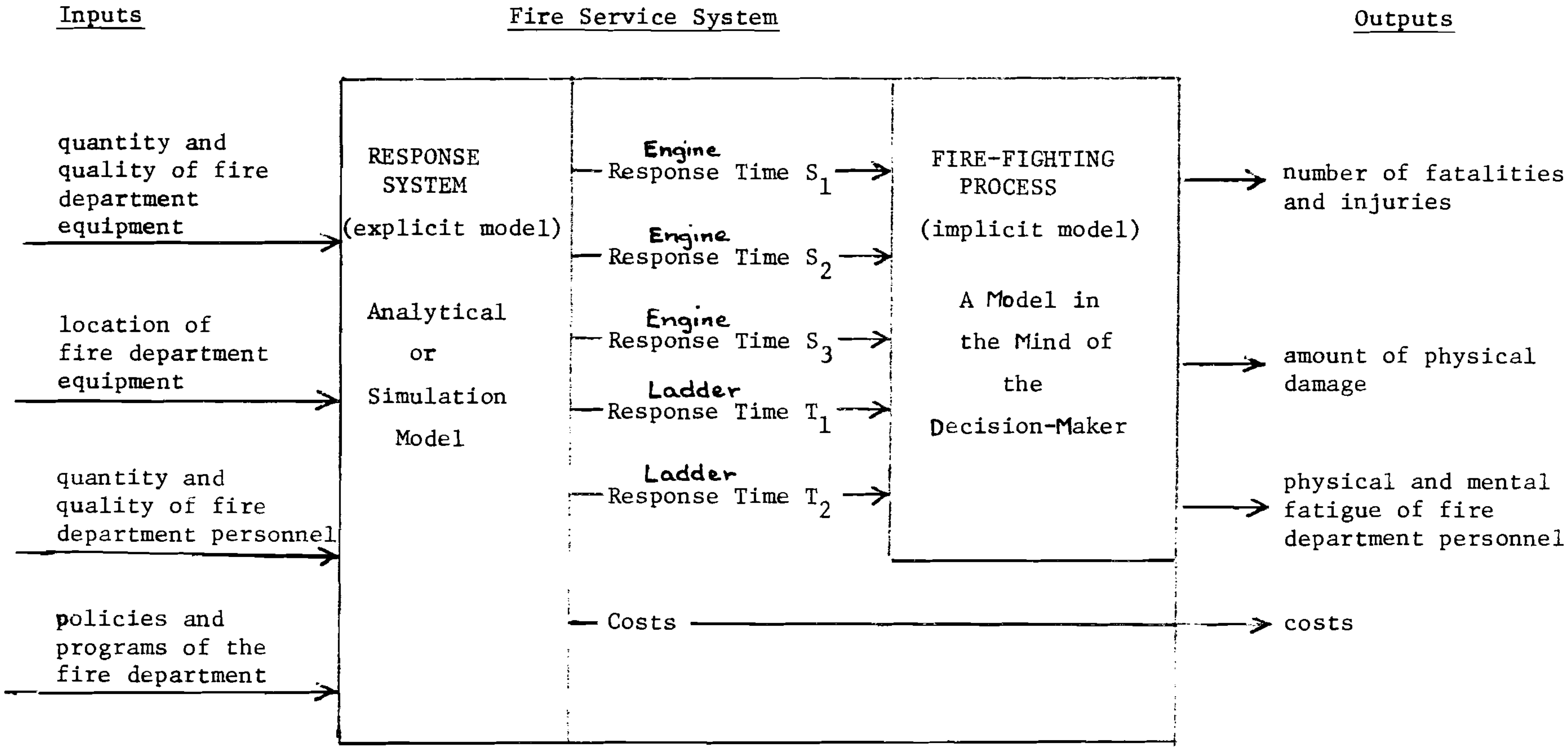
This section, by exploiting the concepts and results of utility theory discussed in earlier chapters, presents an initial attempt to quantify the experience of some New York City Fire Department officials and to investigate means of using this information for evaluating Fire Department policies. This first step involves the preferences of one deputy chief of the New York City Fire Department. A five-attribute utility function is assessed for the response times of the first three engines and first two ladders arriving at a structural fire.

7.3.1 An Overview

Let us step back and try to get an overall picture of where this work fits into Fire Department decision making. It would be desirable to evaluate proposed Fire Department policies and programs in terms of fundamental objectives such as " maximize the quality of fire service provided" and "minimize its cost." Annual cost of the Fire Department measured in dollars may be an appropriate attribute for the second objective, but there is no clear candidate for the first one. Thus, it may be necessary to divide this objective into lower-level objectives such as "minimize loss of life," "minimize injuries," "minimize property damage," "minimize psychological anxiety of the citizens," etc. Reasonable attributes for these first three objectives, are respectively the annual number of deaths, the annual number of injuries caused by fire, and the annual dollar value of lost property, whereas a subjective index would likely be required for the attribute dealing with psychological anxieties. However, these first three attributes are not exactly ideal. It is very difficult to determine what fatalities, injuries, and damage is attributable to the service of the Fire Department and what part is not. For example, an individual who causes a fire by falling asleep while smoking in bed might die before the fire is reported. This and similar fatalities should not reflect on Fire Department services. Such problems with the available data, coupled with the fact that little is known quantitatively about the fire-fighting process, contribute to the non-operational use of these measures. In addition, there are problems about the relative seriousness of different injuries and difficulties of directly placing a value on the life of an individual which further complicates matters.

Fortunately, the response times of the various apparatus responding to fires provide a natural set of proxy attributes for evaluating the level of service for such problems. Figure 7.6 is a simplified model of the fire service system illustrating that response times are inputs to the fire-fighting process, whereas objectives concerning loss of life and property damage relate to outputs.

Firemen are accustomed to thinking in terms of response times in informally evaluating their preferences for various alternative courses of action. In doing this, they use their experience in gauging both the likelihoods of the various possible response times given a particular policy and the effects these response times have on the more fundamental service objectives of the department. Aside from their interpretative appeal, data exist for specifying the probabilities of the response times conditional on a particular course of action. For nearly a decade, the New York Fire Department has kept extensive records on particular aspects of fire occurrence. These data have been analyzed and they provide the necessary input information for developing the simulation model of Fire Department operations, an early version of which is described in Carter and Ignall [1970]. This model is



A Simplified Model of a Fire Department Service System

Figure 7.6

used to generate probability density functions for the response times of any prescribed operational policy.

Our objective here is to relate the various possible response times to the accomplishment of the Fire Department's objectives for fundamental services. We want to distill years of experience of some Fire Department officials by quantifying their subjective preferences for response times to fires in a manner useful for improving the fire-fighter's decision making process. Thus, we are essentially asking the official to consider the implications of a particular set of response times (i.e., the first engine responds in 3 minutes, the second in 5 minutes,..., the first ladder responds in 2 minutes,..., etc.) on the outputs, and then to evaluate his preferences for various sets of response times in light of the respective implications. The result is a subjective model, based on experience, for the fire-fighting process, its consequences, and the relative undesirabilities of these consequences.

Whose preferences should be assessed? By virtue of their experience, it was decided that the operating chiefs were best suited to understand the multitude of implications of various combinations of response times. As a logical first step, the preferences discussed here are those of one deputy chief of the New York Fire Department.

7.3.2 Use of the Response Time Utility Function

The original motivation for assessing a utility function for response times was to develop a model for the escalation of fires. If the originally delegated units cannot control a fire, additional units must be requested, and it is said that the fire escalates. Since such events are very important to the Fire Department's performance, it would be useful to model the escalation phenomenon and include it in the simulation model. Specifically, we would like to know when poor Fire Department service leads to escalation. Since the probability of escalation is clearly related to the quality of deployment and since this quality can be measured by the response-time utility function, it may be desirable to assess the conditional probability of escalation given the quality of the response as summarized by its utility.

The utility function for response times can be useful for guiding decisions concerning operational policy of the department. Examples of such policies concern variation in initial response patterns and dispatching of vehicles, alteration of the areas of responsibility between different pieces of equipment, introduction of "special squads" during high demand hours, and temporary relocation of equipment into areas where resources are almost all working at fires. The simulation mentioned earlier and other models generate, for any given policy, probability distributions for response times. Thus, given an appropriate utility function, one can evaluate policies according to expected utility.

Let $R = \{T_1, T_2, S_1, S_2, S_3\}$ denote the attribute complex dealing with service levels and let C denote the cost attribute. Let

$$u'(c,\underline{r}) = u'(c,t_1,t_2,s_1,s_2,s_3)$$

be the overall utility for cost c and response vector $\underline{r}$. Assuming that R is utility independent of C, a most reasonable assumption, we can define a utility function $u(\underline{r}) = u(t_1,t_2,s_1,s_2,s_3)$ and from the results in Chapter 5,

$$u'(c,\underline{r}) = u'(c,t_1,t_2,s_1,s_2,s_3) = f[c,u(t_1,t_2,s_1,s_2,s_3)] \quad . \tag{7.3}$$

In our discussion, we shall confine our remarks to the assessment of $u(\underline{r})$.

Before proceeding to the assessment of the response-time utility function, let us suggest that the general ideas presented here are relevant to other emergency services such as law enforcement agencies and emergency ambulance systems. In such systems, as discussed and used by Larson [1972], Savas [1969], and Stevenson [1972] to name a few, response times are extensively used to evaluate alternative proposed policies. In all such cases, the question arises, "how much is a minute of response time worth?" The work discussed here is an initial attempt to address such questions.

7.3.3 Assessing the Response-Time Utility Function

During 1970, Deputy Chief Francis J. Ronan of the New York Fire Department and Keeney held a number of discussions to specify Chief Ronan's preferences for response times to fires. These usually lasted between 1 and 2 hours each.

Historically the traditional "standard response" in New York City has involved three engines and two ladders, so it was decided to assess a utility function over five attributes: the response times of the first and second ladders arriving at a fire and the response times of the first three arriving engines. Let us designate these attributes respectively by T_i, $i = 1,2$, and S_j, $j = 1,2,3$, and let t_i and s_j represent specific amounts of T_i and S_j respectively. Thus, we are interested in the response-time utility function $u(t_1,t_2,s_1,s_2,s_3)$.

In discussing the assessment of Chief Ronan's utility function, we will follow the guidelines for the assessment procedure suggested in Chapter 5. Thus, the discussion differentiates into five activities:

(1) familiarization,

(2) verifying the assumptions,

(3) assessing conditional utility functions,

(4) evaluating scaling constants of u,

(5) checking for consistency.

The presentation here will illustrate how the response-time utility function was assessed and what input information was necessary.

Familiarization. Before beginning on this problem, Chief Ronan and Keeney had worked together on a very simple decision-analytic model of a fire-response problem. Also, at an earlier time, Keeney had roughly assessed Chief Ronan's utility function for the response time of the first engine arriving at a fire. The main purpose of this preliminary exercise was to check whether it was reasonable simply to minimize the expected response time of the first engine arriving at a fire. In most analytical studies dealing with emergency services, this linearity assumption (i.e., minimize expected response time) is implicitly used. As a result of this initial exposure, the Chief was introduced to the basic concepts of utility theory. After the first two assessment sessions, which turned out to be learning experiences for both analyst and respondent, the discussions became more productive.

Verifying the Assumptions. To exploit the theory of Chapters 5 and 6, it was necessary to check whether requisite utility independence assumptions were appropriate for this problem.

Specifically, it was verified that it was reasonable* to assume:

(1) engine response times $\{S_1, S_2, S_3\}$ and the ladder response times $\{T_1, T_2\}$ were utility independent of each other,

(2) first ladder response T_1 and second ladder response T_2 were utility independent of each other, and

(3) the j^{th} engine response S_j was utility independent of the other engine responses, for $j = 1, 2$, and 3.

Because of Theorem 5.3 and (1), the assessment of u was broken into two parts: assessments of an engine utility function and a ladder utility function. Analogously, these two utility functions could be broken down into component parts because of (2) and (3).

Let us illustrate the verification procedure with an example. To check if T_1 was utility independent of T_2, Chief Ronan was asked "Given that the response time of the second arriving ladder is fixed at six minutes, what response time t_1 for the first arriving ladder would be indifferent to having a 50-50 chance that the first ladder responds in either one or five minutes?"

---

*The following independence assumptions were deemed to be approximately valid after considerable probing. Some of the dependencies were so slight--contrary to what was first expected--that independence was taken as an innocuous idealization.

Notice that if $t_2 = 6$, then $t_1 < 6$ and this restricts the domain conveniently. A response that $t_1 = 3.4$ minutes was eventually chosen using a "convergence" technique discussed in Section 4.9.

Next we asked the same question only the second ladder response time was fixed at eight rather than six minutes. Again, the indifference response was 3.4 minutes, leading us to believe that the relative preferences for changes in the response time of the first ladder did not depend on the <u>fixed</u> response time of the second ladder. By additional questioning similar to the above, this speculation was confirmed. Thus, it seemed appropriate to assume that T_1 was utility independent of T_2.

<u>Assessing Conditional Utility Functions</u>. Given the assumptions above, utility functions were needed for each of the five response time attributes. Actually these are conditional utility functions since they concern preferences over a single response time given that the other response times are held fixed. However, because of the utility independence conditions, the particular amounts of these other responses are not important, since the utility function should be the same in any case. To illustrate the approach, let us assess $U_1^T(t)$, the conditional utility function for the first arriving ladder.

Through the questioning, we found a 2.2 minute response of the first arriving ladder was indifferent to a 50-50 chance at either a one or three minute response. Similarly, 4.2 minutes was indifferent to a 50-50 chance at three or five minutes, and 6.2 minutes was indifferent to a 50-50 chance at 5 or 7 minutes. In general, a 50-50 chance at either a t or a (t + 2) minute response was indifferent to a (t + 1.2) minute response for certain. As indicated in Chapter 4, such preferences imply the utility function must be of the form

$$u_1^T(t) = d + b(-e^{ct}) \tag{7.4}$$

where d and b, c > 0 are constants. Since utility functions are unique up to positive linear transformations, it was decided to scale u_1^T from minus one to zero. In addition, the response times ranged from zero to twenty minutes, which implied

$$u_1^T(0) = 0 \tag{7.5}$$

and

$$u_1^T(20) = -1 \quad . \tag{7.6}$$

Next, a 4.5 minute response time for the first ladder was found to be indifferent to a 50-50 lottery yielding either one or seven minutes. Hence u_1^T must be such that

$$u_1^T(4.5) = 1/2\ u_1^T(1) + 1/2\ u_1^T(7) \quad . \tag{7.7}$$

Substituting (7.4) into (7.5) through (7.7) yields three equations with three unknowns which can easily be solved to give

$$u_1^T(t) = 0.0998\left(1 - e^{.12t}\right) . \tag{7.8}$$

Similar procedures were used to obtain the other four conditional utility functions.

<u>Evaluating Scaling Constants of u</u>. Given the individual utility functions for the five response times, the next step is to put them together in the appropriate manner to obtain the overall utility function for response times. This requires assessing the scaling constants--that is, the k's--of Theorems 5.3 and 6.3. To illustrate the method, let us use the ladder-response utility function

$$u_L(t_1,t_2) = k_1u_1^T(t_1) + k_2u_2^T(t_2) + [k_1 + k_2 - 1]u_1^T(t_1)\ u_2^T(t_2) \quad . \tag{7.9}$$

Chief Ronan was asked for the response time t_2 of the second ladder such that he would be indifferent between the two ladders arriving in three and eight minutes respectively, denoted by (3,8), and the response $(4,t_2)$. His answer was $t_2 = 5.7$ indicating a willingness to give up one minute of first ladder response in exchange for decreasing second ladder response by 2.3 minutes, <u>given</u> he started from (3,8). This implied

$$u_L(3,8) = u_L(4,5.7) \quad . \tag{7.10}$$

Similarly, we found (2,6) indifferent to (3,4.2) so

$$u_L(2,6) = u_L(3,4.2) \quad . \tag{7.11}$$

Using (7.9) and the individual utility functions to evaluate both sides of (7.10) and (7.11) gives us two equations with two unknowns, the parameters k_1 and k_2, which when solved yields

$$u_L(t_1,t_2) = 0.66u_1^T(t_1) + 0.19u_2^T(t_2) - 0.15u_1^T(t_1)\, u_2^T(t_2) \quad . \tag{7.12}$$

Other parameters of the overall utility function were evaluated in similar ways as covered in Section 6.6. The general idea is to ask questions to obtain equations containing the unknown parameters, and then to solve the set of equations for the parameter values.

Checking for Consistency. Checking the utility function for consistency and reasonableness is obviously very important--both because the assessment is inherently a subjective process and because the synthesis required to obtain the overall utility function can result in the introduction of "errors." It was important to make sure the implications of the utility function agreed with the chief's preferences.

The most important checks concern the conditional utility functions and tradeoffs between the various response times. This involved discussing the implications of the utility function and using the utility function for providing answers to questions like those asked in the assessment process. In all cases where there was a major discrepancy between the implications of the utility function and the chief's preferences, part of the assessment procedure was repeated and his utility function adjusted accordingly. Many parts of the utility function were adjusted in light of consistency checks. The final utility function appears to represent Chief Ronan's responses quite closely.

7.3.4 The Response-Time Utility Function

In this section, we present the final form of the "first-cut" utility function and discuss its implications. From our assessments, we found

$$u(\underline{t},\underline{s}) = 0.24u_L(\underline{t}) + 0.16u_E(\underline{s}) - 0.6u_L(\underline{t})\, u_E(\underline{s}) \quad , \qquad (7.13)$$

where

$$u_L(\underline{t}) \equiv u_L(t_1,t_2) = 0.66u_1^T(t_1) + 0.19u_2^T(t_2) - 0.15u_1^T(t_1)\, u_2^T(t_2) \quad , \qquad (7.14)$$

with

$$u_1^T(t_1) = 0.0998\left[1 - e^{.12t_1}\right] \quad , \qquad (7.15)$$

and

$$u_2^T(t_2) = 0.143\left[1 - 0.5e^{.08t_2} - 0.5e^{.12t_2}\right] \quad , \qquad (7.16)$$

and where

$$u_E(\underline{s}) \equiv u_E(s_1,s_2,s_3) = 0.63u_1^S(s_1) + 0.18u_2^S(s_2) + 0.09u_3^S(s_3)$$

$$- 0.06u_1^S(s_1)\ u_2^S(s_2) - 0.03u_1^S(s_1)\ u_3^S(s_3)$$

$$- 0.01u_2^S(s_2)\ u_3^S(s_3) \quad , \tag{7.17}$$

with

$$u_1^S(s_1) = 0.0845\left[1 - 0.4e^{.1s_1} - 0.6e^{.14s_1}\right] , \tag{7.18}$$

$$u_2^S(s_2) = 0.156\left[1 - e^{.1s_2}\right] \tag{7.19}$$

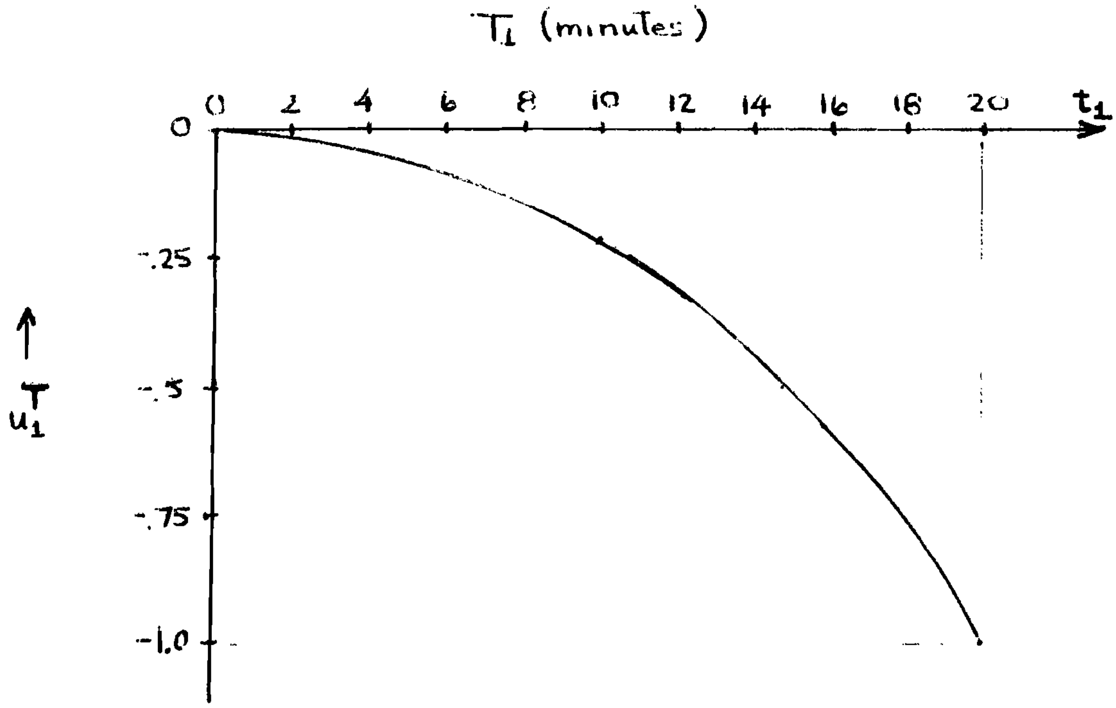
and

$$u_3^S(s_3) = 0.253\left[1 - e^{.08s_3}\right] . \tag{7.20}$$

For illustrative purposes, the utility function in (7.15) is shown in Figure 7.7 and the indifference map implied (7.14) is given in Figure 7.8.

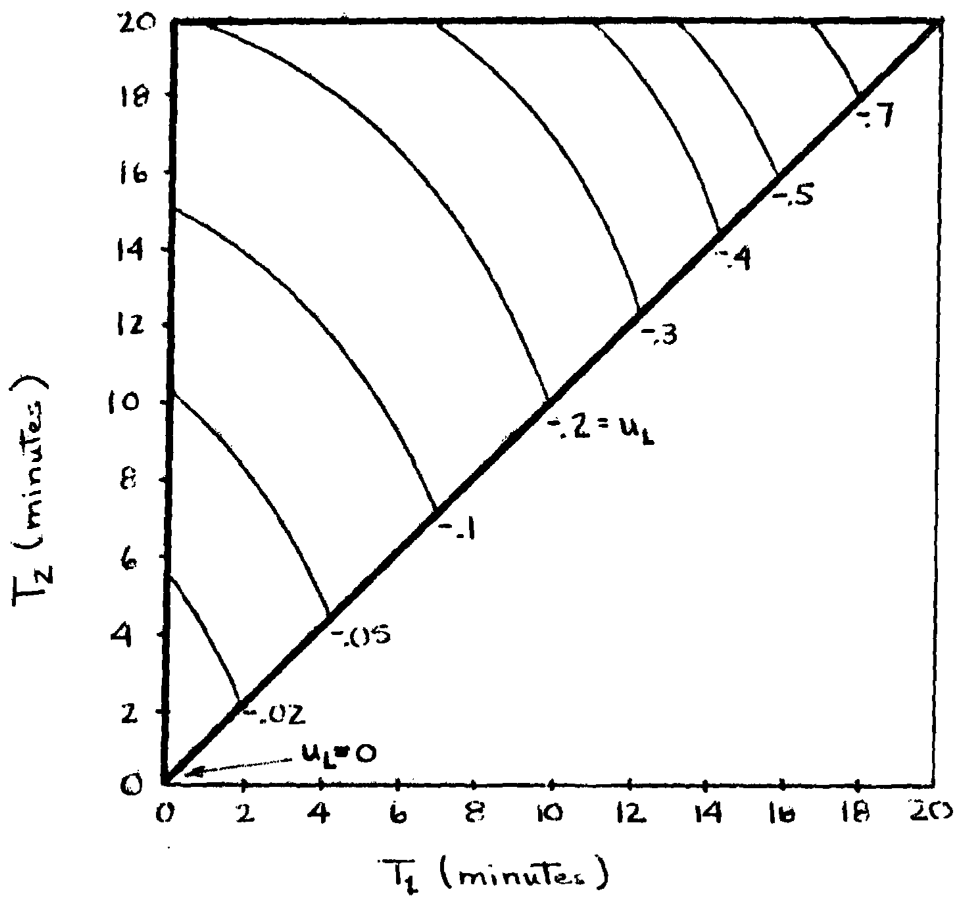
It was decided to evaluate preferences in the unit hypercube from (0,0,0,0,0) to (20,20,20,20,20). Thus, for each of the equations above, the variables may only range from zero to twenty minutes. Furthermore, by definition, we have $t_1 \leq t_2$ and $s_1 \leq s_2 \leq s_3$.

Properties of the Utility Function. The utility function u in (7.13) has several properties which are intuitively appealing and which appear to represent Chief Ronan's preferences. Some of these pertain to u as a whole, some to the utility function for ladders u_L or the utility function for engines u_E, and some to the utility functions of the individual units. Taking the latter first, we have



Utility Function for First Ladder Response Time

Figure 7.7



Indifference Curves for Ladder Response Times

Figure 7.8

(1) u is decreasing in each t_i and s_j. This means the sooner a particular unit arrives, the better, given the response times of other units are fixed.

(2) Each minute of delay of the first arriving engine is more important* than a corresponding minute for the second arriving engine, which in turn is more important than the corresponding delay of the third arriving engine. Similarly, each minute of delay of the first ladder is more important than a corresponding delay of the second ladder. These properties are indicated by the relative values of the coefficients of the u_i^T terms in (7.14) and the u_j^S terms in (7.17).

---

*To clarify the meaning of <u>more important</u>, recall the utility function (7.3) for cost and response times $u'(c,\underline{r}) = u'(c,t_1,t_2,s_1,s_2,s_3)$ and assume the cost attribute C and the set of response attributes $R \equiv \{T_1,T_2,S_1,S_2,S_3\}$ are utility independent of each other. Now select any base level cost c_0 and consider changes $\underline{r}'$ to $\underline{r}''$ and $\hat{\underline{r}}$ to $\hat{\hat{\underline{r}}}$, each of which will be assumed to be for the better. We will say the change $\underline{r}'$ to $\underline{r}''$ is more important than the change $\hat{\underline{r}}$ to $\hat{\hat{\underline{r}}}$ if $c_1 > c_2$, where c_1 and c_2 are defined by $u(c_0,\underline{r}') = u(c_1,\underline{r}'')$ and $u(c_0,\hat{\underline{r}}) = u(c_2,\hat{\hat{\underline{r}}})$. That is, one must be willing to pay more in cost to make the <u>more important</u> change.

(3) The conditional utility function for each attribute is risk averse regardless of the values of the other attributes. This means, for instance, that for T_1, a sure $(t_1'' + t_1')/2$ minute response is preferred to a 50-50 lottery yielding either t_1'' or t_1'. Said another way, the average response time is preferred to the lottery. When this is the case, for each unit, each additional minute of delayed response is more important than the former minute.

Concerning the utility function for ladders, we have:

(4) The relative importance of the response time of the i^{th} ladder increases as the response time of the other ladder increases. Said loosely, this means the slower the first ladder is in arriving, the more important it is that the second ladder arrive soon afterwards. This property is accounted for by the fact that the product term in (7.14) has a negative coefficient.

Similarly, for engines, there is an analogous property:

(5) The relative importance of the response time of the j^{th} engine increases as the response times of the other engines increase. This property is accounted for by the negative coefficients of the product terms in (7.17).

The last two properties concern the entire utility function. They are

(6) A one minute delay in the arrival of the i^{th}

ladder is more important than the corresponding minute delay on the i^{th} engine. Thus, for example, using two minutes responses for both the first engine and the first ladder as a base, we would prefer to have the first ladder respond in two minutes and the first engine in three than to have the first engine respond in two minutes and the first ladder in three. This property is indicated by the fact that the coefficient of u_L in (7.13) is larger than the coefficient of u_E.

(7) The relative importance of the response times of ladders increases as the response times of engines increase. This means the importance of the first arriving engine is less when a ladder has already arrived than it is when no ladders have arrived. The negative coefficient of the product term in (7.13) indicates this property holds.

These properties, each of which is intuitively reasonable, go a long way toward specifying the utility function. That is, the manner in which the shape of the utility function (7.13) can be altered without violating one or more of the above conditions is severely restricted. This fact lends some additional confidence to our assessments.

Although the complexity of assessing a multiattribute utility function increases rapidly as the number of measures of effectiveness increase, the opportunity for "consistency checks" involving properties such as those above also greatly increases. In order to meaningfully represent one's preferences in these complex situations, it is important to exploit such intuitively appealing attitudes toward preference to the fullest extent possible.

7.3.5. Conclusions

The main result of this work is a "first cut" utility function over five response-time attributes, namely, those associated with the first two ladders and the first three engines arriving at a fire. This gives us some means for determining the relative values of a minute of response time for the various pieces of equipment. By looking at the coefficients of the single u_i terms of (7.13), one can get a very rough idea of the relative values of a minute of response time for the different pieces of apparatus. Doing this, if we set the relative value for a minute of response time of the first ladder at 10, the corresponding value for first engine is 7, the second ladder is 3, the second engine is 2, and the third engine is 1.

However, as we have mentioned, the worth of a minute of response time on a specific vehicle depends on the response times of the other pieces of equipment and the time since the alarm was reported. So for instance, using a (2,4;2,4,6) response* as a base, the partial derivatives of u with respect to the five response times are in a ratio of 10:4:5:3:2 implying that if the relative value of a minute of response time of the first ladder is set at 10, the corresponding value of the second ladder is 4, the first engine is 5, the second engine is 3, and the third engine is 2. The point is that the relative values depend on the base response.

The assessment procedure was too time consuming and too complex. Since it was impractical to develop a computer program to help assess one utility function, calculations were done by hand. Thus, there was a lack of immediate feedback to Chief Ronan concerning the implications of his preferences. Often this caused small differences in the chief's responses during different sessions due to the slight variation of his preferences from time to time. But, of course, the involvement over a considerable time span has its merits too. We would like the assessed utility structure to be somewhat stable over time. In the future, an interactive computer program, such as the one discussed in

---

*The first ladder responds in 2 minutes, the second ladder in 4 minutes, the first, second, and third engines in 2, 4, and 6 minutes, respectively.

Appendix 6c, would likely help maintain interest as well as assess the utility function much more quickly with many more consistency checks.

By asking Chief Ronan about his preferences for responses to the "typical structural fire," we essentially asked him to synthesize in his mind all the possible implications of each response aggregated over the possible types of structural fires. This understandably caused some discrepancies in the answers to our questions, because of the tendency to focus on particular types of incidents at different times. Since our major interest in this particular work centers on the first broad cut, rather than details relevant to particular fires, the aggregation requirement may be reasonable.

Our ultimate objective is to obtain a utility function appropriate for the use of the New York Fire Department. This section reports a first step: assessing a utility function of one Deputy Chief of that department. However, the Chief's preferences are his and not necessarily those of the Fire Department, and they should not be interpreted as such. Furthermore, although a serious attempt was made not to lead the Chief to any specific answers, his responses to questions could have in part been shaped by the questioning process, and the results should be interpreted with this possibility in mind.

This assessment exercise was done about five years before the present monograph was sent to the press and if the exercise were to be repeated again, we probably would now proceed somewhat differently. We would attempt to establish some broad, basic, underlying principles, which seemed to govern Chief Ronan's responses and then to deduce more of the structure of his utility function from these basic principles. Essentially we would try to model, to some extent, his motivations based on interviews which would probe more deeply on qualitative matters. This, of course, is easier said than done, and we would like someday to be able to report a good example of this technique. This is the trouble in writing about a subject in its infancy.

7.4 STRUCTURING CORPORATE PREFERENCES FOR MULTIPLE OBJECTIVES*

Every corporation periodically asks itself: "How should we run our business?" More specifically, this raises such questions as: Given the complex social, economic, technological, and political characteristics of our society, which management policies should we adopt now? Are these policies consistent with our personal objectives, with the desires of our shareholders, and with our social value structure? If we choose policy A, will it

---

*We would like to thank the Board of Directors of Woodward-Clyde Consultants for its permission to discuss this work in our book. The assistance of Dr. Keshavan Nair of Woodward-Clyde in writing this section is greatly appreciated. Material in This section is adapted from Keeney [1975].

be possible to account for the contingencies which may arise in the near future and adapt accordingly? How can we best maintain the leadership position in our field and simultaneously, keep the vitality of our organization? All of these are crucial questions which deny the simple dollars and cents answers which are mythically supposed to be appropriate for almost all "business" decisions.

Since early 1972, Woodward-Clyde Consultants, a holding firm for several professional-service consulting firms has used some innovative approaches based on ideas discussed in this book to help them examine questions such as those raised above.* Although this effort is still in progress, it is sufficiently interesting and informative to include here. Two aspects of this effort seem to be unique. First, multiattribute utility functions over attributes measuring fundamental objectives

---

*In November, 1974, Woodward-Clyde made some very broad organizational changes. It is no longer a holder firm but rather one consulting firm with five regional divisions. The work described in this section was done from 1972 through October, 1974, so the organizational structure which prevailed during that period is described. The subsequent organizational changes are briefly summarized at the end of the section.

of the corporation have been assessed for many executives at Woodward-Clyde. Second, this work was done not to evaluate a specific decision, but rather:

- to aid communication among the decision makers,
- to grapple with fundamental issues of the firm,
- to determine and examine differences of opinion in a quantitative fashion, and
- to aid in generating creative alternatives in solving corporate problems.

The affiliate consulting firms of Woodward-Clyde Consultants operate mainly in the geotechnical engineering and environmental areas. Problems they examine include design of earth dams, siting and design of nuclear power plants, geotechnical and environmental studies associated with pipeline systems (e.g., the Trans-Alaska pipeline), and design of structures for earthquake-prone regions. None of the affiliates build any products (e.g. roads, dams, power plants); they are exclusively professional-service consulting firms. Collectively, their fees received in 1973 were approximately 25 million dollars, and historically, this has increased at approximately twenty percent annually. All the shareholders of Woodward-Clyde must be senior professionals on the staff of one of the affiliates.

In 1972, Richard J. Woodward, the Chairman of the Board of Woodward-Clyde Consultants, appointed a long-range planning committee whose assignment included "the development of a long-range plan for Woodward-Clyde Consultants that includes quantified objectives and is responsive to the Statement of Purpose and Standing Policies." After this original committee reported, the 1973 and 1974 Long Range Planning Committees have successively updated the objectives of Woodward-Clyde and examined policy alternatives in terms of these objectives. Douglas C. Moorhouse was the chairman of each of these three committees. Dr. Keshavan Nair, a Vice President of Woodward-Lundgren and Associates, one of the affiliates of Woodward-Clyde was also a member of these committees.

Much of the work discussed here, specifically Sections 7.4.2 through 7.4.5, was done jointly by Dr. Nair and Ralph L. Keeney, working as a consultant to Woodward-Clyde. Section 7.4.1 discusses the original Long-Range Planning Committee's work, which has served as an excellent basis on which to build. The final Section 7.4.6 surveys some of the specific uses being made of Woodward-Clyde's utility function. We just remark here that the purpose in assessing a utility function was <u>not</u> to help management choose amongst action alternatives in a formal manner--e.g., probabilistic analysis was not done in accompaniment of utility analysis--but rather to help management articulate some of its basic assumptions and to facilitate communications amongst the executive group. This, to a large extent, was, and is being, accomplished via the formal assessment procedures described below.

7.4.1 The 1972 Objectives and Measures of Effectiveness

The basic approach taken by the 1972 Long Range Planning Committee to fulfill its mission was (1) to establish the primary objective of the firm, (2) to divide this into subobjectives, and (3) to conduct a deficiency analysis indicating discrepancies between present state and desired state on each objective. By weighting the various objectives, the deficiencies were ranked in order of importance and policies recommended for eliminating these deficiencies.

The overall objective of Woodward-Clyde was provided by a sentence in their Statement of Purpose: "The combined efforts of Woodward-Clyde Consultants and its affiliates are directed toward the creation and maintenance of an environment in which their employees can realize their personal, professional, and financial goals." It was felt that growth was essential in the achievement of this objective.

The hierarchy of objectives developed by the 1972 Long Range Planning Committee is presented in Figure 7.9. This hierarchy has been adaptively revised since that time. The numbers in parentheses in the box with each objective indicates the original division of weight among subobjectives. More will be said about this later. In

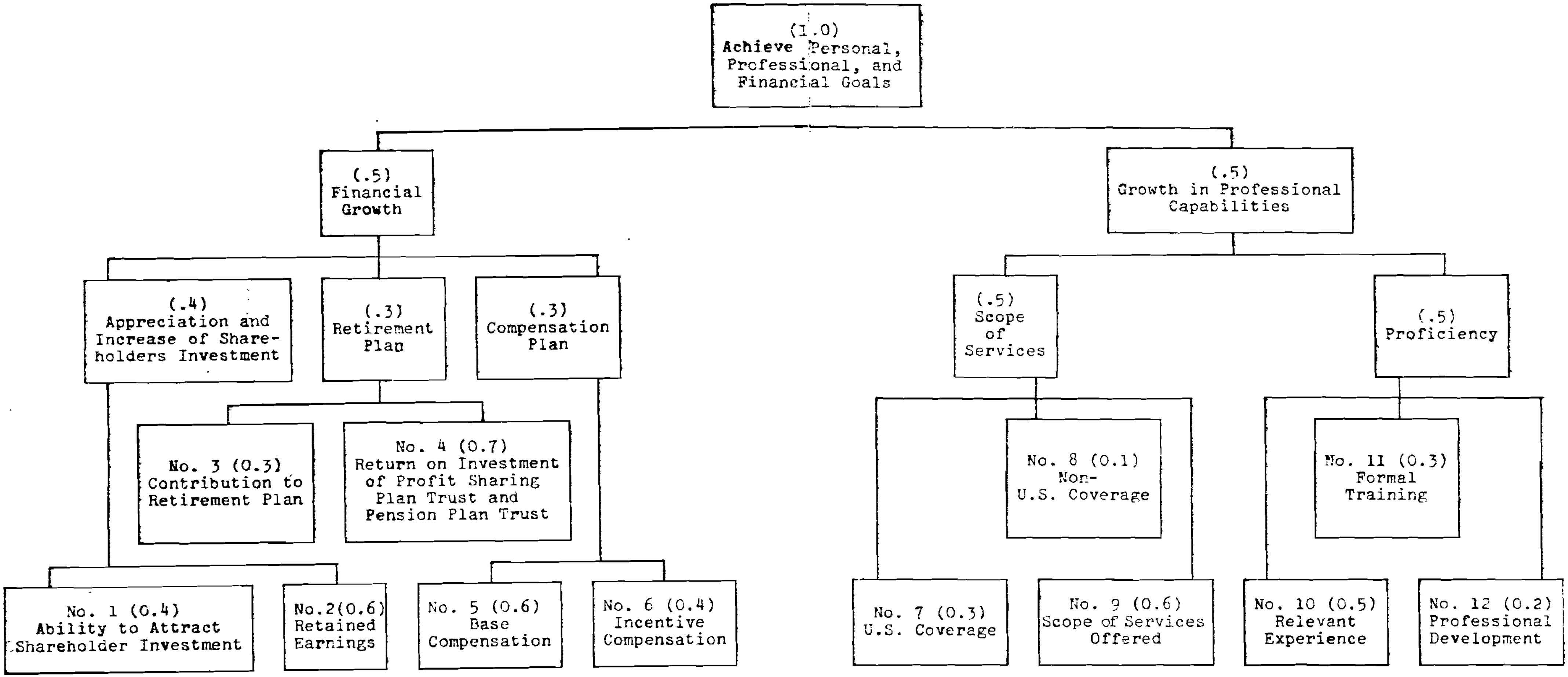


FIGURE 7.9 1972 OBJECTIVES HIERARCHY OF WOODWARD-CLYDE CONSULTANTS

Table 7.5, the weights of each of the attributes associated with the lowest-level objectives and the range of each attribute are identified.

It was implicitly assumed that an additive value function

$$v(x_1,x_2,\ldots,x_{12}) = \sum_{i=1}^{12} k_i v_i(x_i) \quad ,$$

where the x_i's represent levels of the attributes, each v_i is a value function over the i^{th} attribute, v and the v_i's are scaled zero to one, and the weights, that is the k_i's sum to one, was appropriate. For each attribute, component value functions were constructed and present states and desired states, defined as the practical maximum felt to be achievable, were identified. Deficiency on each of these lowest-level objectives was then calculated by multiplying the weight of the objective times the difference in the value of its present and desired states. This indicated "areas" where approvement was needed.

Four shortcomings of the 1972 "quantification of objectives" might be categorized as follows:

(1) the weights were assigned to each objective without explicitly considering the range of the associated attributes,

(2) the component value functions were estimated by a direct value estimation technique independent of each other,

TABLE 7.5

1972 ATTRIBUTES FOR WOODWARD-CLYDE CONSULTANTS

ATTRIBUTE	MEASUREMENT UNIT	RANGE	ATTRIBUTE WEIGHT
Ability to attract shareholders investment	$\frac{\text{Number of shares requested}}{\text{fees}}$ %	0-5	.08
Retained earnings	% of fees	0-8	.12
Contribution to retirement plan	% of fees	0-10	.045
Return on investment for retirement plan	% of investment	0-20	.105
Base compensation	% annual increase	0-20	.09
Incentive compensation	% of fees	0-8	.06
U.S. coverage	$\left(\frac{\text{Geographic centers adequately covered}}{\text{Centers where relevant work can be generated}}\right)$ %	25-100	.075
Non-U.S. coverage	$\left(\frac{\text{Geographic centers adequately covered}}{\text{Centers where relevant work can be generated}}\right)$ %	0-50	.025
Scope of services offered	$\left(\frac{\text{Number of disciplines having threshold capability}}{\text{Number of synergistic disciplines required by society}}\right)$ %	25-100	.15
Relevant experience	$\left(\frac{\text{Existing man-years experience}}{\text{Required man-years experience}}\right)$ %	25-100	.125
Formal training	Number of degrees per professional staff member	1-3	.075
Professional development	% of fees	0-2	.05

(3) the overall objective function, being a value function, was not appropriate for examining policies with uncertain consequences,

(4) the additive value structure did not lend itself to investigating overlap among the objectives.

Even with these weaknesses, the Long Range Planning Committee and the Board of Directors felt this quantification of objectives was a big improvement over informally articulated objectives. This set of objectives and measures has proven to be an excellent basis for modification and improvement, the substance of which we begin to describe in the next subsection.

Before proceeding, let us briefly remark on aspects of the attributes and their measurement units which may not be clear from Table 7.5. For the first attribute, using the number of shares requested divided by fees implicitly assumes the cost of a share is known in order to make the measure readily interpretable. The measure of the scope of services offered is an index meant to indicate breadth in handling the interdisciplinary projects increasingly requested by society. With relevant experience, the idea is to have the staff available to do quality work on those projects which the Woodward-Clyde affiliates would like to do. For formal training, the number of degrees per professional staff member is defined as follows: a doctorate is three, a masters degree two, and a bachelors one. Professional development includes attending management or technical seminars, holding in-house study sessions, etc.

7.4.2 Clarifying the Measures of Effectiveness

One of the first issues Drs. Nair and Keeney jointly considered was whether the measures of effectiveness met the comprehensiveness and measurability criteria discussed in Chapter 2. For each objective, the question "Can a better attribute be found?" was asked. In several cases, the answer was "yes." Let us discuss some examples.

(a) Ability to Attract Shareholders Investment. The measurement unit for this attribute was changed to the dollar value of shares requested divided by the fees. Thus in interpreting trends, and simply in evaluating various levels of the attributes, one does not need to keep the value of the shares implicitly in mind.

(b) Scope of Non-U.S. Coverage. The 1974 Long Range Planning Committee changed this measure to percentage of the United States business in terms of fees received. It was the Committee's viewpoint that the major reason for expanding overseas was to reduce the consequences of a possible recession in the United States and to take advantage of current foreign opportunities. Since Woodward-Clyde will remain primarily a U.S. operation in the foreseeable future, the new measure both is more easily quantifiable than the previous one and also more directly indicates vulnerability to domestic recessions.

(c) Relevant Experience and Professional Development. As demand for Woodward-Clyde services increases, the need to increase their relevant experience grows. The 1972 measure of relevant experience indicated the level at any given time, as opposed to focusing on the increase of relevant experience. Increased relevant experience is funded out of the Professional Development budget and usually consists of opportunities for employees to work on projects under experienced personnel at company expense and to take specialized courses in areas of their practice. Because it is the increase in relevant experience which is currently important at Woodward-Clyde, the measure was changed to percent of fees committed to the relevant experience program.

This change of the relevant experience measure required a redefinition of the components of the professional development measure. In 1972, the latter measure included fees used for obtaining relevant experience. However, with the new relevant experience measure, the professional development measure must explicitly exclude the fees used for acquiring relevant experience.

(d) Formal Training. The measure remained the same for formal training but the desirability of particular levels has greatly changed. The value function in this case is interesting in that it is not monotonic. It is low at a level of 1, since all professionals then only

Table 7.6 An Aid for Evaluating Preferences for the Attribute Formal Training as Measured by the Degrees per Professional Staff Member.

DEGREES PER PROFESSIONAL STAFF MEMBER	5% PhD BS MS PhD	10% PhD BS MS PhD	15% PhD BS MS PhD	20% PhD BS MS PhD	25% PhD BS MS PhD	30% PhD BS MS PhD	35% PhD BS MS PhD	40% PhD BS MS PhD	45% PhD BS MS PhD	50% [illegible] BS [illegible]
1.5	55, 40, 5	60, 30, 10	65, 20, 15	70, 10, 20	75, 0, 25	Not possible	Not possible	Not possible	Not possible	Not possible
1.6	45, 50, 5	50, 40, 10	55, 30, 15	60, 20, 20	65, 10, 25	70, 0, 30	"	"	"	"
1.7	35, 60, 5	40, 50, 10	45, 40, 15	50, 30, 20	55, 20, 25	60, 10, 30	65, 0, 35	"	"	
1.8	25, 70, 5	30, 60, 10	35, 50, 15	40, 40, 20	45, 30, 25	50, 20, 30	55, 10, 35	60, 0, 40	"	
1.9	15, 80, 5	20, 70, 10	25, 60, 15	30, 50, 20	35, 40, 25	40, 30, 30	45, 20, 35	50, 10, 40	55, 0, 45	"
2.0	5, 90, 5	10, 80, 10	15, 70, 15	20, 60, 20	25, 50, 25	20, 40, 30	35, 30, 35	40, 20, 40	45, 10, 45	5[illegible]
2.1	Not possible	0, 90, 10	5, 80, 15	10, 70, 20,	15, 60, 25	20, 50, 30	25, 40, 35	30, 30, 40	35, 20, 45	40, [illegible]
2.2	"	Not possible	Not possible	0, 80, 20,	5, 70, 25	10, 40, 30	15, 50, 35	20, 40, 40	25, 30, 45	[illegible]
2.3	"	"	"	Not possible	Not possible	70, 30	5, 60, 35	10, 50, 40	15, [illegible], 45	[illegible]
2.4	"	"	"	"	"	Not possible	Not possible	0, 60, 40	5, 50, 45	10, 40, [illegible]
2.5	"	"	"	"	"	"	"	Not possible	Not possible	[illegible]

have a bachelors degree, and increases to a peak and then falls rapidly as the level of degrees increases. With a level of 3, the firm would consist entirely of professionals with doctorates. In 1972, the desired state was identified as 2.25, the peak of the value function. On further examination, this level seemed high. If just 25 percent of the professionals of Woodward-Clyde had only a bachelors, a minimum of 50 percent would have to have a doctorate to get the average level to the "desired state" 2.25.

As an aid to thinking about the implications of different levels of "degrees per professional," Table 7.6 was constructed. For evaluating preferences over average degree levels, an individual is meant to select the best distribution of degrees for each average level, and then compare these "best" distributions.

7.4.3 Checking for Independence Conditions

To structure a utility function over the twelve attributes of Table 7.5, modified as indicated in the previous subsection, the process began by examining whether pairs of attributes were preferentially independent of their complements.* In most cases it seemed

*Initial assessments were done using Dr. Nair's preferences. Subsequently, Dr. Nair has assessed the preferences of other members of the Long Range Planning Committee.

appropriate to assume preferential independence, but let us indicate three situations where this was not so.

In examining preferential independence assumptions involving the attribute "ability to attract shareholder investment," the Long Range Planning Committee came to the agreement that it was redundant based on present policy. This attribute was meant to indicate the ability and desirability for principals to invest in the corporation. The Committee felt the desirability aspect was adequately captured by retained earnings. On the other hand, the ability to invest was measured by both incentive compensation and base compensation. For these reasons, the "ability to attract shareholder investment" was dropped from the list of attributes.

In another case it at first seemed advantageous to subdivide the objective concerning base compensation into three groups: senior principals, junior principals and associates, and associate candidates. In effect, the current attribute "base compensation" would have been replaced by three attributes, namely base compensation for senior principals, base compensation for junior principals and associates, and base compensation for associate candidates. It was found that one of these attributes taken together with a different attribute, say retained earnings, was not preferentially independent of its complement. The reason was that the rate at which one would substitute retained earnings for base compensation for associate candidates depended on the level of base compensation increases to the principals and associates. If these latter groups received large increases in base compensation, it seemed reasonable to give up more retained earnings to bring increases in base compensation for associate candidates up to some comparable level, than one would give up to make the same increase for associate candidates if in fact the other groups received low increases in base compensation. The concept of equity among the three groups made it inappropriate to assume preferential independence in this case.

There were two other possibilities investigated. Each pair of the three base compensation attributes was found to be conditionally preferentially independent of the third given all other attributes are fixed at an arbitrary level. This would have allowed us to construct an additive component value function over the three attributes. The alternative was to use the original aggregated base compensation attribute. It was felt that members of the Long Range Planning Committee could keep the equity considerations in mind when using the aggregated attribute. Therefore, since it is simpler to use one attribute than the three component attributes, the former was chosen.

Base compensation and incentive compensation do have some overlap in purpose and, because of this, the latter paired with, for instance, retained earnings is not exactly preferentially independent of its complement. However, the overlap is not great since the function of the former is to provide a solid salary for competent work within the "normal" call of duty, whereas the function of the later is to provide motivation and reward for efforts 'beyond' the call of duty. Hence after considerable checking, it was decided that it was a reasonable approximation to assume the preferential independence condition. This "appropriateness" decision was taken in conjunction with the decision to eliminate the attribute "ability to attract shareholder investment" from the list in Table 7.5.

It was decided that the two attributes concerning retirement plan should be aggregated into one called "growth in retirement plan," since in fact both seemed to meet the same fundamental objective. Woodward-Clyde desires that any participant in their retirement plan receive a combined amount from the plan and social security equal to 50 percent of his or her last five years average salary. The new measure for "growth of retirement plan" is the annual increase of assets in the retirement plan. Its range is zero to thirty percent, and it should be clear that this excludes the social security benefits. In effect, this change is simply moving up the objectives hierarchy of Figure 7.9 for a quantitative assessment of retirement plan consequences.

7.4.4 The 1974 Objectives and Measures of Effectiveness

The objectives and attributes updated from the original 1972 list are given in Table 7.7. After considerable examination, Dr. Nair felt that it was appropriate to assume that for the ranges given in the table, each pair of attributes was preferentially independent of its complement. The reasonableness of this assumption has been preliminary accepted by each of the other members on the 1974 Long Range Planning Committee.

7.4.5 Assessing the Utility Function

The preferential independence conditions imply that an additive value function exists over the ten attributes in Table 7.7. From Theorems 6.1 and 6.2, by verifying that just one attribute is utility independent of its complement, either a multiplicative or additive utility function is appropriate to quantify preferences. It was verified that retained earnings was in fact utility independent of its complement, and utility independence was also verified for other attributes to serve as consistency checks. For future reference, it turned out, the final utility function over the attributes in Table 7.7 was multiplicative, and thus expressible in the form

$$1 + ku(\underline{x}) = \prod_{i=1}^{10} [1 + kk_i u_i(x_i)] \quad , \qquad (7.21)$$

where u and the u_i's are scaled zero to one, $0 < k_i < 1$,

TABLE 7.7

1974 ATTRIBUTES FOR WOODWARD-CLYDE CONSULTANTS

ATTRIBUTE	MEASUREMENT UNIT	RANGE
$X_1 \equiv$ Retained earnings	% of fees	0-8
$X_2 \equiv$ Growth in Retirement Plan	% of existing assets	0-30
$X_3 \equiv$ Base Compensation	% annual increase	0-30
$X_4 \equiv$ Incentive Compensation	% of fees	0-8
$X_5 \equiv$ Scope--Geographic (U.S.)	$\left(\frac{\text{Geographic centers adequately covered}}{\text{Centers where relevant work can be generated}}\right)$	25-100%
$X_6 \equiv$ Scope--Geographic (Outside U.S.)	% of U. S. business	0-50
$X_7 \equiv$ Scope--Services Offered	$\left(\frac{\text{No. of disciplines having threshold capability}}{\text{No. of synergistic disciplines required by society}}\right)$	25-100%
$X_8 \equiv$ Relevant Experience (annual increment)	% of fees	0-1
$X_9 \equiv$ Formal Training	No. of degrees per professional staff member	1.5-2.5
$X_{10} \equiv$ Professional Development (excluding relevant experience)	% of fees	0-1

and k is a non-zero scaling constant greater than minus one which can be evaluated from the k_i's.

The task remaining was to assess the component utility functions, assess their scaling factors, and then evaluate the k-value for the multiplicative form.

Assessing the Component Utility Functions. All the ten utility functions were assessed on a zero to one scale using the techniques discussed in Chapter 4. Let us briefly consider those for retained earnings and formal training, attributes X_1 and X_9 in Table 7.7.

The range of retained earnings is zero to eight percent, so since preferences are monotonically increasing, we set

$$u_1(0) = 0 \quad , \qquad u_1(8) = 1 \quad ,$$

where u_1 is the utility function for retained earnings. Next, by checking certainty equivalents for a number of lotteries, it was verified that Dr. Nair was risk averse in terms of retained earnings. It was found that $2 \sim \langle 0,8 \rangle$. $0.75 \sim \langle 0,2 \rangle$, $4 \sim \langle 2,8 \rangle$, $5.5 \sim \langle 4,8 \rangle$, and for a check, that 4 for certain was indifferent to a 0.75 chance at 8 and a 0.25 chance at zero. The utility function consistent with these assessments is shown in Figure 7.10.

The assessment of the utility function for formal training led to some surprises. What was not a surprise

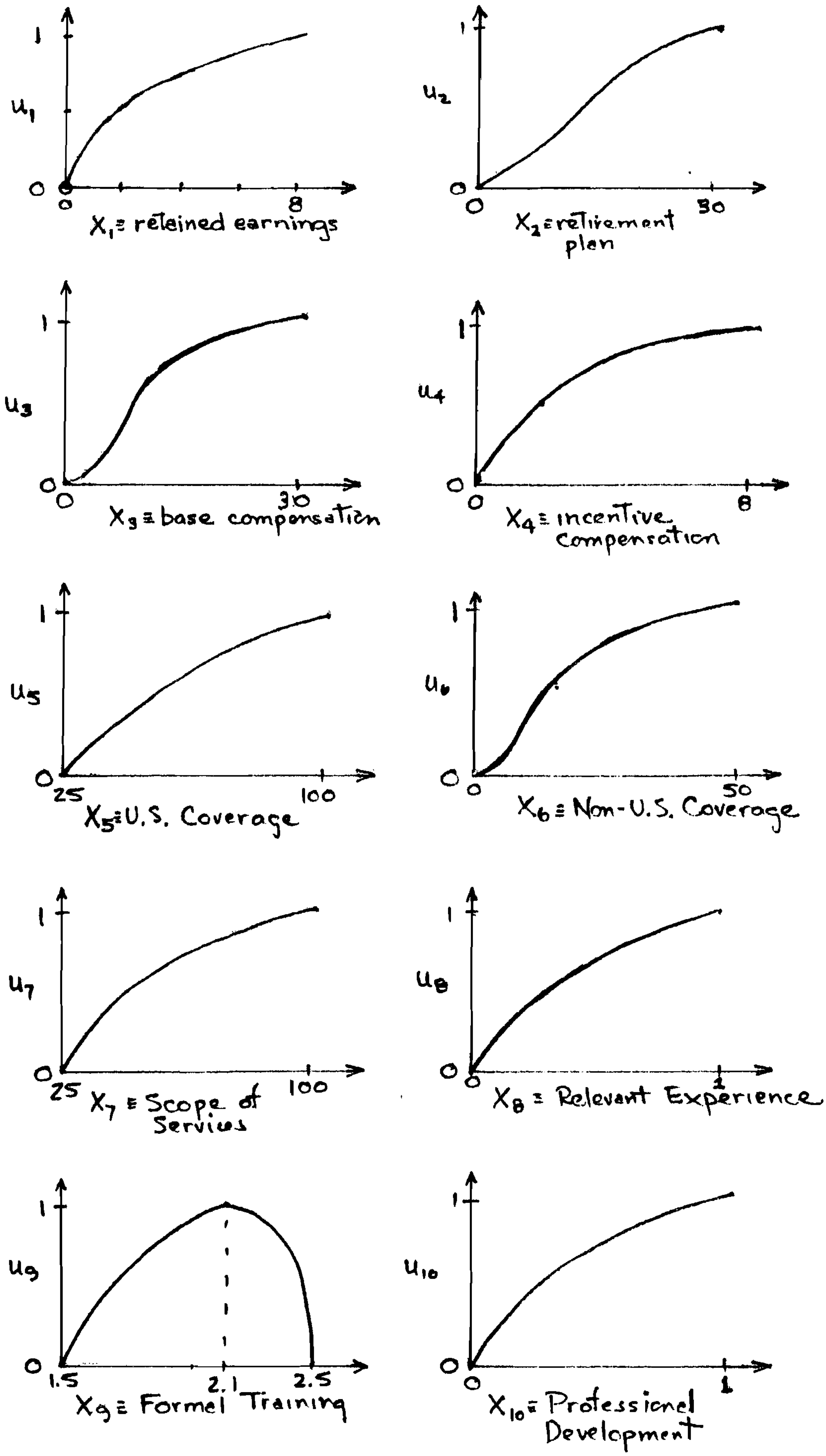


Figure 7.10. Woodward-Clyde's Component Utility Functions

was that preferences for levels of this attribute are not monotonic; they increase up to a maximum point and then decrease. Originally, it was the thought to assess preferences from 1 to 3 degrees per professional staff member. However, once we began this task, it became clear that with levels between 1 and 1.3 and 2.7 and 3, Woodward-Clyde could not exist in a form similar to the present. Hence our viable range was changed from 1.5 to 2.5, which were practical limits for the foreseeable future.

Next, by using the Table 7.6, it became clear that the previously felt optimum level of 2.25 was too high and 2.1 was chosen as an alternative after some consider-ation. It was also felt that the undesirability of 1.5 or 2.5 degrees per professional was about equally as bad so u_9, the utility function for formal training was scaled by

$$u_9(1.5) = u_9(2.5) = 0 \quad , \qquad u_9(2.1) = 1 \quad .$$

Again with the aid of Table 7.6, it was concluded that $1.7 \sim \langle 1.5, 2.1 \rangle$, $1.8 \sim \langle 1.7, 2.1 \rangle$, and $2.3 \sim 1.8$. The resulting utility function is shown in Figure 7.10.

<u>Assessing the Relative Scaling Factors</u>. The ranking of the ten attribute scaling constants of the multiplicative utility function--that is, the k_i's in (7.21)--is given

in Table 7.8. To specify their relative magnitude, Dr. Nair considered the relative desirability of consequences with one attribute at its most preferred level and all other attributes at their worst levels. He decided that the one he would most like to have at its best level was retained earnings. Thus the scaling factor associated with retained earnings is the largest. The attribute he would next prefer to have alone at its most desirable level was formal training so its scaling factor is second largest. Repeating this procedure led to the ranking of the scaling factors indicated in Table 7.8.

To quantitatively establish the relative values of the scaling factors, tradeoffs between pairs of attributes were explicitly assessed. Dr. Nair was asked, for nine pairs of attributes, questions such as:

> "Assume all attributes other than retained earnings and retirement plan are fixed at convenient levels. Now, how high would retained earnings have to be, given the retirement plan is at its lowest level, in order for you to be indifferent between this option and an alternative option with the retirement plan, at its most desirable level of 30 and retained earnings fixed at its lowest level?"

The responses are shown in Table 7.8 in the column labeled "indifference equivalent." Thus if we designate the scaling factor of X_1 as k_1, the scaling factor for

Table 7.8 Evaluating the Scaling Factors in Woodward-Clyde's Utility Function

Attribute	Ranking of Scaling Factor	Range	Indifference Equivalent	Relative Scaling Factor	Scaling Factor
$X_1 \equiv$ retained earnings	1	0-8	—	k_1	.67
$X_2 \equiv$ retirement plan	7	0-30	30 of $X_2 \sim$ 3 of X_1	$k_2 = .66k_1$	.44
$X_3 \equiv$ base compensation	5	0-30	30 of $X_3 \sim$ 4 of X_1	$k_3 = .77k_1$	.517
$X_4 \equiv$ incentive compensation	9	0-8	8 of $X_4 \sim$ 2.5 of X_1	$k_4 = .58k_1$	.391
$X_5 \equiv$ U.S. coverage	6	25-100	100 of $X_5 \sim$ 3.5 of X_1	$k_5 = .72k_1$	.482
$X_6 \equiv$ non-U.S. coverage	10	0-50	50 of $X_6 \sim$ 50 of X_5	$k_6 = .5k_5$	.241
$X_7 \equiv$ scope of services	3	25-100	55 of $X_1 \sim$ 100 of X_5	$k_5 = .75k_7$	.634
$X_8 \equiv$ relevant experience	4	0-1	1 of $X_8 \sim$ 50 of X_5	$k_8 = .5k_5$	.241
$X_9 \equiv$ formal training	2	1.5-2.5	2.1 of $X_9 \sim$ 7 of X_1	$k_9 = .97k_1$	.647
$X_{10} \equiv$ professional development	8	0-1	1 of $X_{10} \sim$ 50 of X_5	$k_{10} = .5k_5$	.241
					4.505

X_2, for instance, must be $.66k_1$ since, using u_1 in Figure 7.10, the utility of a retained earnings of 3 percent is 0.66. This follows since the utility of 3 percent retained earnings, with the growth in retirement plan at its least desirable level, must equal the utility of 30 percent growth in retirement plan, with retained earnings at its minimum level. Because of the preferential independence assumptions, the levels of the attributes other than retained earnings and retirement plan do not matter. The relative values of the scaling constants are also shown in Table 7.8.

Selecting a Utility Function. We felt fairly confident about the relative values of the scaling constants, but to get their absolute magnitudes requires the answer to a difficult question. Dr. Nair was asked:

> "What probability π_1 would you select such that you would be indifferent between option 1 which retained earnings at 8 percent and all other attributes at their least desirable levels and an alternative option 2 consisting of a lottery yielding all attributes at their most desirable level with probability π_1 or otherwise all attributes at their least desirable level?"

Those two options are illustrated in Figure 7.11. Using the "converging method" discussed in Section 4.9, a value of 2/3 for π_1 was selected. This implied that

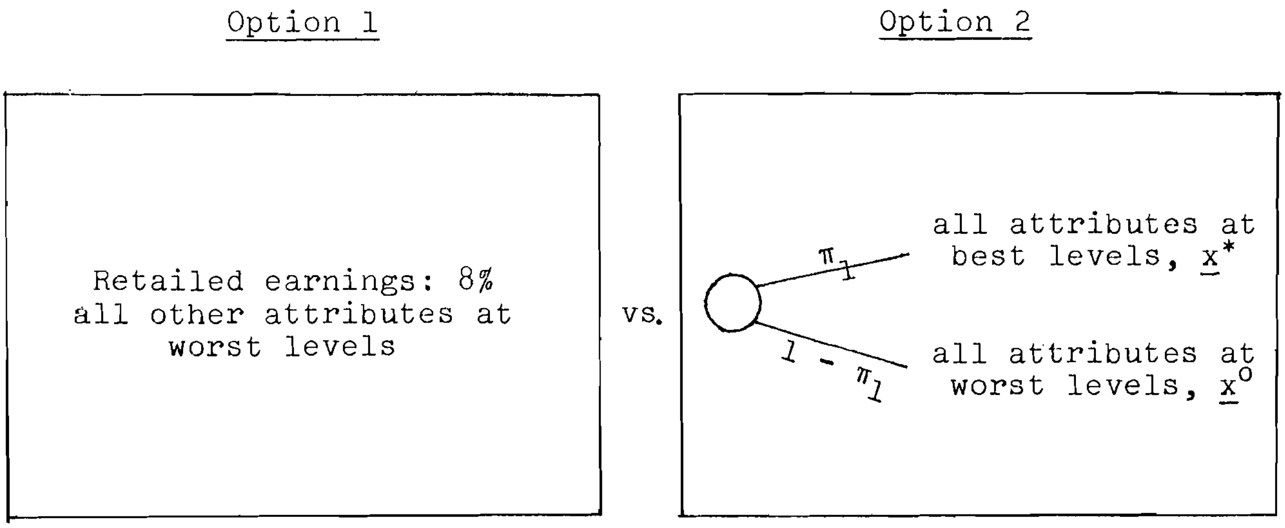


Figure 7.11 Adjust π_1 to get indifference!

the scaling factor k_1 should be 0.67, from which the values of the other scaling factors indicated* in Table 7.8 follow:

Since the sum of the scaling factors is 4.505, we knew the multiplicative utility function (1) was appropriate to express Dr. Nair's preferences. Evaluating (1) for the most desirable consequences one finds

$$1 + k = \prod_{i=1}^{10} (1 + kk_i) \quad , \qquad (7.22)$$

which was solved using the routine of Appendix 6B to yield $k = -.998$. Such a low level for k (it must be greater than -1) indicates a high level of complementarity among preferences for the attributes. It is the general feeling of the Long Range Planning Committee that if retained earnings are at a high level, one can "take care of" the other attributes if proper policies are implemented. However, this feeling weakens as the time frame of reference increases. That is if our attributes represent one-year levels, Woodward-Clyde could stand a bad year with most attributes and make it up in the next year. On the other hand, if the attributes of Table 7.8 designate five-year averages, the desirability of waiting five years to "redistribute" high retained earnings to attributes at their lowest levels is understandably much

*The sensitivity of the analysis to π, is discussed shortly.

less. This situation, which became apparent during the assessment process, is clearly important to recognize in discussions of options affecting the future vitality of Woodward-Clyde. The original preference assessments were made using a one-year period. The results reported here are made using annual averages over a three-year period.*

Sensitivity Analysis. Because of the importance of the probability π_1 assessed to specify k_1, a small sensitivity analysis was made of this parameter using the same relative values of the scaling constants in Table 7.8. Recall that $\underline{x}^*$ defines the consequence with all attributes at their best levels and $\underline{x}^o$ the consequence with all attributes at their worst levels. To assist in examining the implications of the various π_1 values, let us make two definitions:

$\pi' \equiv$ the probability such that a lottery with a π' chance at $\underline{x}^*$ and a $(1 - \pi')$ chance at $\underline{x}^o$ is indifferent a consequence with retained earnings and formal training at their best levels and all other attributes at their worst levels,

---

*For reference, the indifference probability π_1 for the options in Figure 7.11 was 0.75 when a one-year period was considered, whereas it was 0.67 for the three-year period.

$\hat{\pi} \equiv$ the probability such that $\langle \underline{x}^*, \hat{\pi}, \underline{x}^0 \rangle$ is indifferent to the sure consequence with each attribute at its level of 0.5 utility.

The results, which were calculated using the computer program discussed in Appendix 6C, are shown in Table 7.9, where π_1 is first specified. Then, using the relative scaling factors from Table 7.8, the individual k_i's are fixed. Using these, k, π', and $\hat{\pi}$ were calculated. Further reflection and examination of Table 7.9 led Dr. Nair to stay with his original estimate of $\pi_1 = 0.67$ for the three-year period. Thus, the final scaling constants are those shown in the last column of Table 7.8.

Table 7.9 A Sensitivity Analysis of the Scaling Factor k

π_1	Σk_i	k	π'	$\hat{\pi}$
.87	5.86	-.999	.98	.973
.74	4.96	-.999	.925	.947
.67	4.5	-.998	.884	.928
.60	4.06	-.996	.836	.903
.47	3.15	-.979	.714	.835
.34	2.25	-.900	.561	.733

7.4.6 Uses of Woodward-Clyde's Utility Function

Since the original assessments, Dr. Nair has essentially repeated the assessment procedure just described with each of the members of the 1974 Long Range Planning Committee. These assessments included verification of assumptions, assessing single-attribute utility functions, and specifying scaling constants. This resulted in some minor changes to Dr. Nair's utility function (already integrated into the previous subsections) to achieve what may be referred to as a consensus corporate utility function. This obviously does not mean the Board of Woodward-Clyde will blindly make decisions with this utility function. It is being used to facilitate communication among officers of Woodward-Clyde and to help professional intuition.

The assessment process forced individuals to be a bit more precise in deciding why they felt certain levels of specific attributes were important. As previously mentioned, it also served to indicate how tradeoffs among attributes depended on the time frame of reference. The general feeling of those involved in the utility function assessment may be summed up by the comment of one individual, "I've had to make tradeoff decisions like this all my life, but until now the process has always been somewhat fuzzy and left me with the feeling that I didn't completely comprehend all the implications of my subjective judgements. The use of utility theory and explicit tradeoffs helps considerably." With a better understanding of one's own tradeoffs and preferences, it is a small wonder that it becomes easier to communicate these and discuss the issues with one's colleagues.

The process of assessing a utility function has also led to minor, but important, modifications in the overall evaluation process for long-range plans. Some objectives have been deleted or aggregated, and in other cases, several attributes have been altered to better indicate the concerns of Woodward-Clyde. Changing the attribute measure for relevant experience to reflect the yearly increase in experience is one such example.

Since several of the attributes concern distribution of income available (i.e., percent of fees), it is a simple task to use the utility function to help select the best distribution among salaries, retained earnings, incentive compensation, professional development, relevant experience, and contribution to retirement plan. With any fixed percentage of fees available, the technically feasible surface of fee distribution, as well as the distribution with maximum utility, is easily specified.

As before, the component utility functions can still be used to conduct a deficiency analysis by indicating the difference between the present state and a desired state, representing what is technically feasible in a specified time span. A bit more broadly, by calculating the gradient of the utility function in each attribute for the present state position and combining this with subjectively assessed changes in the state of each attribute for an equivalent amount of effort (time and money), one gets an indicator of policies which may be particularly fruitful to pursue.

The utility function discussed here will no doubt go through additional metamorphosis in the future years, as needs and preferences of individuals at Woodward-Clyde adjust to better reflect their position in society, the external environment, and so on. For example, the Pension Reform Act of 1974, because of certain provisions with regard to the ability of Pension and Profit Sharing Plan Trusts to invest in company stock, is likely to alter the present relative value of the attribute "growth in retirement plan" among the attributes. Woodward-Clyde Consultants is presently examining the effect of this and other external changes on the utility functions for the various individual attributes and the tradeoffs between the attributes. This will be a continuing activity.

The current function does overcome the original shortcomings on the 1972 quantification of objectives outlined in Section 7.4.1. It is being used to examine present decisions which effect the future existence of the company. In addition, the Woodward-Clyde objectives hierarchy partially provides an underlying and unifying basis for evaluating long-range plans and operational activities of the affiliated firms. It is not an overstatement to say that several individuals at Woodward-Clyde find the multiattribute utility concept interesting and helpful. Perhaps more importantly, they are enthusiastic about potential future uses. In this regard, partially as a result of the work discussed here, a special group within Woodward-Clyde Consultants has been set up and funded to begin to transfer the concepts and techniques of decision analysis into their professional practice.

As an interesting anecdote, in 1974, Woodward-Clyde Consultants reorganized its operations from that of a holding company subsidiary relationship to an operating company with five regional divisions, each division having geotechnical and environmental capabilities. The more significant reasons given for this reorganization were to better serve its clients in terms of providing integrated geotechnical and environmental capability, establish a one company image for improved marketing, and increase efficiencies by eliminating various subsidiary management structures. In evaluating the desirability of the organizational changes, many members of the Board of Directors made a subjective determination as to whether the changes would increase the companies ability to improve their level of performance over the various attributes. The explicit statement of attributes made it possible to make this evaluation.

7.5 EVALUATING COMPUTER SYSTEMS

How should management select a computer system? How should the management of a computer facility evaluate the quality of its service? When and how should a time-sharing system be altered to provide better service to its users and to attract additional users? These are representative questions facing various participants, including both managers and users in todays computer industry. It seems that responsible answers to such questions require the consideration of a number of factors: availability of the system, its reliability, response times to different requests, costs, as well as many less tangible aspects. These problems are inherently multidimensional.

In this section we will report on some work of Grochow [1972,1973], which deals with such questions using the concepts and methodology discussed earlier in this book. Grochow assessed a three-attribute utility function for users of time-sharing systems. To illustrate the usefulness of such information for decision making by the management of these systems, we first describe what Grochow did and then discuss its relevance to the questions posed at the beginning of this section.

7.5.1 Preferences of Systems Programmers

To begin, Grochow interviewed a number of users of general time-sharing systems to determine their usage patterns and objectives of importance. His subjects were computer system programmers concerned mainly with the input and editing of programs and the compilation and testing of these. Their ratio of editing sessions to compiling and testing sessions was approximately five to one. Four attributes of the system important to this class of users were

(1) Response time to trivial requests, i.e., editing,

(2) Response time to compute-bound requests, i.e., compiling,

(3) Availability,

(4) Reliability.

Grochow assessed utility functions over the first three of these attributes conditional on reliability being at a high level.

Before beginning the assessment process, Grochow discussed the basic ideas of utility theory with each user and presented a scenario indicating the importance of the three attributes and establishing that reliability was at a high level. For measures of effectiveness he used, for the first two attributes the average number of seconds to satisfy requests, and for the third, the percentage of successful log-ins.

By assessing various conditional utility functions over one attribute at a time given that the other two attributes were held fixed, he established the appropriateness of different utility independence conditions and thus, restricted the form of the utility function. Let us define attributes

$X \equiv$ average response time to trivial requests in seconds,

$Y \equiv$ average response time to compute-bound requests in seconds, and

$Z \equiv$ percentage of successful log-ins.

In terms of this notation, the conditions that Grochow verified as appropriate for the class of users under consideration were

(i) X is conditionally utility independent of Y given Z,

(ii) X is conditionally utility independent of Z given Y,

(iii) Y is conditionally utility independent of Z given X.

It follows directly from Theorem 6.17 in subsection 6.11.4 conditions (i) and (ii) imply that

(iv) X is utility independent of $\{Y,Z\}$.

Using Theorem 5.6, from condition (iv), we know

$$u(x,y,z) = u_X(x,y^O,z^O)\ u(x^*,y,z) + [1 - u_X(x,y^O,z^O)]\ u(x^O,y,z) \tag{7.23}$$

where u and u_X are scaled from zero to one with superscripts ^{O} and * indicating respectively the least and most desirable level of an attribute. Then using condition (iii) and the analogous result to Theorem 5.6 for conditional utility functions, we can further break down (7.23) to yield

$$
\begin{aligned}
u(x,y,z) = {} & u_X(x,y^O,z^O)\ \left[u_Y^*(x^*,y,z^O)\ u(x^*,y^*,z) \right.\\
& \left. + \{1 - u_Y^*(x^*,y,z^O)\}\ u(x^*,y^O,z)\right]\\
& + \left[1 - u_X(x,y^O,z^O)\right]\ \left[u_Y^O(x^O,y,z^O)\ u(x^O,y^*,z)\right.\\
& \left. + \{1 - u_Y^O(x^O,y,z^O)\}\ u(x^O,y^O,z)\right] \quad , \qquad (7.24)
\end{aligned}
$$

where u_Y^* and u_Y^O are also scaled from zero to one.

One can note that given these scaling conventions,

$$u_X(x,y^O,z^O) \equiv \frac{u(x,y^O,z^O)}{u(x^*,y^O,z^O)} \qquad (7.25a)$$

$$u_Y^O(x^O,y,z^O) \equiv \frac{u(x^O,y,z^O)}{u(x^O,y^*,z^O)} \qquad (7.25b)$$

$$u_Y^*(x^*,y,z^O) \equiv \frac{u(x^*,y,z^O) - u(x^*,y^O,z^O)}{u(x^*,y^*,z^O) - u(x^*,y^O,z^O)} \ . \qquad (7.25c)$$

If one plugs (7.25) into (7.24) we see that (x,y,z) is completely specified by assessing the seven consistently scaled one-attribute conditional utility functions illustrated by heavy lines in Figure 7.12.

The actual verification of conditions (i), (ii), and (iii) was iterative in nature. Each additional conditional utility function contributes to a better understanding of the overall structure of the utility

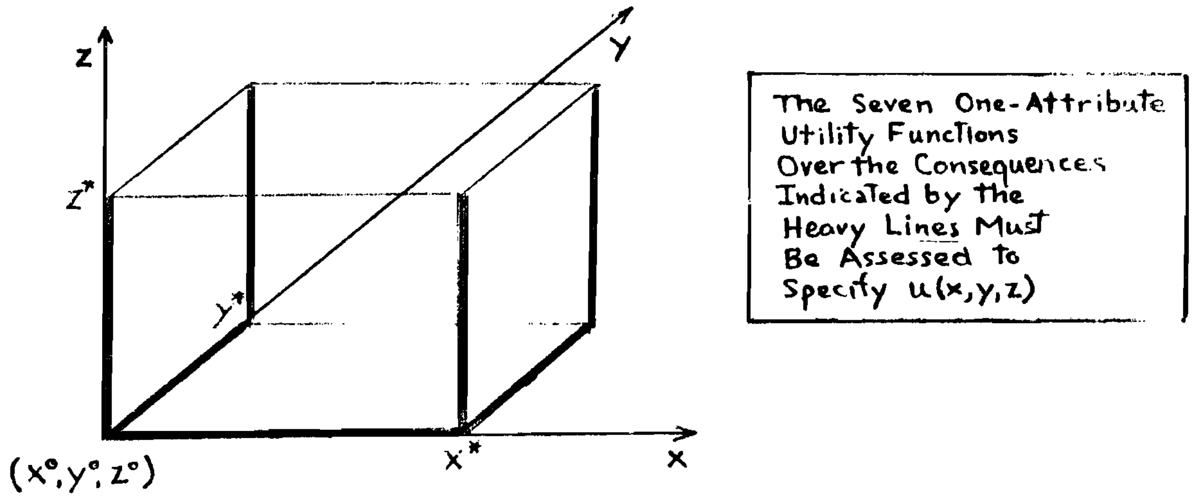


Fig. 7.12 Assessments Required for Evaluating a Utility Function for a Time-Sharing Computer System

function u(x,y,z). The implications of these were discussed with the user throughout the assessment procedure. Whenever there were inconsistencies in the responses of a user, they were pointed out and part of the procedure redone. In all, the utility independence conditions (i), (ii), and (iii) were verified for eight different individuals in the class of users described earlier.

An actual utility function was assessed for only one of these users. The general procedure discussed in Chapters 5 and 6 was used for this purpose. The utility function was assessed over the space $2 \leq x \leq 9$ (seconds), $2 \leq y \leq 120$, and $10 \leq z \leq 100$ (percent).

It turned out that Z was not utility independent of {X,Y} or conditionally utility independent of either X or Y. Grochow states the reason for this: When either response time is at an unfavorable value, for instance, the programmer will be spending most of his or her time contending with the slow response, and consequently will not be as concerned about logging in as when response times are at more desirable levels. The stated reason why Y is not conditionally utility independent of X given Z is that the users may set their relative preferences for response time to compute-bound requests in terms of the response time to trivial requests they are experiencing.

Let us now consider how one might use Grochow's results for making decisions in the computer industry.

Suppose our user was trying to choose among different time sharing facilities which differed not only in terms of X, Y, and Z, but also in terms of their reliability R and their monthly subscription cost S. A proper evaluation here would require a utility function $u'(r,s,x,y,z)$ for the user. However, if $\{X,Y,Z\}$ is utility independent of $\{R,S\}$, then of course, from Theorem 5.6, u' can be expressed as a function of r, s, and u so

$$u'(r,s,x,y,z) = f[r,s,u(x,y,z)] \quad .$$

The original utility function u can be used in a similar fashion if $\{R,S\}$ is utility independent of $\{X,Y,Z\}$ and $\{X,Y,Z\}$ is not utility independent of $\{R,S\}$. Given this assumption, Theorem 5.6 says u' may be expressed as a function of one utility function over $\{R,S\}$ and two utility functions over $\{X,Y,Z\}$ given different levels of $\{R,S\}$. One of these utility functions can be $u(x,y,z)$.

Going one step further, suppose our user (or firm) must decide whether to <u>buy</u> a computer or <u>rent</u> such services. If the choice is made to buy a computer, there may be many options. Clearly, such a decision would involve a time horizon of at least a few years. To remain simple, let us assume that attributes X, Y and Z and a cost attribute are sufficient for the decision. With a five-year horizon, this cost attribute might be $C \equiv \{C_1,C_2,\ldots,C_5\}$ where C_i represents costs in the i^{th} year. Then, as before, with necessary utility independence assumptions between C and {X,Y,Z}, the original utility function u can be used.

Switching gears, suppose the management of a time-sharing service has two objectives: maximize profits and provide the best possible service to customers. A reasonable measure of the quality of service to a user may be its utility function over attributes X, Y, and Z. Hence, given many users, the firm may select a utility function which is a function of annual profits, for instance, and the individual utility functions of its users.

By including potential user's utility functions as arguments of its utility function, the firm may have a tool to help select pricing and service policy. That is, if prices are too high, many users will select competitors and thus reduce the firms profit. If the subscription prices are too low, the firm will also do poorly financially. By maximizing its expected utility, the firm can find the "optimal" price.*

---

*This brief discussion has neglected actions by competitors. The utility functions discussed are applicable in conjunction with game theory, a discipline concerned with these competitive aspects. A basic introduction and survey of game theory is Luce and Raiffa [1957]. A more recent survey is Shakun [1972].

7.6 SITING AND LICENSING OF NUCLEAR POWER FACILITIES

The siting of nuclear power facilities is an extremely complex process. There are many concerned interest groups, each with their own set of multiple objectives, trying to influence the decision making process. The stakes are large, involving hundreds of millions of dollars, possible energy shortages and "blackouts," the possibilities of severe environmental damage, and in some situations, heavy dependency on foreign fuels, to mention a few of the relevant considerations.

In the United States the power company has to prepare its case advocating a particular site or sites and submit these plans for review by governmental regulatory authorites (e.g. power, environmental impact) and by the federal Atomic Energy Commission. These bodies try to reach a decision by weighing the available data, considering the broad tradeoffs, and examining diverse viewpoints: of the power company, of environmentalist groups, of the public as energy consumers, and of local groups, such as the communities near the suggested sites. How can these governmental authorities rationally integrate all the available information in a manner useful for aiding their decision process?

The power companies themselves have difficulties in dealing with the multiple objectives they face. They are, however, mainly concerned with competitive business positions and engineering factors, such as transmission facilities design and network reliability, which directly affect their financial returns. But when a power company is asked, by the regulatory boards, its position on broader questions such as the impacts of its proposals on the environment and local communities, it too must address a broader set of objectives and often their analyses depend on informal and intuitive reasoning. Perhaps with a better understanding and presentation of the fundamental tradeoffs among the conflicting objectives necessitated by each of the alternatives under consideration, the power company might be in better position to select the best alternative in view of its economic objectives, its public responsibility, and the public's requirements. A formal analysis of these considerations may contribute toward reducing the required time of the now lengthy process necessary for approval of nuclear power facilities. The big question is, what are the characteristics of such an analysis and how does one get it done? The literature on this general subject is voluminous but of direct relevance to the techniques and framework introduced in this book, we single out the works by Gros [1974], Papp et al. [1974], Nair et al. [1975], and by Keeney and Nair [1975].

In this section, we do two things:

(1) speculate on the appropriateness of multiattribute utility theory for examining the questions raised in the preceding paragraphs, and

(2) discuss the work of Jacques Gros, who attempts to quantify preferences for nuclear siting problems using results discussed in this book.

The work described below is far from definitive--perhaps it could more appropriately be described as "suggestive research." We do not dwell on important issues, such as: Is the attribute set complete? Whose preferences should be assessed? How does one introduce political relevancies? How does the analysis help (or hinder) conflict resolution? Our purpose is merely to focus on the concepts of the suggestions and to worry little about their pragmatic implementation. Our excuse for speculating on possible uses of a theoretical nature in a so-called "applications" chapter is that we feel the ideas introduced here are important and the framework of analysis may be appropriate to carry out in practice. In this regard, we feel that Gros' accomplishments are encouraging. At the time of this writing, Woodward-Clyde Associates (see Section 7.4) is evaluating the siting and design of nuclear power plants using these same concepts and techniques.

7.6.1 Objectives For Nuclear Power Siting

Each party interested in siting nuclear power facilities will have its own objectives. By and large, however, in each case these objectives might fall under the five categories: environmental, human safety, consumer well-being, economic, and national interest. Let us suppose that the set of objectives listed in Table 7.10 is sufficient for analysis by any of the interested parties, although clearly, there is overlap in this crude list and all of these objectives are not needed by all of the parties. Those objectives of primary interest to the concerned parties are indicated in the table. Also, for future reference the associated attribute--possibly a vector attribute--is designated notationally. No attempt is made to specify specific attributes at this time.

Table 7.10 Some Objectives for Siting Nuclear Power Facilities

Attribute	Category	Objective	Parties Primarily Concerned*
X_1	Environmental	Minimize Pollution	E,L
X_2	Environmental	Provide Aesthetically Pleasing Facilities	E,L
X_3	Human Safety	Minimize Human Health Hazards	E,L,P,S,F
X_4	Consumer Well-Being	Provide Necessary Power	C,E,P,S
X_5	Consumer Well-Being	Minimize Consumer Power Costs	C,S
X_6	Economic	Maximize Economic Benefits to Local Community	L
X_7	Economic	Maximize Utility Company Profits	P
X_8	Economic	Maximize State Revenues	S
X_9	Economic	Improve Balance of Payments	F
X_{10}	National Interest	Reduce Dependency on Foreign Fuels	F

*C ≡ consumers; E ≡ environmentalists; L ≡ local communities; P ≡ power company; S ≡ state agency; F ≡ federal agency.

7.6.2 A Conceptual Framework for Analyses by the Interested Parties

The utility functions briefly discussed here are mainly to suggest a conceptual framework for thinking about crucial preference aspects of the nuclear power siting problem and for communicating these preferences to other interested parties. For brevity, we skip a discussion of the utility functions of the consumers, environmentalists, and local community interests. These are, in theory, more straight forward than the cases we do consider.

The Power Company's Point of View. One might simply say that a power company is concerned only with maximizing its own profits. If such were the case, it would be appropriate to assess the company's utility function $u_P(x_7)$ over attribute X_7 and use this in evaluating the power company's alternatives. However, in this era of broader corporate interest and responsibility, it is more likely the case that the company is also interested in satisfying its consumers preferences for energy, minimizing the detrimental environmental impact of its facilities, and maximizing the net benefits of its facilities on local communities in which facilities are to be built. Let us designate attributes for these three additional objectives as U_C, U_E, and U_L, respectively, and note that they can be measured by the respective utility functions u_C, u_E, and u_L. The power company, at least informally, is concerned with its utility function $u_P(x_7,u_C,u_E,u_L)$ over the four attributes x_7, u_C, u_E, and u_L, in order to analyze which of its possible options is most attractive to pursue. Conceptually, one might define utility u_L to be a function of $u_1,\ldots,u_\ell,\ldots,u_N$, where u_ℓ is the ℓ^{th} community's utility function and N communities are considered as possible sites. The power company must weight its subjective judgments about the relative desirability that community 1 has for proposed plant A against the relative desirability that community 2 has for proposed plant B. Such tradeoffs, although terribly difficult, must be formally or informally addressed by the power company.

The State Agency's Point of View. Let us oversimplify once again and assume there is only one state agency concerned with licensing nuclear power facilities, whose main responsibility is nuclear safety. Thus, the objectives of the agency might be to minimize danger due to nuclear radiation, to provide state revenue, and to satisfy the interested groups. Attributes X_3 and X_8 from Table 7.11 may be useful for measuring the first two objectives, whereas u_C, u_E, u_L, and u_P might do for indicating interest group satisfaction. Thus, the state agencies preferences might be conceptualized by $u_S(x_3,x_8,u_C,u_E,u_L',u_P)$, where u_L' is the state agencies aggregation of the N communities' utility functions.

The Federal Agencies Point of View. The main federal agency concerned with nuclear power plants in the United States is the Atomic Energy Commission. Its problem is quite similar to that of the state agency just outlined. The major difference might be the federal concern for the balance of payments, indicated by attribute x_9, and the national dependency on foreign fuels, measured by attribute x_{10}. It may be useful for the federal agency to conceptualize its preferences with the utility function $u_F(x_3, x_9, x_{10}, u_C, u_E, u_L'', u_P)$, where u_L'' measures the federal agencies concern for the local community impact of nuclear facilities.

7.6.3 Empirical Assessments of Gros

Gros [1974] studies nuclear facility siting from a slightly different viewpoint and in the process has generated evidence that the utility functions postulated in the proceeding section can be meaningfully assessed.* Specifically Gros investigates the usefulness of what he refers to as Paretian environmental analysis in nuclear siting decisions. Generally stated, Paretian analysis attempts to identify the benefits accruing to each of the

---

*Ovi's [1973] results using multiattribute decision analysis for evaluating nuclear versus fossil power plant alternatives, and nuclear siting and decision studies in progress at Woodward-Clyde Associates also lend support to this contention.

various parties involved in a decision making process and to illuminate the tradeoffs among these groups. To illustrate his approach, Gros examines the deployment of 1000 megawatt nuclear baseload units to possible sites along the New England coast.

In the terminology of this book, Gros assessed multi-attribute utility functions for four parties involved in nuclear power plant siting in New England: power companies, environmentalists, regulatory agencies, and local groups. These utility functions were each assessed over four attributes:

$Y_1 \equiv$ Capacity at a site, measured by the number of 1000 megawatt units at a coastal site,

$Y_2 \equiv$ Incremental dollar costs, measured by the cost of thermal abatement equipment plus transmission costs expressed as a percent of the minimum cost facility,

$Y_3 \equiv$ Radiation hazard, measured by the population within fifteen miles of the nuclear facility times the number of units at the site, and

$Y_4 \equiv$ Thermal pollution level, measured in degrees Fahrenheit at the outfall of the nuclear facility.

These attributes were generated after interviewing a number of individuals who had previously participated in siting controversies. Rather than focusing on questions of whether or not this set of attributes is appropriate for the problem considered, let us consider the assessment procedure.

For each of the four interest groups, a knowledgeable observer, who had an intimate knowledge of many of the group members preferences, was chosen based on recommendations of group members. The knowledgeable observer's utility function was assessed and his preferences were used as those of the appropriate representative group. The results were verified for reasonableness with other group members. For each of the four utility functions, necessary utility independence conditions were verified to invoke Theorem 6.1 implying the appropriateness of either the multiplicative or additive utility functions.

Gros was also interested in preferences over the forty year design horizon. For each of the knowledgeable observers he verified that preferences for lotteries in any individual year were utility independent of preferences for lotteries over the other years. Also he found preferences in each pair of years to be preferentially independent of preferences in other years. Hence the forty-attribute utility function, representing the forty-year period, was again either multiplicative or additive. Because of the desire on the part of the knowledgeable observers to spread risks over the years, the multiplicative form was selected as appropriate.*

---

*Some interesting assessments of preferences over time indicating some of these issues are found in Chapter 9.

Gros' efforts and empirical assessments help to illustrate something that we firmly believe. Namely it is possible to develop meaningful utility functions, such as those postulated in the preceding subsection, for the various participants in the complex decision processes concerning the siting and licensing of nuclear power facilities. The assessments briefly discussed here are an important first step toward characterizing utility functions directly useful in making nuclear power siting decisions. The task is difficult and the effort required to obtain these preferences is substantial. However, to avoid these problems relegates the crucial tradeoff issues and the preference evaluation of the risks involved to informal analysis.

7.7 OTHER APPLICATIONS

Experience with formal quantification of preferences in multiattribute contexts is growing. Let us briefly mention a number of decision problems, in addition to those in earlier sections in this chapter, where the concepts of Chapters 2 through 6 were utilized.

7.7.1 The Safety of Landing Aircraft

The safety of landing an aircraft depends on many factors: wind, visibility, ceiling, other aircraft in the vicinity, etc. Yntema and Klem [1965] attempted to quantify the safety of various situations which differed in terms of ceiling, visibility, and amount of fuel that would remain at touchdown given a normal landing. Other relevant factors were fixed at a standard value.

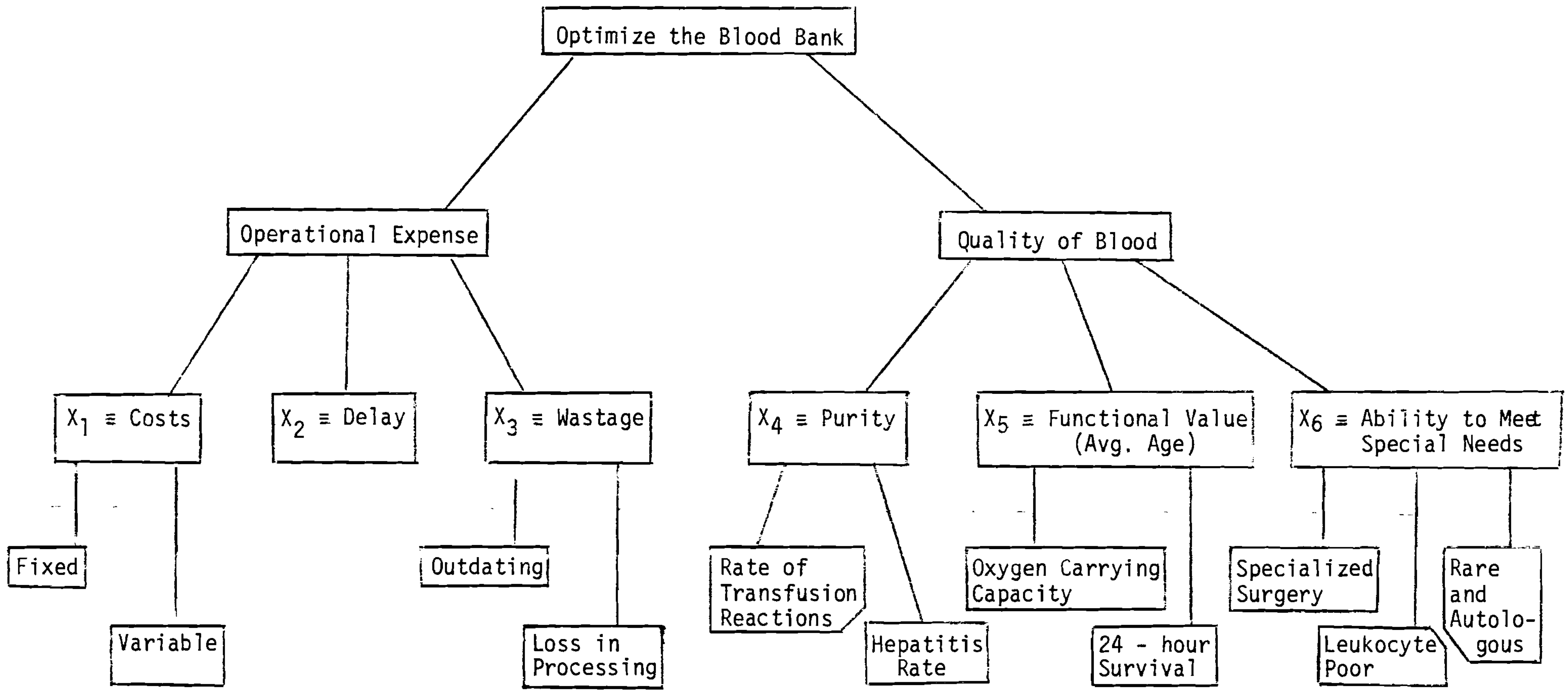
The decision makers for this study consisted of twenty Air Force pilots, each of whom had a good deal of experience in landing aircraft under a wide variety of situations. Using the form of the three-attribute quasi-additive utility function discussed in Result 2 of Section 6.2, utility functions over the attributes ceiling, visibility, and remaining fuel were assessed. In the attribute space, ceiling varied from 100 to 5,000 feet, visibility from 0.25 to 5 miles, and remaining fuel from 15 to 250 gallons. Each decision maker was also presented with forty pairs of consequences and asked to pick the preferable one of each pair. These responses were compared with the implications of each decision maker's utility function. Yntema and Klem concluded "The results were satisfactory."

It should be pointed out that the utility independence assumptions requisite for Result 2 of Section 6.2 were not empirically verified. In fact, the assessments of Yntema and Klem were completed a few years before the formal theory was developed. In spite of this, the resulting utility functions did seem appropriate to represent the preferences of the pilots. Yntema and Klem's pioneering effort gave some support to the contention that it was reasonable and practical to quantify preferences in multiattribute situations.

7.7.2 Strategic and Operational Policy Concerning Frozen Blood

Should a hospital blood bank or system of blood banks invest in expensive blood freezing equipment? And for systems with such capabilities, what are the most desirable proportions of frozen and non-frozen blood? These questions were addressed in a thesis by Bodily [1974]. He also conducted a preliminary investigation of national strategies in blood research and in the usage of frozen blood.

First, after considerable consultation with blood bankers, objectives and measures of effectiveness were specified for evaluating frozen blood issues. The resulting list, given in Table 7.11, indicates the depth at which preferences and probabilities were initially going to be assessed. However, to help the respondent's thinking about the implications of various levels of the attributes, the objectives hierarchy was developed and qualitatively extended as illustrated in Figure 7.13.



A possible configuration for aggregation of attributes for a frozen blood decision problem.

Figure 7.13

Table 7.11 Objectives of a Hospital Blood Bank

Objectives	Measures of Effectiveness
Meet all requests for blood	Average delay or frequency of delay above some acceptable cutoff
Provide high quality blood	Average age at transfusion
Minimize disease	Rate of hepatitis
Minimize cost	Cost/unit
Minimize transfusion reactions	Rate of transfusion reaction
Minimize wastage	Outdating plus processing loss
Provide bloods for special uses	Fraction of special needs that are met with frozen blood or an equivalent unit.

In analyzing the problem of what proportion of frozen blood should be selected for a particular blood bank and the problem of whether or not such a bank should invest in blood freezing equipment, Bodily used a variety of approaches to obtain reasonable probability distributions over the attributes for each alternative. These approaches included utilizing empirical information from blood banks, projections using simple mathematical models of the operations of such blood banks, judgmental estimation of experts, and statistical data available in journal articles, etc.

Bodily tried to assess preferences over the six attributes labelled X_i, $i = 1,2,\ldots,6$ in Figure 7.13 for a number of individuals concerned with blood banking. A first conclusion was that attribute X_6 could be eliminated in considering the question of proportion of blood to freeze. The reasoning was that if a blood bank froze any blood, it would certainly freeze enough to satisfy special needs, and so, the objective "meet special needs" would be equally satisfied with all the viable alternatives. Hence, it could be dropped from the list.

Next attributes X_1, X_2, and X_3 were aggregated since each pair of these was preferentially independent of its complement and substitution rates were constants. Wastage and delay were translated into economic terms using a simple additive value function

$$v(x_1,x_2,x_3) = x_1 + dx_2 + wx_3 \quad ,$$

where d is the equivalent cost per unit of blood delayed and w is the cost per unit wastage. If attribute Y is defined as $X_1 + dX_2 + wX_3$, then what is needed is a utility function $u(y,x_4,x_5)$ over Y, X_4, and X_5.

In the assessment process, it became clear that blood bankers considered the possible range of the average age transfused much less important than the ranges of economic and purity considerations. Hence X_5 was dropped and utility functions $u(y,x_4)$ were completely assessed for one blood banker and one individual with a public health graduate degree and a knowledge of decision analysis. In both cases Y and X_4 were mutually utility independent and so, from Theorem 5.2, the quasi-additive utility function was appropriate. In addition, Bodily ascertained that in a paired comparison of two simple lotteries with identical marginal probability distributions, the blood banker was indifferent. Thus from Theorem 5.4, it follows that the respondent's utility function was additive. Details of these assessments are found in Bodily [1974].

Many of the concepts of Chapters 2 through 6 were explicitly used in the overall assessment process. First, a first-cut hierarchy of objectives was articulated as discussed in Chapter 2, and one objective was then dropped since it was not important enough to influence decisions. Then using preferential independence conditions and the concepts of Chapter 3, a value function over three of the attributes was specified to achieve an aggregation and reduction of dimensionality. Next quantitative considerations led to the exclusion of attribute X_5. Finally, utility independence and the unidimensional assessment techniques surveyed in Chapter 4 were used to specify the final utility functions.

This case illustrates well a typical evolutionary process which starts from a listing of objectives--in this case the specification of subobjectives extended further down the hierarchy than the quantitative analysis--and terminates with the quantification of the final utility function.

7.7.3 Sewage Sludge Disposal in the Metropolitan Boston Area

In Boston, the Metropolitan District Commission (MDC) has responsibility for water and sewage works for the forty-three cities and towns within its jurisdiction. As of 1971, one hundred tons of sewage sludge was being discharged daily into Boston Harbor by the treatment plants of the MDC. Because of increasing public concern and the interest of the U.S. Environmental Protection Agency, the Division of Water Pollution Control of the Massachusetts Department of Natural Resources requested the MDC to make a comprehensive study of new and better alternatives to the present sludge disposal practice. The MDC organized a committee named the Boston Harbor Pollution Task Force (BHPTF) to study the problem and make recommendations. At the suggestion of the Massachusetts Office of Environmental Affairs and with the consent of the BHPTF, Dennis Horgan, then a graduate student at M.I.T., worked with this task force and conducted an independent decision analysis of the sludge disposal alternatives. This subsection briefly surveys Horgan's work.

The viable alternatives for Boston sludge disposal could be categorized as being either marine disposal or land disposal. In the former category one could either barge sludge to a dumping ground ten miles offshore or extend a submerged sludge line approximately seven miles out to sea. In the latter category one could directly spread the sludge on available land and till it into the soil, or alternatively, the sludge could first be incinerated--thus reducing its volume approximately seventy percent and then disposed of at a land site. There are variations of these four basic alternatives, such as different processes of incineration, etc., but, these were felt to be second-order considerations and not explicitly considered in the analysis.

The analysis by Horgan specified four major objectives: minimize costs, minimize water pollution, minimize land pollution, and minimize air pollution. Thus, the classic question concerning tradeoffs of one kind of pollution against another was explicitly addressed. The net present value of costs was used as the measure of effectiveness of the cost objective. Air pollution was measured in tons of particulate matter and gases due to sludge incineration and land pollution was measured in terms of the total area required for sludge disposal sites. To indicate water quality, Horgan defined a subjective index, as discussed in Section 2.3, scaled from zero to ten, based on state water quality standards.

Exploiting probabilistic independence properties where appropriate, probability distributions were specified over the four variables for each of the four basic alternatives.* Concerning preferences, Horgan verified with members of the BHPTF that each of the four attributes was utility independent of its respective complement, and also, that pairs of attributes were not preferentially independent of their complements. Hence, by Theorem 6.3, the multilinear utility function was appropriate. The specific utility function and probability assessments, as well as sensitivity analysis of the results, are found in Horgan [1972].

7.7.4 Selecting a Job or Profession

A critical decision facing each of us from time to time concerns the selection of a job. This problem is different in one important respect from many of the other illustrations in this book in that it is essentially a personal decision. Most of the other problems dealt with a decision maker as representative of his company or as representative of a branch of the government. Here we will briefly summarize two philosophical approaches to job selection, both of which utilize the general ideas discussed in earlier chapters. The works of Miller [1966,1970] and Teweles [1972] will serve as models for our discussion.

---

*For an incineration alternative, air pollution and land pollution, for example, were not probabilistically independent, since they both depend on the volume of sludge, Horgan's model explicitly included such dependencies.

Miller developed and tested a procedure for evaluating the "worth" of various situations described by multiple attributes. One of the problems to which it has been applied involved a graduate student faced with numerous employment offers immediately following graduation. After preliminary analysis, this number was reduced to four viable contenders. The objectives hierarchy and attributes associated with each of the lowest-level objectives which were identified by the graduate student are illustrated in Figure 7.14.

An additive "worth" function,

$$w(x) = \sum_i k_i w_i(x_i) \quad , \tag{7.26}$$

where w_i measures the worth of an amount x_i of attribute X_i, was used to evaluate the alternatives on a zero to one scale. The scaling factors k_i were determined using conditional assessments as described in Section 3.7. For instance, first weights of 0.33, 0.17, 0.17, and 0.33 were assigned to monetary compensation, geographical location, travel requirements, and nature of work, respectively. Then, for instance, of the monetary compensation, a 0.7 weight went to immediate compensation and 0.3 to future compensation. Of the future compensation, 0.65 and 0.35 went to anticipated three-year salary and anticipated five-year salary respectively. Then, the total effective weight assigned to anticipated three-year

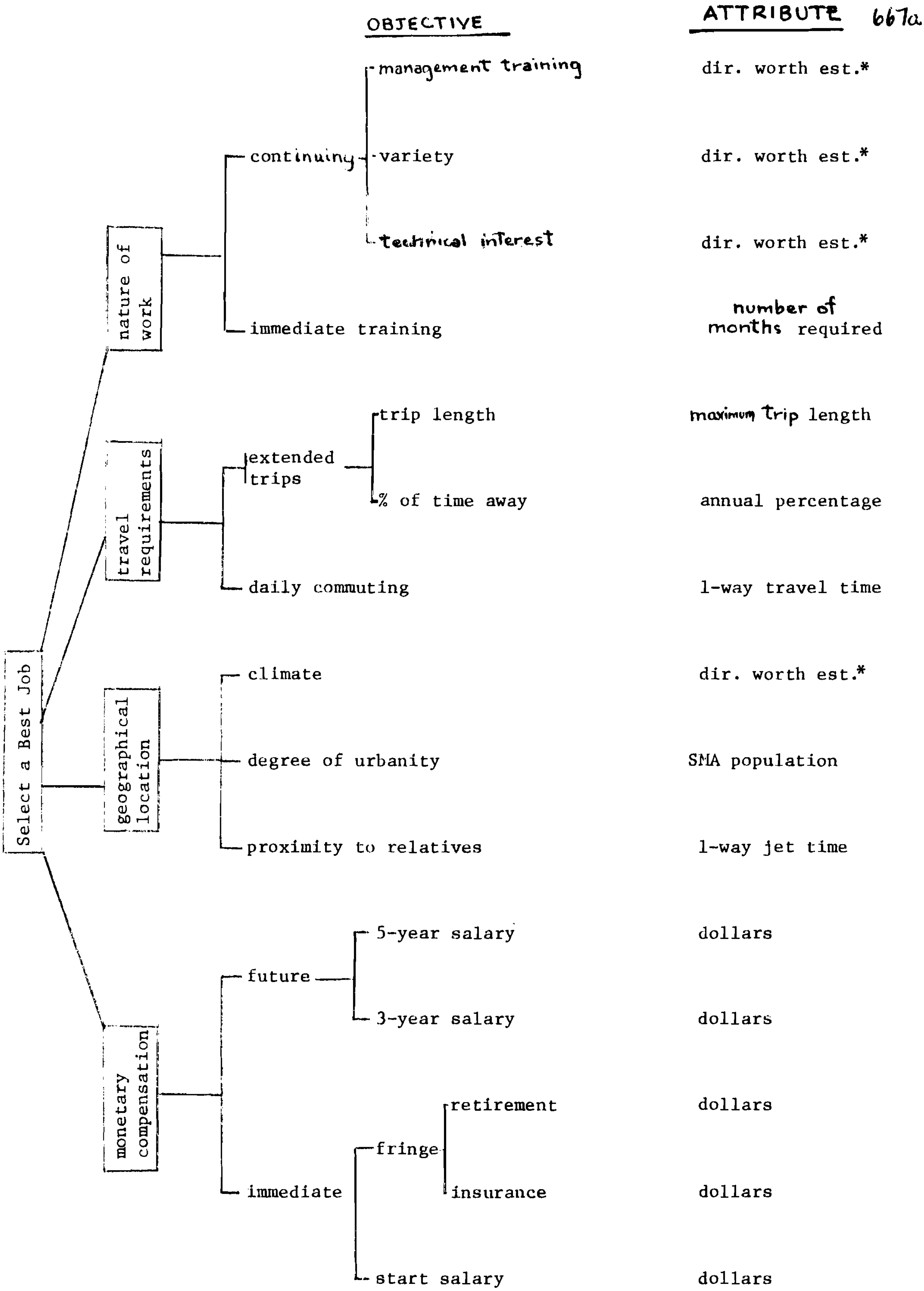


Fig. 7.14. Miller's Objectives for Evaluating Employment Options

*Direct worth estimate: a subjective assessment.

salary was (0.33) (0.3) (0.65) or 0.064. These effective weights were then adjusted to account for the degree to which the attribute actually measured to achievement on an objective. Finally for each of the fifteen attributes, individual worth functions w_i were determined.

The four alternatives were then each represented as a fifteen-attribute vector and the worth of each calculated using (7.26). Uncertainties were not explicitly considered in the problem.

Notice that all the attributes in Miller's problem are in some sense proxy attributes. Presumably, they are proxy for the quality of the decision maker's life. Because of this it was possible to identify many objective measures for these attributes.

Teweles' approach was very different in this respect. He attempted to establish a more direct set of attributes to indicate the desirability of various alternative careers open. Teweles' objectives are given in Table 7.12 along with a short description of the meaning of each.

Table 7.12 Teweles' Objectives for Evaluating Professions

Job Satisfaction--enjoyment derived from doing the type of work you have chosen. Direct benefits of a job such as the opportunity for travel, meeting interesting people, and means of self-expression are included in this factor.

Wealth--the financial remuneration which can be expected from working and the accumulation of capital which can be earned from investment of excess funds. As money is, in a sense, a means of obtaining other goods and services the utility of these products can be substituted for wealth in determining its value.

Security--a condition of relative safety which results from being able to continue your job if you wish to do so. Also included in this factor is the risk to one's health associated with a particularly dangerous occupation.

Family considerations--this factor is an amalgamation of the possible influence a particular career might have on the other members of your family. A wife's attitude, mother's sentiment, child's future, or other considerations should be accounted for in career planning.

Independence--refers to the ability one has of being his own boss and scheduling his own activities. Independence also refers to the short-term flexibility to do what is most important to the individual at a particular time.

Self-esteem--is the self-respect one gains from his own achievements. The self-esteem one could anticipate from a job is very dependent on his ability to be successful at his work.

Prestige--the reputation one acquires within a group as the result of competence, character, power, wealth, etc. The professional respect of one's colleagues may be an important factor to some individuals.

For each of the objectives, except wealth, a subjective index was defined, ranging from zero to one hundred, which was used to indicate the degree to which the corresponding objective was achieved.*

The job alternatives evaluated by Dr. Teweles+ were (1) a private general dentist, (2) a military dentist,

---

*Miller's and Teweles' work illustrates a tendency mentioned in Chapter 2. Namely, as the attributes become more direct indicators of fundamental preferences--as opposed to proxy attributes--it is more difficult to identify suitable objective measures, and one must define subjective indices.

+Dr. Teweles is a dentist, and at the time he wrote his paper, he was completing a Masters of Science in Business Administration and reaching the end of his initial military commitment.

(3) an orthodontic dental specialist, (4) an investment analyst, and (5) a management consultant. These five occupations were evaluated using an additive utility function. Using available data on various professions in addition to personal judgment, Teweles was able to assess probabilities about the degree to which each objective would be met conditional on each alternative. Expected utilities were calculated for each alternative and sensitivity analyses performed.

In Dr. Teweles' report, he states, "The major difficulty in all career planning decisions is for an individual to gain sufficient insight into his own future goals and then learn enough about each alternative to evaluate it objectively." Among Dr. Teweles' conclusions are "As a result of my career analysis, I feel more capable of making the proper career decision at this time. There is no doubt that I understand the factors which motivate me a little better than I did before the analysis." The authors know of many eases where similar personal analyses have been conducted. Some of these resulted in similar conclusions as Dr. Teweles'; other self-analyses, as you might expect, were abortive and useless. We also know of one medical doctor who used this personal self-evaluation technique on a mental patient in a hospital and he reported a surprising success. This doctor took our vernacular phrase, "a framework, for straightening out one's mind," quite literally.

7.7.5 Transporting Hazardous Substances

During the past decade there has been a large growth in the type and the amount of hazardous materials transported within the United States. Shipment of such materials is achieved via all ground modes--rail, highway, water and pipeline. Private citizens, industry, and governmental agencies have become increasingly concerned about the risks associated with transporting these hazardous materials. Aspects of the risk might be divided into two factors:

(1) the likelihoods of various accidents occurring, and

(2) the damage caused by an accident which does occur.

Too often, one has a tendency to assume that "reducing the risks" can always be accomplished by reducing the probability of an accident occurring.* However, one must clearly also include the possible consequences when attempting to reduce risk. Said another way, the <u>risk</u> of the circumstance: "There is one chance in 1,000,000 that a gas leak will lead to a moderate-sized explosion in a populated area next year" seems much greater intuitively than the circumstance: "There are 4 chances in 1,000,000

---

*One can investigate "fail-safe" as well as "safe-fail" techniques.

that a gas leak will lead to a large explosion in the desert next year."

Some pioneering work of Brooks and Kalelkar at Arthur D. Little is currently attempting to measure the relative undesirability of the consequences of various accidents which may result from transporting hazardous materials. In addition, they are investigating which modes of transport are safer for which specific substances.

The aspect of Brooks and Kalelkar's efforts of most interest here concerns their attempts to assess a three attribute utility function over the attributes: human deaths, property damage, and environmental damage. The first attribute ranged from zero to 1200, and the second attribute ranged from zero to ten million dollars. The third attribute was measured by a subjective index scaled from 1 to 13, as defined in Table 7.13.

Table 7.13. Environmental Effects from Hazardous Chemical Spills

Note: This scale applies equally well to water and to land.

1. No effect.
2. Residual surface accumulation of harmless material such as sugar or grain.
3. Aesthetic pollution (odor-vapors).
4. Residual surface accumulation of removable material such as oil.
5. Persistent leaf damage (spotting, discoloration) but foliage remains edible for wildlife.
6. Persistent leaf damage (loss of foliage) but new growth in following year.
7. Foliage remains poisonous to animals (indirect cause of some death upon ingestion).
8. Animals become more susceptible to predators because of direct exposure to chemicals and a resulting physical debilitation.
9. Death to most smaller animals.
10. Short term (one season) loss of foliage with emigration of specific animals that eat the foliage. Eventual reforestation.
11. Death to foliage and emigration of animals.
12. Death to foliage and animals.
13. Sterilization of total environment with no potential for reforestation or immigration of species.

The person whose preferences were assessed by Brooks and Kalelkar was an experienced worker in the field of safety who attempted to take the viewpoint of society as a whole in indicating preferences. It was verified that each of the single attributes were utility independent of the remaining two. Hence, Theorem 6.3 held and the three one-attribute utility functions and the requisite scaling constants necessary for specifying the three-attribute utility function were assessed. The three utility functions are illustrated in Figure 7.15. Details of these assessments are found in Kalelkar et al. [1974].

This analysis raises deep ethical concerns and should be examined critically and constructively by analysts concerned with such problems. At least Kalelkar articulates a utility structure that others can criticize and this is a step forward. Pious, vacuous rhetoric does not help in making such horrendous tradeoffs. We feel that in cases such as the one examined by Kalelkar, implicitly used value and utility structures should be of public concern and should not be suppressed.

7.7.6 Treatment for Cleft Lip and Cleft Palate*

Cleft lip and cleft palate is the second most common congenital deformity in the United States. Treatment for

---

*Roughly speaking, a cleft lip is a failure of the upper lip to grow together. It usually results in a gap in the lip approximately below one nostril. Cleft palate refers to a split in the palate at birth.

this condition is very involved: it requires many different medical specialists, coordinating from birth to adulthood, not only to correct surgically the physiologic defect, but also to address the child's psychological, social, and mental development. The effects of the treatment of clefts and the effects of the clefts themselves are not completely distinguishable. Both are serious and should be considered in selecting an approach for treatment. With this, a critical issue surfaces, namely: what is the best procedure for treatment in a given situation? Value judgments are essential to answer this question, but because survival of the child is not a factor, various concerned individuals--parents and professionals--often disagree more in their value structures in this situation than in cases where survival is an issue. The best treatment should depend on a number of characteristics, such as the physical features of the child after treatment, the cost, the effects on hearing and speech, etc. Pathbreaking results of Jeffrey Krischer [1974] of Harvard University constitute a very interesting attempt to address some of the critical value issues concerning treatment of cleft lip and cleft palate. Here, we briefly describe his work.

In discussing the importance of cleft lip and cleft palate, Krischer states, " Rarely are there defects so handicapping to the child or so disturbing to the family, yet so amenable to treatment." One major objective of treatment is to correct the physical deformities and provide a normal-looking lip and nose. There are usually uncertainties about the surgical success one will have in this process and there is always the possibility of resulting scars. Defective speech often accompanies those with cleft palate, which can be attributed to both physical and psychological factors. Another complication is the possibility of hearing loss due to a variety of factors. Thus, clearly two other important objectives of treatment are to improve future speech skills and to improve hearing.

Krischer has quantified the preferences of over one hundred people, including surgeons, orthodontists, speech therapists, audiologists, pediatricians, and parents of children with clefts, all of whom are actively involved with individuals having clefts. The four objectives and associated attributes which he explicitly considered are given in Table 7.14 along with the range of these attributes. One unique aspect of these assessments was the attribute evaluating physical effects. Krischer had segments of children's faces showing the nose and mouth area superimposed on a sketched face of a child. These pictures illustrated various degrees of physical deformity after treatment for the cleft. The individuals were asked to assess subjectively their preferences for these pictoral displays. Also note that the hearing attribute only had two values. This, of course, could be generalized. For speech, word intelligibility was measured as the percent

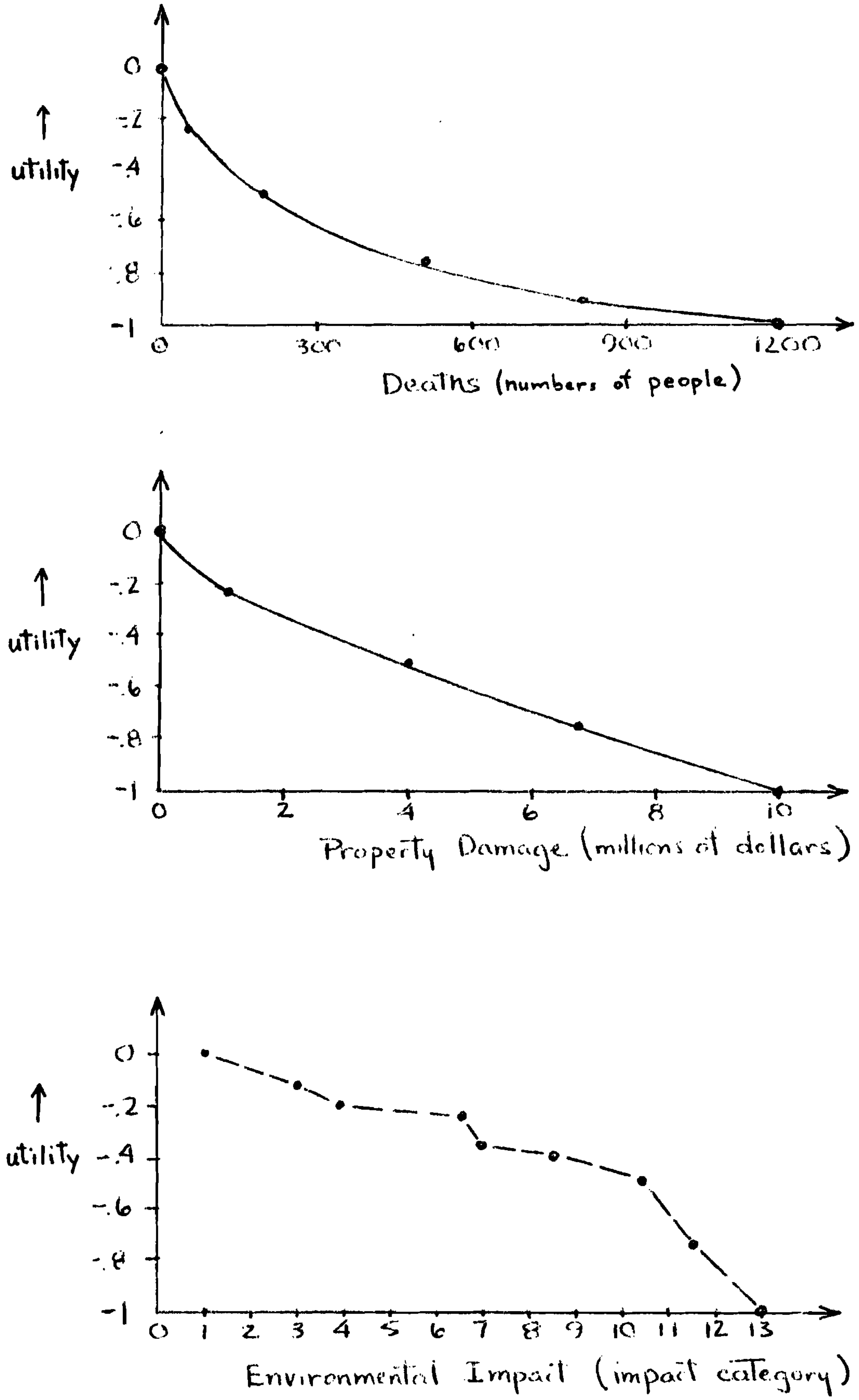


Fig. 7.15. Utility Functions for Evaluating the Transportation of Hazardous Substances

of words accurately identified by a group of listeners with normal hearing. Here 90% is completely adequate, 75% causes mild difficulty in understanding, 50% requires frequent repetition, and 35% is unintelligible.

Once these objectives and attributes were specified, Krischer, working with medical specialists concerned with clefts, developed a questionnaire to assess preferences over the four attributes. This was mailed to medical specialists at numerous cleft-lip and cleft-palate treatment facilities in the United States and through these facilities to some parents of children with clefts. Part of the questionnaire concerned utility independence assumptions and the conditional utility functions for the four attributes and another part concerned preferential independence assumptions and tradeoffs among attributes. Of the first one hundred twenty-five responses, approximately seventy-five percent appear to have accepted requisite assumptions to invoke Theorem 6.1 in formalizing preferences. Details of these assessments, a copy of the questionnaire, and an interesting discussion of individual differences of preferences are found in Krischer [1974].

Table 7.14. Krischer's Objectives for Evaluating Cleft Lip and Palate Treatment

Objective	Attribute	Range
Provide normal looking lip and nose	Pictoral	(see text)
Improve speech	Percent word intelligibility	35 to 90
Improve hearing	Hearing aid required	yes or no
Minimize treatment costs	Dollars	0 - 10,000

7.7.7 Development of Water Quality Indices

Recent work by O'Connor [1973] illustrates some important considerations relevant to specifying and using social indices. O'Connor utilized a modified Delphi procedure (Dalkey [1969]) to combine the judgments of several experts in constructing two separate indices of water quality. One concerned the quality of water to be used as a public water supply and the other described the quality of water for sustaining fish and wildlife populations. Eight experts* were used to (1) specify attributes that should be included in each of the water quality indices and to (2) prescribe a value function over these attributes that would indicate water quality. Since these indices are value functions, they have the property that higher values indicate higher water quality. However, it is not necessarily appropriate to use the expected value of these indices in making decisions when uncertainty is involved.

O'Connor sent questionnaires to and personally visited each of the experts to discuss the attributes which should be explicitly included in some aggregate water quality index and the form of this aggregation function. An additive model was chosen for both the public water supply and fish and wildlife indices. O'Connor emphasizes that an additive model is not appropriate for instance when certain toxic substances enter the water at an unacceptable level or when some of the other attributes, such as pH, reach extreme levels. Thus O'Connor's models are meant to be valid subject to

---

*O'Connor describes the experts as follows: "Eight experts were chosen from an initial set of 20 contacted. Two experts were high-ranking members of The Environmental Protection Agency. Two members were heads of state engineering services departments, and four were university professors in the areas concerned with environmental quality."

the condition that toxic substances are under recommended limits and other attributes are within specified ranges. However, many normal situations probably meet these restrictions. The final attributes used in the public water supply index and in the fish and wildlife index are given in Table 7.15. Details about procedures used and the final value functions are in O'Connor [1973].

Table 7.15. O'Connor's Final Attributes in the Water Quality Indices

Public Water Supply	Fish and Wildlife
Fecal Coliforms	Dissolved Oxygen
Phenols	Temperature
Dissolved Solids	pH
pH	Phenols
Flourides	Turbidity
Hardness	Ammonia
Nitrates	Dissolved Solids
Chlorides	Nitrates
Alkalinity	Phosphates
Turbidity	
Dissolved Oxygen	
Color	
Sulfates	

7.7.8 Examining Foreign Policy

What are the advantages and disadvantages to the U.S. of a Mideast agreement sought to ensure the continued availability of Mideast oil and an increased production to meet the world demand? An exploratory policy analysis done by Decisions and Designs, Inc.* examined how a multiattribute decision analysis might clarify the reasoning and simplify the presentation of conclusions for such a complex problem.

The first phase of the analysis produced a flexible decision model and used it initially to evaluate three sharply different negotiating strategies regarding a possible Mideast agreement. A "base option" involving no change now or later in U.S.-Mideast policies was used primarily as a reference point for purposes of comparison. A maximum option involved an agreement which went most of the way toward what certain Mideast oil-producing countries want. A moderate option was an intermediate strategy reflecting a moderate change in U.S. policy,

---

*Decisions and Designs, Incorporated is an independent research and development company located in McLean, Virginia specializing in decision analysis for the United States Government and industry. Much of their work is devoted to problems involving multiple objectives.

which would be attractive to the Mideast oil-producing countries but not politically difficult for the U.S.

The decision model evaluated the impact of various negotiating postures on Mideastern oil supply and the associated political and economic costs and gains to the U.S. Specifically, the attributes concerned balance of payments, the way Western Europe and Japan would perceive a Mideast agreement, the impact on U.S.-foreign relations, the resulting public sentiment in the U.S., and finally, the effect an agreement would have on other oil producers.

Various sub-models were used to elicit probabilistic judgments and preference assessments at differing levels of complexity and aggregation. The uncertainty side of the analysis was based on judgments elicited from policy makers and substantive experts. Alternative approaches used direct unconditional assessments of oil volume, joint assessments of volume and price, and indirect assessments conditioned on possible political developments. Where different approaches led to inconsistent results, those inconsistencies were resolved by interacting with the respondents.

The preferences used in the problem were solicited from policy analysts charged with making recommendations. For a first analysis, the utility function chosen was additive. The single attribute utility functions for attributes such as "oil volume" were constructed in the manner described in Chapter 4. Tradeoffs were addressed by eliciting statements like, "All other factors held constant, an increase in Mideast oil supply to the U.S. of from .5 to 2.5 million barrels a day at $12 a barrel is indifferent to a gratuitous saving of $4 billion in the federal budget (independent of its level)."

The next phase of the ongoing decision analysis used the model developed, with several variations, to explore a much richer set of realistic options and to update continually the inputs in the light of changing circumstances or perceptions of individual decision makers. More details can be found in Brown and Peterson [1975].

7.7.9 Other Applications

As one can see from the examples described, there is a wide variety of settings in which multiattribute value or utility analysis is being employed. Still our collective experience is not so large that the theory and 'art' of such analyses is anywhere near standardized. Indeed, practically each new analysis contributes to the 'art' of assessing multiattribute preferences, if not to the theory aspects also. For space considerations, we have unfortunately not been able to review many such interesting 'groundbreaking' analyses.

Some of these are Bauer and Wegener's [1975] examination of urban development plans; Gearing, Swart, and Var's [1973,1974] measure of tourist attractiveness and selection of touristic projects for the Turkish Ministry of Tourism; Lorange and Norman's [1973] investigation of risk attitudes of Scandinavian shipowners; Gustafson Feller, Crane, and Holloway's [1971] development of a severity of pain index; Boyd, Howard, Matheson, and North's [1971] decision of whether to seed hurricanes; Dyer, Farrell, and Bradley's [1973] development of curriculum planning information for elementary school principles; and Collins [1975] evaluation of solid waste disposal alternatives in Southeastern Michigan. Huber [1974a] reviews a number of studies that used multi-attribute utility models. Two more general articles which survey the applications of decision analysis in industry are Brown [1970] and Longbottom and Wade [1973].

CHAPTER 8

AIRPORT DEVELOPMENT FOR MEXICO CITY: A CASE STUDY*

This chapter describes the application of decision analysis to a large scale public decision problem--selection of a strategy for developing the major airport facilities of the Mexico City metropolitan area. The purpose of discussing this study here is twofold. First, many of the techniques and procedures developed in earlier chapters of the book are utilized on a very important "typical" problem. Of course, it's typical of those one-of-a-kind strategic decisions which always concern many atypical aspects. Second, although the analysis stresses the value side of the multiattribute problem, it also deals with structuring the problem, aspects of modelling the possible impacts of various alternatives, and the larger framework within which the analysis occurred.

Many people contributed significantly to the study. It was done in the summer of 1971 for the Government of Mexico under the auspices of the Secretaria de Obras Publicas (Ministry of Public Works) and directed by F.J.

---

* This chapter closely follows the development in, and at times takes sections almost verbatim from deNeufville and Keeney [1972] and Keeney [1973a].

Jauffred, Director of the Center for Computation and Statistics, and F. Dovali, Head of the Department of Airports. Richard deNeufville of Massachusetts Institute of Technology and the two of us were consultants assisting SOP on the project. The total time spent by the consultants on the project was fifty man-days.

8.1. THE PROBLEM

Rapid growth in the demand for air travel, combined with increasingly difficult operating conditions at the existing airport facilities compelled the Mexican Government to address the question: "How should the airport facilities of Mexico City be developed to assure adequate service for the region during the period from now to the year 2000?" This was the essential question addressed by the study team.

Our initial problem was not this one however. Two previous studies for developing the airport facilities of Mexico City had recommended very different alternatives. One concluded that the current airport, five miles from the city center should be greatly expanded,* whereas the other suggested moving all aircraft operations to a new airport to be built twenty-five miles north of the city.+

* See Ipesa Consultores and the Secretaria de Communicaciones y Transportes [1970].

+ See Secretaria de Obras Publicas [1967] or Wilsey y Ham de Mexico [1967].

Our initial charter was to evaluate the various alternatives, in light of this discrepancy, and to recommend the most effective program for airport development.

For this more limited development decision, one needed to be concerned with the following:

(1) the location of the airport (or airports);

(2) the operational policy defining which services are to be performed and where they will be located; and

(3) the timing for development of different airport facilities.

Because of severe environmental constraints, the two sites previously mentioned are the only ones adequate for a large international airport in the Mexico City metropolitan area. The configurations possible at either site, with respect to the runways for example, were not really significant in this particular problem.

Many different ways of operating the airports -- with substantial differences in the quality of service provided -- were possible, however. In particular, it was necessary to decide what kinds of aircraft activity (international, domestic, military or general) should be operating at each of the two sites.

The question of timing is very important, since failure to act at a given time may preclude future options. For example, land available now may not be available in the future when one might want to develop it. On the other hand, premature action can significantly increase total costs to the nation. The timing issue and operational policies were the most important aspects of this initial airport problem.

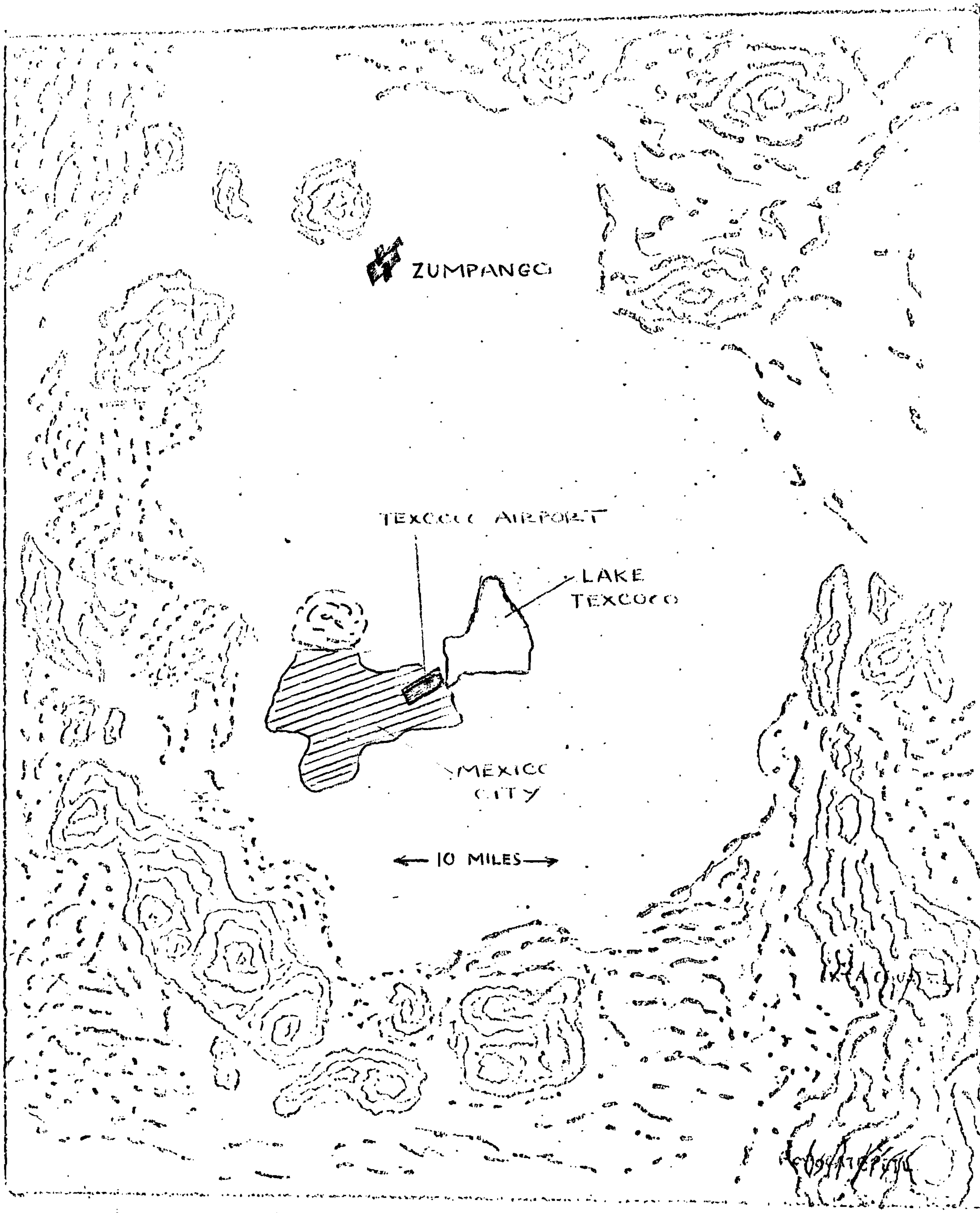
8.2. BACKGROUND INFORMATION

The existing airport is about five miles east of the central part of Mexico City, but still within the city limits on the edge of Lake Texcoco. The other site is 25 miles north of the city in an undeveloped farming area, near the village of Zumpango. The relative location of the two feasible sites is indicated in Figure 8.1.

8.2.1. The Physical Environment in Mexico City

Mexico City is situated at an altitude of about 7,400 feet in a valley ringed with high mountains ranging to over 17,000 feet above sea level. The mountains are very high in all directions except the northeast, where the range lowers to around 10,000 feet. Most flights entering or leaving the Mexico City area fly over these lower mountains to the northeast, although some do proceed through a smaller and higher pass to the south.

The maneuverability of the aircraft at high altitudes is low, especially in hot climates. This requires that the flight patterns over Mexico City be broader than usual and prevents aircraft from safely threading their way through mountainous regions. Thus there are considerable



GENERAL GEOGRAPHY OF THE MEXICO CITY METROPOLITAN AREA
Figure 8.1

restrictions on the usable airspace around Mexico City. This constraint, which principally affects the capacity of the Texcoco site, is serious since Mexico City already handles over 2 million passengers a year and ranks among the the busiest airports on the continent.

When the Texcoco Airport was organized in the 1930's, it was out in the country, but the population of the metropolitan area has grown at the rate of about 5% a year, passing from five million in 1960, to eight million in 1970. During this time, Texcoco has been surrounded on three sides by mixed residential and commercial sections. This has created problems of noise, social disruption, and safety.

Should a major accident occur on landing or takeoff toward the city it would likely cause hundreds of casualties. The area is densely populated and, for example, a large school is located under a flightpath only 500 feet from the end of a runway. Since the approach pattern passes directly over the central parts of the city, high noise levels affect many thousands of people. These noise levels are bound to persist for at least the next 15 years until "quiet" engines are developed and installed on all aircraft. In addition, major expansion at Texcoco could result in displacements of up to 200,000 people. A compensating advantage for the Texcoco site is that major facilities already exist. However they do not meet the standards found in the major airports of other large developed countries.

The location of Mexico City on a former lake bed makes construction especially expensive at Texcoco. Heavy facilities such as runways not only sink rapidly, but at different rates in different locations, depending on their loads. Each of the two major existing runways at Texcoco require levelling and resurfacing every two years. Such repairs closed down half the airport for four months when they were done in 1971. Because the Zumpango site is on higher and firmer ground, it is not expected to have the same kind of difficulties.

Access to the airport by ground transportation appears to be reasonable for both sites. The Texcoco site is near the main peripheral highway which can distribute traffic around the suburbs. It is not, however, especially well connected to the center of the City, to which one has to proceed through congested city streets. The Zumpango site has the clear disadvantage of being further away, but it can be linked directly to the tourist and business areas via an existing north-south expressway.

8.2.2. The Institutional Setting

The government of Mexico has been in the hands of a single party, the Partido Revolucionario Institucional, for almost forty years. Political power tends to be concentrated in the federal government and, for major decisions such as the location of the capital's airport, in the President himself. Any decision about a new airport during 1970-76 will require the approval of President Luis Echeverria. The debate about this decision has been carried on by three major governmental bodies:

(1) The Secretaria de Obras Publicas, SOP, (the Ministry of Public Works);

(2) The Secretaria de Communicaciones y Transportes, SCT, (the Ministry of Communication and Transport), and

(3) The Secretaria de la Presidencia, a body with functions similar to those of the Office of Management and Budget in the United States.

8.2.3. Previous Studies

Both SOP and SCT have commissioned rival large-scale studies of the airport problem within the past few years. The SOP study (SOP [1967], Wilsey y Ham de Mexico [1967]) done for its Department of Airports between 1965 and 1967, recommended that a new airport be built at Zumpango and that all commercial flights be shifted to this facility. The master plan then proposed was not adopted at that time.

The study commissioned by SCT in 1970 (Ipesa Consultores and SCT [1970]) resulted in a master plan for expanding the airport at Texcoco by adding new runway and terminal facilities. Interestingly, this report assumed that aircraft could take off away from the city toward the east, and could land coming into the city from the east in opposing streams of traffic aimed at adjacent parallel runways.

While this proposal "solves" the noise and displacement problems, its implications for safety are extremely serious at any significant level of traffic, and are unlikely to be acceptable for the expected volumes. This report assumed that "quiet" engines would completely eliminate any noise problems outside the airport boundaries by 1990. The SCT study was prepared and submitted during the closing months of the 1964-1970 administration of the previous President. It was not accepted in 1970. The Government of Mexico did, however, wish to resolve the issue. In early 1971 the new administration committed itself to a restudy. As stated by the President in his State of the Union Message of September 1, 1971, "Construction of a new international airport in the metropolitan area (of Mexico City) is also under study at this time." The study referred to is the one presented here.

8.3. EVOLUTION OF THE ANALYSIS

During the short three-month period --the summer of 1971-- in which we analysts were associated with the "airport" problem, it took on many forms. One might say that much of the time was taken defining the problem, but it seemed to be more than this. There wasn't a single problem, but many interrelated problems: What is the best manner to provide acceptable air service for Mexico City? How can one contribute to a reconciliation of differences of judgement, "facts", and opinion of independent government agencies concerned with airport development, in order to improve quality of information available to the decision makers? What strategies for developing the airport facilities are best in light of the financial and political realities facing the government? And so on. The focus of the analysis shifted as the SOP became more sensitized to issues we felt migt be important, as we became more familiar with the total environment in which this analysis was situated, and as segments of the study felt to be important were completed.

Because of the conflicting recommendations of previous reports, the original directive given to our colleagues in SOP was to evaluate various master plans for developing Texcoco and Zumpango. Therefore, this aspect of the problem had to be completed first. Before we entered the scene, SOP had been formulating this problem for a few months. The alternatives were specified and objectives and preliminary measures of effectiveness were defined. Our main effort concerned helping SOP (1) to synthesize the volumes of relevant information in the previous reports, as well as results from additional studies, and to indicate the degree to which various alternatives met objectives; (2) to meaningfully aggregate the effects occurring in different time periods; (3) to quantify a value structure appropriate for the problem; and (4) to develop a system for doing sensitivity analysis and for reporting results.

As this work progressed, the original problem began to be "solved", thus meeting the original directive and freeing the team to address other important issues. Perhaps the most crucial one was to attempt to reconcile the differences of viewpoint held by various parties, especially SOP and SCT, involved in airport development and operation.

8.3.1. An Attempt at Reconciliation through Shared Analysis

It is expected that impartial experts might disagree on many aspects of a complex analysis. It is crucial to know what aspects of the problem they agree or disagree on and why. For instance, there may be agreement on the structuring of the problem, but disagreement on the possible impacts of the various alternatives and disagreement on the value structure. The reasons may simply be that different experts have incomplete information or conflicting information or that traditional viewpoints due to political and professional orientation have been "cast in concrete". The decision analysis model, along with a graphical input-output display developed to assist in the analysis, seemed to offer a useful framework for analyzing these differences of opinion.

Input-output consoles were installed in offices of the study team, the Secretary and Under-Secretary of SOP, the Presidencia, and the President's own office. Our hope was that both SOP and SCT would agree on the basic framework for analyzing the airport problem and that this framework could then help highlight just where fundamental disagreements lay. The Presidencia would then be in a position to better understand the root causes of the different viewpoints, hear the rationalizations of each side, and then commission its own studies if required to clarify critical aspects of the problem. The SOP felt sure that if this reconciliation process were carried out, they would be shown to be right and they were prepared to be quite open -- even about their uncertainties on some inputs. A major problem, of course, lay in the fact that it was SOP who was suggesting the framework (not the Presidencia) and understandably, but regretfully from our point of view, the reconciliation process was never engaged.

Hence SOP had to proceed on a new tack. Clearly their minds were made up about the merits of Zumpango and now their efforts turned to amassing an argument that would convince the President and the Presidencia -- over and above the objections of the SCT. We thus proceeded in the preparation of an advocacy document that was meant to be impressively scientific. Some strange things happened.

8.4. THE STATIC MODEL

Because of the history of the previous studies, the alternatives, objectives, and measures of effectiveness for the static analysis were firmly specified by our clients, the Secretaria de Obras Publicas.

8.4.1. The Alternatives

The alternatives specified what types of aircraft would operate at each of the two possible sites over the rest of the century. In abstracting these, because of similarities in operating characteristics and functions, SOP had categorized aircraft as follows: International (I), domestic (D), general (G), and military (M). It was assumed that at any one time, each category of aircraft could operate at only one of the two sites.

To account for changes in operating arrangements over the thirty-year horizon while keeping the problem manageable, we decided to focus on the three years 1975, 1985, and 1995 as times when changes in the classes of aircraft operating at a site could occur. Thus, an alternative might be "develop the Zumpango site and move general aircraft to it in 1975, shift international to Zumpango in 1985, and operate all classes of aircraft at Zumpango by 1995." Of course, this discretization into three time epochs was done solely to keep the analysis tractable and the actual timing of moves would not be so constained in implementation. We are still discussing a rough-cut level of analysis with presumably more refined tuning coming at a later stage.

Notice that this gives us $(2^4)^3 = 4096$ alternatives. However, many of these were very similar in nature since, for instance, military operations accounted for less than five percent of the aircraft volume. Other alternatives defined as above were unreasonable. One would not move all operations from Texcoco to Zumpango in 1975 and back again in 1985, for example. In the final analysis, the total number of alternatives which were evaluated was approximately one hundred.

8.4.2. Objectives and Measures of Effectiveness

To evaluate the alternatives, one needs to specify some measures of effectiveness which explicitly describe their possible impacts on each of the important groups concerned about the problem. For this problem, the groups might be characterized as (1) the government, as builder and operator of the airports, (2) users of the air facilities, and (3) nonusers. Based on the previous reports of SOP and SCT and lengthy discussions the following six objectives were selected by SOP.

(1) Minimize total construction and maintenance <u>costs</u>;

(2) Provide adequate <u>capacity</u> to meet the air traffic demands;

(3) Minimize the <u>access time</u> to the airport;

(4) Maximize the <u>safety</u> of the system;

(5) Minimize <u>social disruption</u> caused by the provision of new airport facilities; and

(6) Minimize the effects of <u>noise</u> pollution due to air traffic.

Although there is obviously much overlap, the first two objectives account for the government's stake as operator;

objectives two, three, and four for the user's; and the last three objectives for the nonusers. Measures of effectiveness for these objectives were defined as follows:

$X_1 \equiv$ total cost in millions of pesos; with "suitable" discounting;

$X_2 \equiv$ the practical capacity in terms of the number of aircraft operations per hour;

$X_3 \equiv$ access time to and from the airport in minutes, weighted by the number of travelers from each zone in Mexico City;

$X_4 \equiv$ number of people (including non-passengers) seriously injured or killed per aircraft accident;

$X_5 \equiv$ number of people displaced by airport development; and

$X_6 \equiv$ number of people subjected to a high noise level, in this case to 90 CNR or more.*

Clearly, these six measures of effectiveness are not unique or completely comprehensive. For instance, air pollution considerations are absent. However, SOP felt the list did include all the important factors (other than political factors, prestige, etc., which we will discuss later on in

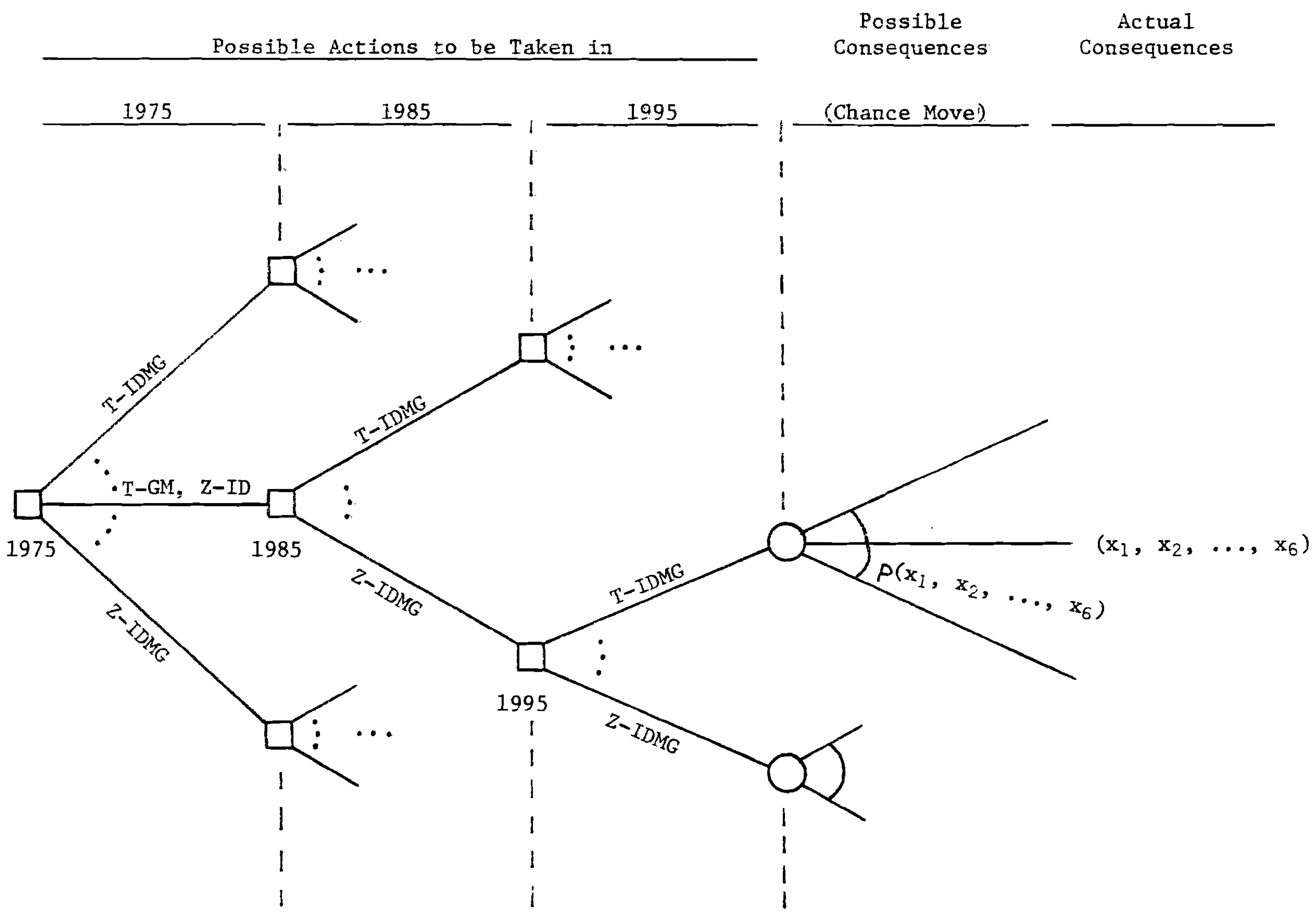
* The Composite Noise Rating, CNR, is a standard index of noise which combines decibel level and frequency of occurence. The 90 level was selected by the SOP Department of Airports.

this chapter) for evaluating effectiveness of the proposed alternatives.

8.4.3. The Basic Decision Model

The basic model is illustrated by the decision tree in Figure 8.2. An alternative is specified by defining what classes of aircraft will operate at which site in each of the three time epochs. As a result of the alternative chosen and events which occur (e.g., demand changes), a consequence $(x_1,x_2,...,x_6)$ will eventually result. However, at the time the decision must be made, uncertainties about this consequence for each possible alternative must be quantified by a probability distribution over the consequences.

The most important point to note about this model is that the alternatives are master plans. They are not designed to adapt to the unfolding of critical events (e.g., demand changes, technological changes, increasing environmental concerns of citizens, etc.) which might occur over the thirty-year period formally considered in the model. Clearly such considerations are essential to any analysis purporting to assist the Government of Mexico in deciding which actions to take in airport development. This was done in the dynamic analysis of options available in 1971 described in Section 8.8. There were two main reasons for first completing a formal analysis of this static problem:



Note: The notation T-GM, Z-ID means operate General and Military aircraft at Texcoco and International and Domestic aircraft at Zumpango.

THE BASIC DECISION MODEL

Figure 8.2

(1) the original request to study the "airport problem" required identifying discrepancies between previous studies, both of which were static analyses, and

(2) without such a study, SOP was very vulnerable to potential criticism of the analysis for excluding the details of such considerations.

The complete description of the probabilistic assessments are given in Section 8.5 , the preference structure is described in Section 8.6 , and the computer input-output along with the results of the anlaysis are given in Section 8.7.

8.5. SPECIFYING THE POSSIBLE IMPACTS OF EACH ALTERNATIVE

The probabilistic assessments were made using the volumes of relevant information from previous studies, the results of parallel studies being conducted by SOP, and the professional judgment of administration within the Mexican Government connected with airport construction, operation, and maintenance. Both reports for SOP (SOP [1967], Wilsey y Ham de Mexico [1967]) and SCT (Ipesa Consultores and SCT [1970]) contain many volumes including detailed demand studies for future air travel, soil mechanics and engineering studies at possible sites, pollution studies considering noise effects, analysis of ground traffic and airport access interaction, cost estimates and projections for various considered airport alterations, etc.. To help in the cost estimates, for each of the sixteen arrangements for aircraft operation at the two possible sites, in each of 1975, 1985, and 1995, general construction plans were outlined indicating where runways, support facilities, and access facilities would have to be built. These plans were used to translate the feasible alternatives specified in Section 8.4 into designs meaningful to airport planners and government officials.

To gain insight on the impacts of alternatives, various experiments were conducted by the SOP. One, designed to gather data on access times, involved dividing Mexico City into ten zones on the basis of residents' pattern of airport patronage, and then studying the driving times to the two airport sites from each zone in different weather conditions at different times of day, etc. This data on travel times and usage characteristics provided the information necessary to assess reasonable distributions for access times for the various alternatives.

In a similar way, detectors were located at various spots in the city to determine the noise levels caused by aircraft. By analyzing current and projected flight paths, superimposed on aerial photos of the city, and the population densities of the affected areas, one acquired a good indication of the noise impacts of various alternatives. These were used in assessing distributions for the number of people subjected to specified noise levels.

By superimposing the various plans for construction on aerial photos of the city, one could easily identify the areas in which people would have to be relocated given that a particular alternative were adopted. The population of those areas was tabulated providing information for assessing the number of people who would be displaced.

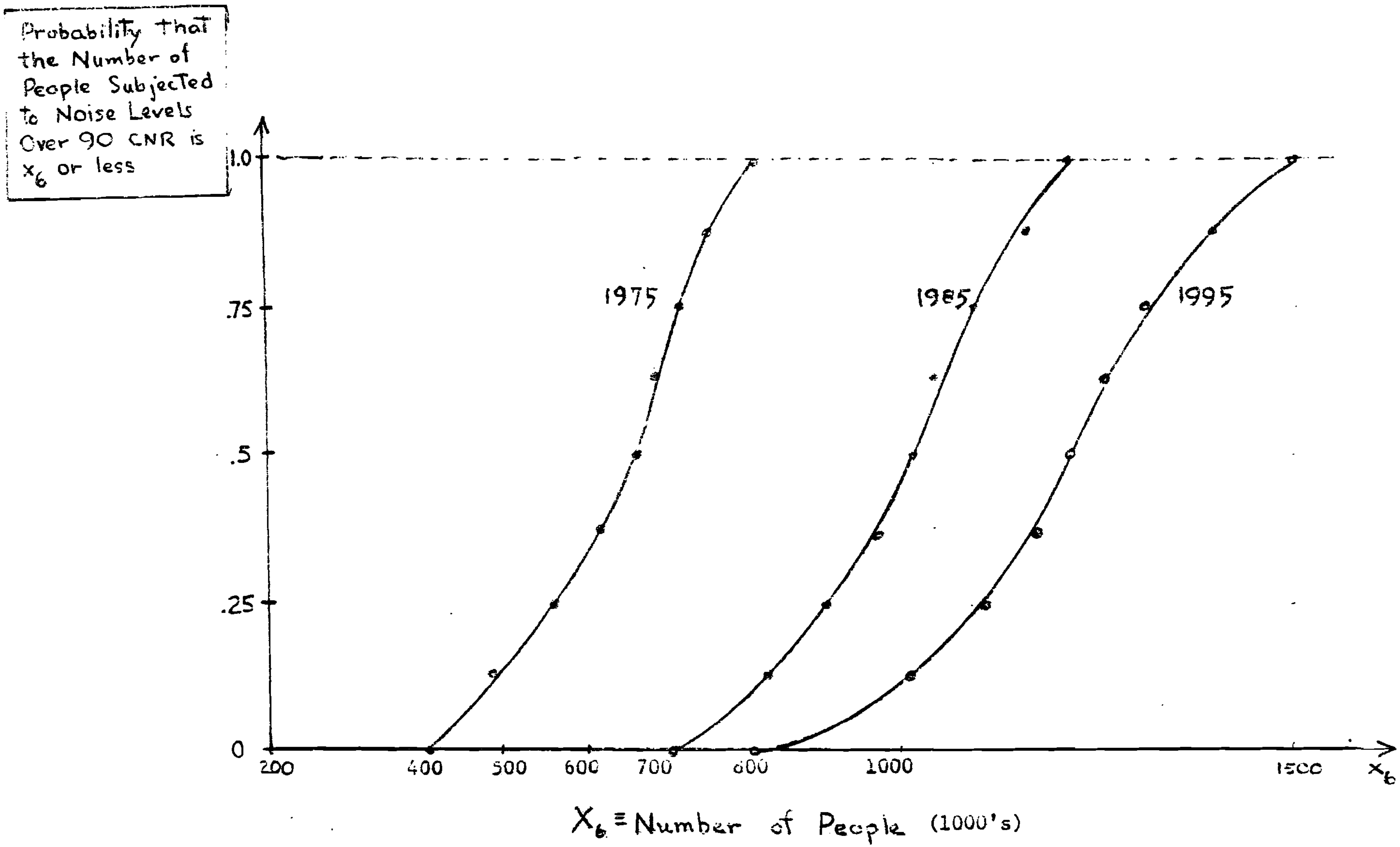
The results of all the previous studies and the data of the concurrent experiments of SOP needed to be integrated to provide meaningful estimates of the impacts of various plans. This integration was done using the professional judgment and experience of members of the Secretaria de Obras Publicas, including the Director of Airports , who is responsible for building and maintaining all the airports in the country of Mexico, the director of the Center for Computation and Statistics, and members of their staffs. The assessments were made in group sessions, where differences in judgments were discussed to arrive at a consensus. The fact that there were no problems in reaching a consensus can probably be attributed to a number of factors: all the professionals had the same information available, all had similar technical training in engineering, they were accustomed to working with each other and knew how each other thought, and the subordinates tended to agree with their superiors.

Having said how in general the probabilistic assessments were conducted, let us get to the specifics. First, the single year assessments will be described, and then the time effects will be accounted for.

8.5.1. One-Year Assessments

The probability density functions were assessed using the fractile method described in Raiffa [1968]. Let us use Figure 8.3 to illustrate the method by example. Consider the possible 1975 noise impact of the operating arrangement "all classes of aircraft at Texcoco." First, the maximum and minimum number of people subjected to 90 CNR or greater was specified as 800,000 and 400,000. Next to 0.5 fractile was evaluated as 640,000. This meant, in the judgment of SOP, the probability that the number of people impacted by 90+ CNR, denoted by X_6^{75}, would be less than 640,000 is one-half. Said another way, it is equally likely that the number of people subjected to the high noise level will be less than or greater than 640,000. The interval between 400,000 and 640,000 was then divided into equally likely parts by choosing the 0.25 fractile as 540,000. The 0.75 fractile was 700,000. Finally, each of the quartiles were divided into equally likely parts in a similar manner.

The fractiles which were assessed are indicated by the dots on Figure 8.3. and the smoothed lines are the cumulative probability distributions describing possible noise impacts for the "all Texcoco" option in years 1975, 1985, and 1995. For any given year, the probability that the impact is between any two adjacent fractile points should be the same, namely 0.125. Thus, to check consistency of the assessments, we asked SOP if in fact their



Assessed Distributions of the Number of People Subjected to Noise Levels Over 90 CNR With the "All Texcolo" Option

Figure 8.3

judgemental probabilities of falling into any of the eight ranges of impact were the same. SOP adjusted their assessments until no more discrepancies could be found. Figure 8.3 indicates the final adjusted curves.

One might ask what are the basic uncertainties which must be considered when assessing the possible noise influence of each airport. First, there is the uncertainty of the population in the flight path area. Current population is known rather accurately, but there is more uncertainty about the population in the future. There is uncercainty about when noise suppressors for jet engines will become operational and incorporated on most jets and about the level of impact of such suppressors. And there is uncertainty about the volume of air traffic in future years. Previous SOP and SCT studies, census figures, SOP experiments, etc., all provided useful information on these basic uncertainties. This information was both formally and informally used by SOP in making their combined assessments for the possible noise impacts.

8.5.2. Incorporating Time Effects

Each of the measures of effectiveness needed to account for the impact over the thirty-year period to the year 2000. Different adjustments seemed appropriate for different measures as indicated:

Costs. The costs that were considered in the model included building and maintenance, but excluded operating costs since it was felt these would be approximately the same for any alternative. As is normal practice for SOP, the present value of the costs was taken as the time dependent attribute of importance. The discount rate used was twelve percent, the standard for the Mexican Government. Sensitivity analysis indicated the choice of a discount rate was not critical for identifying effective strategies.

Noise. For noise, the average number of people annually subjected to aircraft noise levels above 90 CNR was used as the measure of effectiveness. This assumes that it is equally undesirable to have one person subjected to these noise levels for two years or to have two different people subjected in the different years. Furthermore, it assumes the undesirability to an individual of a certain noise level in any year is the same.

Safety. As previously mentioned, safety is measured in terms of the number of people killed or seriously injured per air crash. To adapt this, we chose the average number of people killed or seriously injured per crash averaged over the thirty-year time period. Clearly this measure does not account for the different likelihoods of crashes with various arrangements. SOP was aware of this and of the need to make adjustments to account for this factor. However, they felt it was not prudent to formally include the likelihood of crashes in the model, and chose instead, to make adjustments of the impact per crash in the sensitivity analysis to indicate the effect of differential crash likelihoods.

Access Time. For access time, an average of the possible access times in the various years weighted by the expected number of users in those years was used. This assumes each trip to or from the airport by any individual in any year is as important as any other such trip and that one's preferences for the various access times are stationary over time.

Social Disruption. By reasoning that on the average it would be just as undesirable for a random individual to be moved from his home due to airport development in one year as any other year, we chose the total number of such people displaced to be the measure of social disruption for the analysis.

Capacity. Capacity (maximum possible operations/hour) could not be aggregated in any reasonable way to combine impacts in the different years. This is due mainly to the fact that the relative desirability of various levels of capacity would be very different in different years since demand would probably be larger in later years. Increasing capacity from 80 to 100 in 1975 may be worth very little, since the additional capacity would rarely be needed. However, this same change in 1995 could be extremely important. Thus in the thirty-year model, separate measures of effectiveness for the capacity of 1975, 1985, and 1995 were included.

8.5.3. The Thirty-Year Assessments

By aggregating the three yearly assessed impacts for each measure of effectiveness, except capacity, in the manner just described, we could calculate the probability density functions over the measures to account for impact over time. For instance, with noise, if we difine

$$X_6 \equiv \frac{X_6^{75} + X_6^{85} + X_6^{95}}{3} \tag{8.1}$$

where X_6^i is the number of people subjected to noise levels over 90 CNR in year i, then by using the probability distributions assessed for the X_6^i for a particular strategy, it is straightforward to derive the probability distribution for X_6. This represents what we've taken to be the overall impact of a particular strategy in terms of noise.

8.5.4. Probabilistic Independence Assumptions

In conducting the assessments over one attribute at a time, we were explicitly assuming that for each alternative, the six attributes were probabilistically independent. For some of the attributes, this assumption seems appropriate. For instance, for any given alternative, noise and access time considerations are probably independent of the other attributes. On the other hand, safety considerations may be dependent on capacity, for instance. The lower the capacity, the more often the airport will be operating under hazardous conditions.

The more important assumption with regard to these assessments was that impacts in separate years were probabilistically independent conditional on the given alternative. This is clearly not true. For instance, for the "all Texcoco" alternative, if we found that 800,000 people were subjected to high noise levels in 1975, we would likely feel that more people will be affected by noise in 1985 than we would have if 400,000 had high noise levels in 1975.

Our analysis was designed in an iterative fashion. First simplifying assumptions (e.g. probabilistic independence) were adopted with the intention at a later stage of recycling back with more realistic assumptions. It turned out, however, that the delicacy taken in modelling the probabilistic part of our analysis was not a critical factor since other considerations dominated, and if we had more time, we would have dressed up the probabilistic analysis to be more credible to the reader. But it would have been mere "window dressing" because the action recommendations we finally suggested could not have been reversed by acknowledging the joint dependence of the random variables involved. It would not have been too difficult to incorporate this complexity -- if not analytically, at least through a simulation mode of analysis. It simply was not worth it in this case.

One could argue that given the oversimplifying probabilistic assumptions and the insensitivities, it might have been just as accurate and simpler to use point estimates of the impacts rather than probability distributions. In retrospect, this seems quite reasonable. However, this does not avoid any of the assumptions made in our analysis, and in addition, no account is made for the possible uncertainty of impact for the single attributes. Our approach forces an explicit recognition of this uncertainty by the decision makers. Also, before our analysis, the lack of sensitivity of the types of effective strategies to the attribute levels was not known. A sensitivity analysis using point estimates could have indicated this, however. The strongest reason for maintaining the detail of using probability distributions was that SOP wanted to avoid potential criticism of the analysis due to exclusion of the uncertainties.

8.6. ASSESSING THE MULTIATTRIBUTE UTILITY FUNCTION

Once we had probability assessments which adequately described the impact of alternate strategies in terms of our six measures of effectiveness, the next step was to assess a utility function $u(x_1,x_2,\ldots,x_6) \equiv u(\underline{x})$ over these measures. Proceeding as suggested in Section 6.6 , we began by exploring the decision maker's preference structure in a qualitative manner. This was to build up SOP's and our own experience in thinking directly about $(x_1,x_2,\ldots,x_6)$ consequences, but more importantly, to ascertain whether any of the preferential independence or utility independence assumptions discussed in earlier chapters were appropriate for this problem. Then we formally verified a sufficient set of such assumptions which allowed us to define for each i, $i=1,2,\ldots,6$, a conditional utility function u_i over X_i and then to construct u as a function of the conditional utility functions. That is,

$$u(\underline{x}) = f[u_1(x_1),\ u_2(x_2),\ldots,u_6(x_6)], \tag{8.2}$$

where f is scalar valued. To specify $u(\underline{x})$, the six u_i's and necessary scaling factors were assessed.

The utility assessments incorporated the best professional judgments of both the SOP Director at Airports and the SOP Director of the Center for Computation and Statistics, and members of their staffs. A serious attempt was made to analyze the problem from the point of view of the Government of Mexico.

8.6.1. The Assumptions

To refresh our memory, let us briefly and informally review the concepts of preferential independence and utility independence. Recall that preferential independence concerns only ordinal preferences and no probabilistic elements are involved. Partition the set of attributes into Y and Z. If the rankings of consequences, which differ only in the level of attribute Y, are the same regardless of the fixed level of attribute Z, when Y is preferentially independent of Z.

Utility independence, on the other hand, concerns the cardinal preferences of the decision maker. If the rankings of all lotteries, which differ only in the possible levels of Y which may occur, are the same regardless of the fixed levels of Z then Y is utility independent of Z.

8.6.2. Verifying the Assumptions

Let us illustrate how we verified the preferential independence assumptions used in our work. As an example, consider whether safety X_4 and noise X_6 are preferentially independent of the other attributes. First, we fixed the other attributes at a desirable level and asked what amount of safety x_4 was such that (x_4; 2,500) was indifferent to (1;1,500,000). That is, x_4 people seriously injured or killed given an accident and 2,500 people subjected to high noise levels is indifferent to one person seriously injured or killed and 1,500,000 subjected to a high noise level. After "converging," the amount of x_4 was chosen as 300. The exact number is not important for verifying the assumptions, but our interest is in whether it changes as the other four attributes vary. So we next set these four attributes at undesirable levels and asked the same question and again elicited 300 as response.

Then we asked if this would in general be true for any values of the other four attributes, and the response was "The answer would always be the same given the other attributes were in a static condition." In fact the respondent stated this would be the case concerning any tradeoffs between safety and noise. Hence, we concluded safety and noise were preferentially independent of the other attributes.

By going through identical procedures, we verified that capacity and cost were preferentially independent of the remaining attributes, as was displacement and access time. By this time, the man answering the questions, who was an assistant to the Director of Airports, was in a position to state that ordinal preferences over any two attributes did not depend on the amounts of the other attributes. These conditions were then also verified with other staff members of SOP, including the Director of Airports.

The same general approach was used in verifying the utility independence assumptions -- that X_i was utility independent of its complimentary set $\overline{X}_i$ for all $i = 1,2,\ldots,$ 6. As an example, consider whether access time X_3 was utility independent of $\overline{X}_3$. The other five attributes were set at desirable levels, and the conditional utility function over access time from 12 to 90 minutes (the range originally specified by SOP) was assessed. We found 62 minutes indifferent to a 50-50 lottery yielding either 12 or 90 minutes. Then we changed the amounts of the $\overline{X}_3$ attributes to less preferred amounts and repeated the question. Again, an access time of 62 minutes was indifferent to a 50-50 lottery yielding either 12 or 90 minutes. A general question indicated this would be true for any fixed amounts of $\overline{X}_3$. We found that relative preferences for any consequences and lotteries involving uncertainties only about access time were indeed independent of the other five attributes.

This condition was verified for all six attributes with both the Director of Airports and members of his staff. In all of these verification procedures, an attempt was made not to lead the respondent to answers he would not have arrived at otherwise. Our opinion is that this was done successfully. Since preferences may vary with time, such questioning of the same people may lead to different conclusions at another point in time. However, the preferences indicated by the individuals questioned appeared to represent their "true" preferences at that time, and hence, the assumptions we made were deemed appropriate for the problem.

8.6.3. Forms of the Utility Function

The main theoretical results used in obtaining the utility function were Theorems 6.1 and 6.2 given in Section 6.3. Informally, these results state that if each pair of attributes is preferentially independent of its complement and if each attribute is utility independent of its complement, then $u(x_1,x_2,\ldots,x_6)$ is either an additive or a multiplicative function of the component utility functions $u_1(x_1),u_2(x_2),\ldots,u_6(x_6)$. Actually, as indicated in Section 6.3 , this same result is implied by a much weaker set of assumptions - i.e. only one attribute X_i needs to be utility independent of its complement and each pair of attributes including X_i needs to be preferentially independent of its complement. Therefore, many of the assumptions that were verified are redundant, and they can be thought of as consistency checks on the appropriateness of our results.

The exact form of the utility function u, scaled zero to one, is

$$u(x_1,x_2,\ldots,x_6) = \sum_{i=1}^{6} k_i u_i(x_i) + k \sum_{i=1}^{6} \sum_{j>i} k_i k_j u_i(x_i) u_j(x_j)$$
$$+ k^2 \sum_{i=1}^{6} \sum_{j>i} \sum_{n>j} k_i k_j k_n u_i(x_i) u_j(x_j) u_n(x_n)$$
$$+ \cdots + k^5 k_1 k_2 k_3 k_4 k_5 k_6 u_1(x_1) u_2(x_2) \cdots u_6(x_6), \qquad (8.3)$$

where u_i is a utility function over X_i scaled from zero to one, k_i is a scaling factor for u_i, and k is another scaling constant. Each k_i must be between zero and one and can be interpreted as the utility u assigned to a consequence with all its attributes except X_i set at their least preferable amount and X_i set at the most preferable amount.

The value of k can be found from the values of the k_i's. When $\Sigma k_i = 1$, then $k = 0$ and (8.3) reduces to the additive form

$$u(x_1,x_2,\ldots,x_6) = \sum_{i=1}^{6} k_i u_i(x_i). \qquad (8.4)$$

When $\Sigma k_i \neq 1$, then $k \neq 0$ so we can multiply each side of (8.3) by k, add one to the results, and factor to get the multiplicate form

$$k\ u(x_1,x_2,\ldots,x_6)+1 = \prod_{i=1}^{6} [kk_i u_i(x_i) + 1]. \qquad (8.5)$$

8.6.4. Assessing the u_i's

Each of the single attribute utility functions was assessed using the techniques discussed in Chapter 4. Let us illustrate this by assessing preferences for access time.

The first step involved obtaining maximum and minimum values for access time. From probabilistic assessments of SOP, we found that the range should go from 12 minutes to 90 minutes, where shorter access times were preferred to longer ones. Thus, to remain consistent with our scaling convention where the utility functions ranged from zero to one, we set

$$u_3(90) = 0 \qquad (8.6)$$

and

$$u_3(12) = 1. \qquad (8.7)$$

From questions to check whether X_3 was utility independent of $\overline{X}_3$, recall that we found 62 minutes for sure was indifferent to a lottery, which we will denote by $\langle 12,90 \rangle$, yielding either 12 or 90 minutes, each with probability 1/2. Hence, the utility assigned to 62 minutes, the certainty equivalent for the lottery, is

$$u_3(62) = 0.5u_3(12) + 0.5u_3(90) = 0.5 \qquad (8.8)$$

Since 62 is greater than the expected access time 51 of the lottery <12,90>, this original assessment indicated that the utility function might exhibit risk aversion. In this context, risk aversion means that the expected amount $\frac{x_3 + x_3'}{2}$ of any lottery $\langle x_3, x_3' \rangle$ would be preferred to that lottery. By asking a couple of questions including specific lotteries and then one concerning the general case, we found that the decision makers were risk averse in the attribute access time. This implied the utility function would be concave as indicated in Fig. 8.4.

By asking more questions to find certainty equivalents of additional lotteries, other points on u_3 were specified. For instance, we found 40 minutes indifferent to <12,62> and 78 minutes indifferent to <62,90>, so

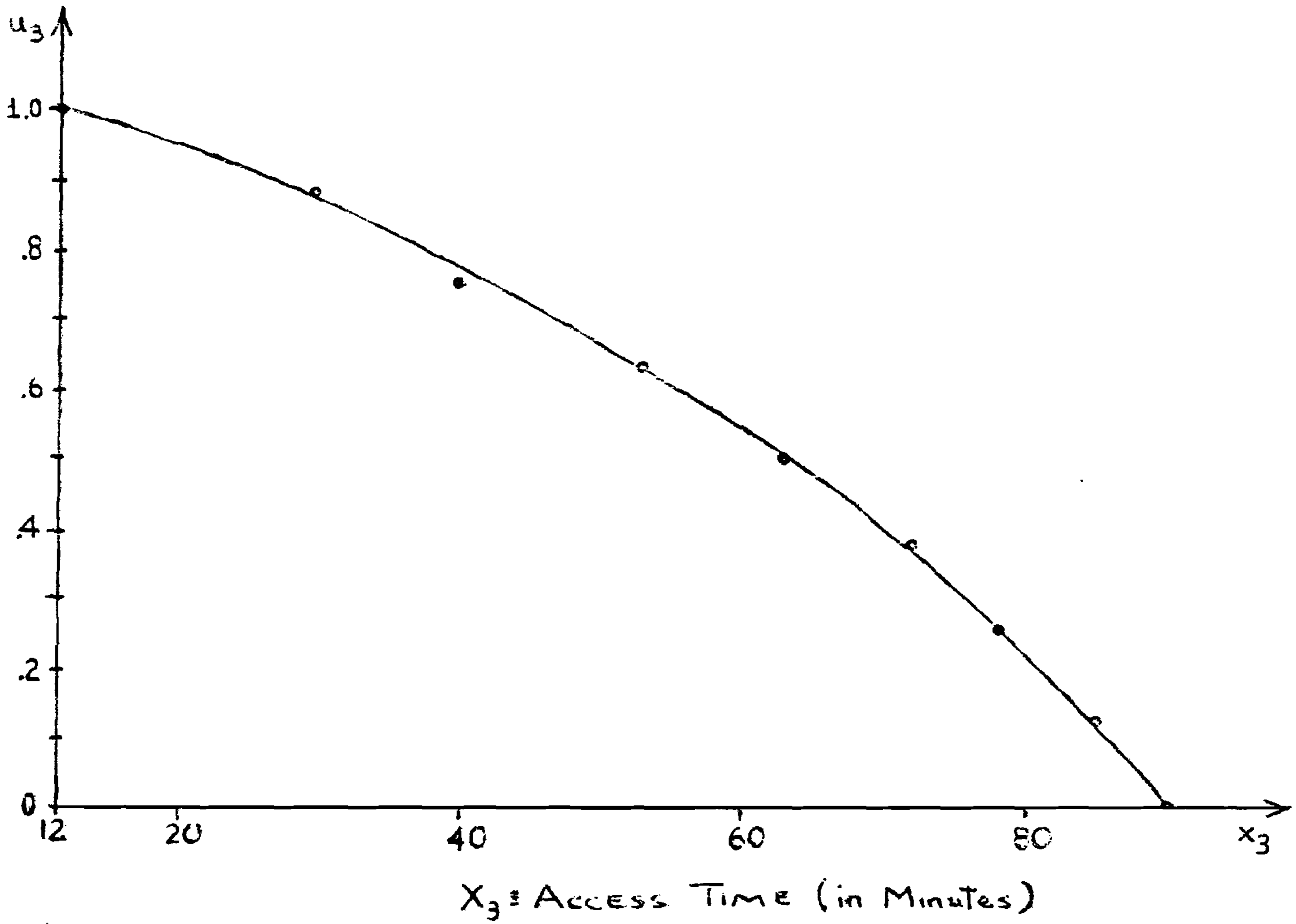
$$u_3(40) = 0.5u_3(12) + 0.5u_3(62) = 0.75, \tag{8.9}$$

and

$$u_3(78) = 0.5u_3(62) + 0.5u_3(90) = 0.25. \tag{8.10}$$

Then an exponential utility curve was fitted to the empirically assessed points.

At this stage, we did not immediately try to ascertain and exploit "higher order" risk properties such as decreasing risk aversion. Such properties represent rather fine tunings in a multiattribute utility function relative to the scaling constants "weighting" the levels of the different attributes and more basic properties such as monotonicity and risk aversion of the separate u_i's. If later in the analysis,



The Utility Function for Access Time

Figure 8.4

it had turned out that the precise form of some of the u_i's were important, we would have returned to this aspect and reiterated our evaluation of alternatives. This did not happen to be the case.

Procedures similar to those described above were also used to assess utility functions for cost, safety, displacement, and noise. The results are illustrated in Figure 8.5. However, as mentioned earlier, no single measure was found to combine capacities in different years. Thus, it was necessary to assess the capacity utility function u_2 differently.

Although the general shapes of the utility functions for access time, cost, and noise seem intuitive, the fact that the curves for safety and displacement are linear is not. For instance, concerning safety, one might expect that since governments usually abhor large numbers of deaths resulting from single tragedies the utility function for safety would be risk averse. The reason for this attitude is usually the political impact due to such tragedies. However, our measure of effectiveness in this problem was not meant to capture these political factors. Roughly speaking, if one says each life is equally important, then alternatives with the same expected number of people killed or seriously injured should be equally undesirable in this respect. This was the attitude taken by SOP in the assessments, and so u_4 is linear.

It was important, before proceeding, to do consistency

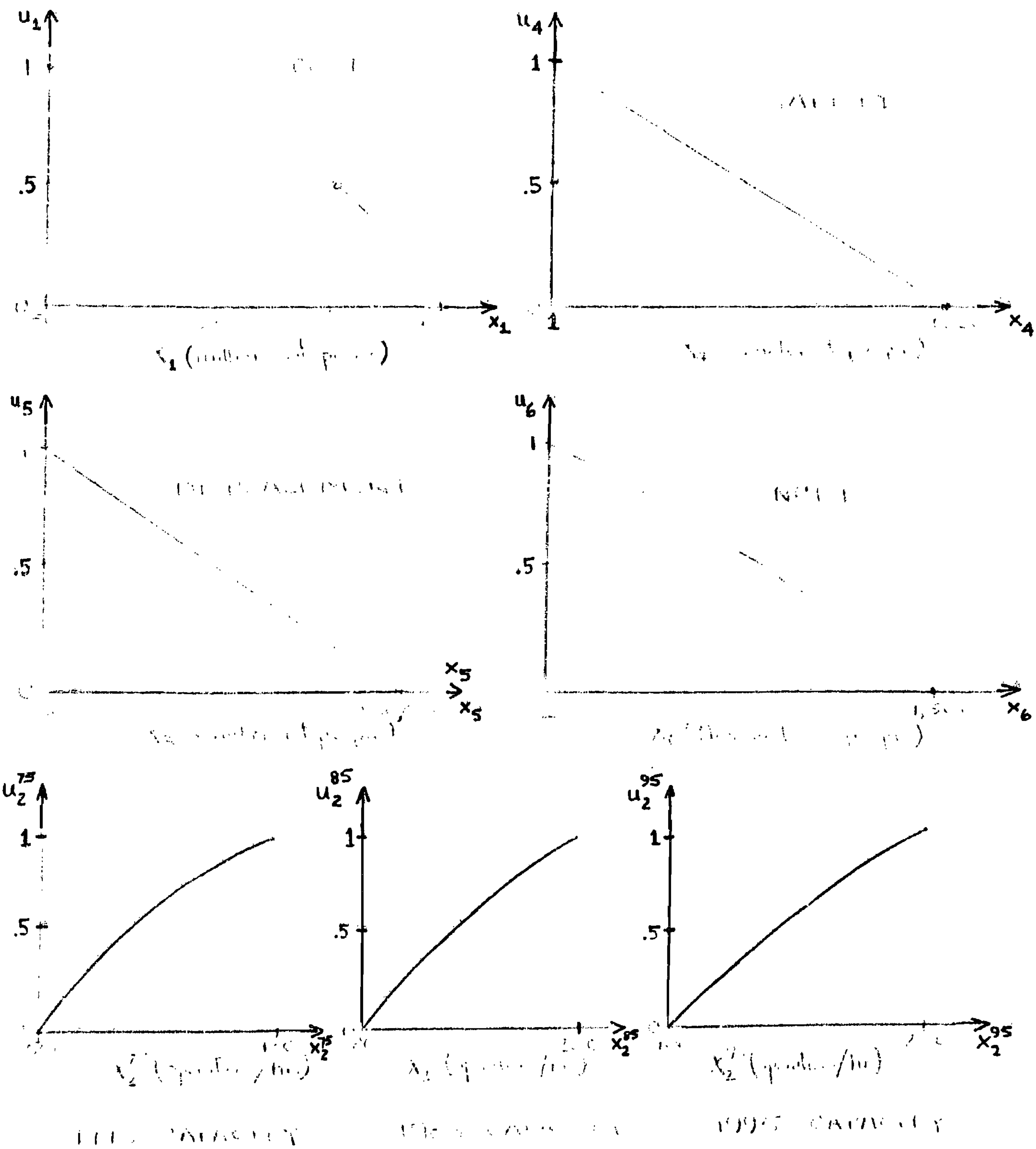


Fig. 8.5 For The Mexico City Airport Study

checks on the reasonableness of the exponential and linear utility functions. This was done by asking additional questions about the decision maker's preferences, and comparing his responses to the implications of the "fit" utility function. When they were consistent with each other, we developed more confidence in the utility function. When they were inconsistent, the inconsistencies were discussed, and part or all of the assessment repeated.

8.6.5. The Capacity Utility Function

Capacity x_2 is a vector $(x_2^{75}, x_2^{85}, x_2^{95})$, where x_2^{75} is the capacity in 1975, etc.

The first step in assessing u_2 was to identify the minimum and maximum possible airport capacities for each year: 1975, 1985, and 1995. There were 50, 80, 100 and 130, 200, 250 operations per hour respectively. Clearly more capacity in any given year was preferred to less capacity, so to scale u_2 from zero to one, we set

$$u_2(50,80,100) = 0 \qquad (8.11)$$

and

$$u_2(130,200,250) = 1. \qquad (8.12)$$

It was verified that each pair of capacity attributes was preferentially independent of the third, and that each attribute was utility independent of the other two.

Thus, we know from Theorems 6.1. and 6.2. that either

$$u_2(x_2^{75}, x_2^{85}, x_2^{95},) = \sum_{j=75,85,95} c_j u_2^j(x_2^j) \qquad (8.13)$$

or

$$cu_2(x_2^{75}, x_2^{85}, x_2^{95}) + 1 = \prod_{j=75,85,95} [cc_j u_2^j(x_2^j) + 1], \qquad (8.14)$$

where the u_2^j are the utility functions over X_2^j assessed on a zero to one scale as illustrated in Fig. 8.5 and c and the c_j are scaling constants. Notice that the forms of (8.13) and (8.14) are analogous to the utility functions expressed in (8.4) and (8.5). Since the following discussion concerns how the k_i's and k in (8.5) are assessed, we will not indicate the assessment of the c_j's and c in (8.14) as the procedures are indentical.

8.6.6. Assessing the k_i Scaling Factors

To illustrate the technique for assessing the k_i scaling factors, let us take cost X_1 as an example. We asked the decision makers to compare a consequence with cost at its most preferred amount, and all the attributes at their least preferred amount, to a lottery yielding the consequences with all attributes at their most preferred amount with probability p or the consequence with all attributes at their least preferred amount with probability 1-p. The object is to find the value of p, call it p_1, such that the decision maker is indifferent between the lottery and the consequence. Then, as shown in

Section 6.6, by using $u(\underline{x})$ from either (8.4) or (8.5) and equating expected utilities, k_1 must equal p_1.

Using this procedure involving questions concerning lotteries, we arrived at an initial estimate for the k_i values. Then we used nonprobabilistic questions as consistency checks. For example, we set all attributes at their least desirable level and asked, "Would you prefer to have capacity or cost changed to its most desirable level?" Capacity was the response implying k_2, the coefficient of capacity utility, had to be greater than k_1, the coefficient of cost utility. Then we found a level of capacity, call it $\underline{x}_2^I$, which was indifferent to the best level of cost, denoted by $x_1{}^*$. Then using either (8.4) or (8.5), we see that $k_2 u_2(\underline{x}_2^I)$ must equal k_1. Since we have u_2 assessed, this gives us a relationship between k_1 and k_2. Pairwise comparison of the k_i's in this manner provided many consistency checks, redundant with others, and forced a readjustment of the k_i values. After several iterations, we ended up using the values of k_i indicated in Table 8.1.

8.6.7. Assessing Parameter k.

Since the sum of the k_i is 1.89, we know the utility function is multiplicative rather than additive; it is additive only if $\Sigma k_i = 1$. Therefore the value of k in (8.5) must be determined by evaluating (8.5) at $(x_1^*, x_2^*, \ldots, x_6^*)$ where x_i^* is the most preferred amount of X_i. This gives us

$$k\,u(x_1^*,x_2^*,\ldots,x_6^*) + 1 = \prod_{i=1}^{6}[kk_i u_i(x_i^*) + 1], \tag{8.15}$$

but from our scaling conventions, we know both $u(x_1^*,x_2^*,\ldots,x_6^*)$ is 1 and the $u_i(x_i^*)$ are all one so

$$k + 1 = (k\,k_1 + 1)\;(k\,k_2 + 1)\;\cdots\;(k\,k_6+1). \tag{8.16}$$

Since the k_i are known, parameter k can be evaluated by solving (8.16). As shown in the Appendix 6B, since $\Sigma k_i > 1$, the value which k must assume is the solution to (8.16) such that $-1<k<0$. Using the k_i values from Table 8.1, we found $k = -0.877$. Of course, if this were redone from scratch a new k would be found. But it would probably fall closer to $-.80$ (say) than to .00 or to +.80. In the final analysis, it is important to do sensitivity studies on k and the k_i's.

Table 8.1 SCALING FACTORS FOR THE MEXICO CITY AIRPORT STUDY

Attribute X_i	Scaling Factor k_i
X_1 = Cost	0.48
X_2 = Capacity	0.6
X_3 = Access Time	0.10
X_4 = Safety	0.35
X_5 = Displacement	0.18
X_6 = Noise	0.18

8.6.8. The Utility Function

Procedures identical to those just illustrated were used to evaluate the c_j and c in (8.14). It was found that $c_1 = 0.3$, $c_2 = 0.5$, $c_3 = 0.4$, and $c = -0.46$. These parameters, together with Table 8.1 and $k = -0.877$ and the utility functions illustrated in Figures 8.4 and 8.5 represent the information necessary to specify the utility function $u(x_1, x_2, \ldots, x_6)$. The next section describes how it was used.

8.7. THE ANALYSIS

A computer was programmed to assist in evaluating the alternatives. Computationally, the program was quite simple: given any set of probability distributions and a utility function, it calculated the expected utility for specified alternatives.

To keep the calculations at a reasonable number, as mentioned earlier, many alternatives were eliminated before going through expected utility calculations. For instance, since military aircraft represent a relatively insignificant amount of the total air traffic, most alternatives differing only in terms of the airport for military operations were not considered separately. Secondly, alternatives which shifted certain types of aircraft from the Texcoco site to Zumpango and back again at a later date were excluded.

8.7.1. The Input-Output Display

Graphical input-output consoles were used as an efficient and accessible system for sensitivity analyses and communicating results of the study. This capability was used daily by the SOP, and could also be used by the other interested parties to examine the relative merits of alternative developmental policies. The input-output system allowed any user to use his own probability and utility estimates for evaluating any specified alternatives. There were two options for doing this. Option 1 provided the standard estimates that SOP used in evaluating the alternatives on the console screen. To change these, one just typed in the changes over the SOP estimates. This option was particularly useful for sensitivity analyses. Option 2 allowed the user to enter his own estimates without seeing any others.

The probabilistic estimates of possible impact could be altered by changing the upper and lower bounds on these impacts. For instance, as illustrated in Figure 8.3, SOP's lower and upper bounds on the possible number of people subjected to noise above 90 CNR in 1975 were 400,000 and 800,000, respectively. Merely by typing on the console, one could look at the overall effect on strategy if these were 600,000 and 1,200,000.

To alter the utility function, one changed the scaling factors listed in Table 8.1. Because the meaning of these constants can be easily misunderstood (as discussed in section 5.9) and because of the difficulty in specifying a consistent set of estimates, a short subroutine was developed to assist the user. This routine essentially asked the user on the screen the same questions that we asked SOP in initially assessing the scaling constants. Once a reasonable consistency was achieved among the k_i's, the constant k in (8.3) was calculated. If k=0, the additive form (8.4) was used, and if k≠0, the multiplicative utility function was used to evaluate strategies. As was the case with the general shape of the probability densities, the individual utility functions u_i could not be changed by graphical input-output. These changes required adjustments in the programs. However, although important, these changes represent fine tunings relative to the options provided for graphically.

Another particularly useful feature of the computer program was a routine which calculated certainty equivalents. Using this routine, the overall possible impact of any alternative could be reduced to an equivalent impact described by a vector of certainty equivalents. Since we assumed probabilistic indepencence* and first-order utility independence (i.e., each X_i is utility independent of its complementary set), from the marginal probability distribution of X_i and the component utility function u_i, it is possible to define the certain equivalent $\hat{x}_i$ by

$$u_i(\hat{x}_i) = E[u_i(\tilde{x}_i)], \ i = 1,2,\ldots,6. \qquad (8.17)$$

Notice that the certainty equivalent $\hat{x}_i$ is independent of the possible impacts on other attributes. Also notice that the certainty equivalent vector $\underline{\hat{x}} = (\hat{x}_1,\ldots,\hat{x}_6)$ does not commit one to any determination of the scaling constants k_i's or k.

If two alternatives A and B are reduced to certainty equivalent vector impacts $\underline{\hat{x}}_A$ and $\underline{\hat{x}}_B$, it is easy to check for dominance. Also, for example, one could investigate exactly how large a change in the impact on attribute X_i of alternative A would be required before it would be less preferred than alternative B.

* If k=0 (or close to zero), then u can be taken to be (approximately) additive and only the marginal probability distributions are of relevance. If k≠0, and joint probabilistic dependence is warranted, then the analysis by certainty equivalents must be considerably modified. One could, however, employ the notion of "<u>conditional</u> certainty equivalence" to some advantage. This was not done.

8.7.2. Effective Strategies

Of the alternatives we did evaluate using expected utility, the top ten according to SOP are indicated in Table 8.2. In the table, the expected utilities are calculated on a scale from zero to one-hundred, where zero utility was assigned to a hypothetical alternative generated by taking the least desirable probability distribution for each attribute from the set of all alternatives. The utility value of one hundred units was assigned to a hypothetical alternative generated by taking the most desirable probability distribution for each attribute from the set of all alternatives. On this scale, the alternative of keeping all aircraft in Texcoco in all three years has an expected utility 5.20.

By looking at Table 8.2, it is clear that two types of strategies are effective. One type might be categorized as the "all Zumpango" alternative and represents building a major new airport at Zumpango as soon as possible. The alternatives in the table involving both International and Domestic aircraft operating at Zumpango in all three years make up this category. The other type of effective strategy is the "phased development at Zumpango" characterized by either International or Domestic aircraft operating in Zumpango in 1975 and then both by 1985 and 1995. All strategies which included keeping a part of the International or Domestic traffic operating out of Texcoco through 1985 did not appear competitive in terms of effectiveness with

Table 8.2 THE BEST TEN ALTERNATIVES

Alternative						Expected Utility	Rank
1975		1985		1995			
Z	T	Z	T	Z	T		
D	IMG	ID	MG	ID	MG	91.23	1
IDMG	-	IDMG	-	IDMG	-	90.90	2
I	DMG	ID	MG	ID	MG	90.79	3
ID	MG	ID	MG	ID	MG	89.30	4
ID	MG	IDMG	-	IDMG	-	88.10	5
ID	DMG	ID	MG	IDMG	-	86.75	6
I	DM	IDMG	-	IDMG	-	86.55	7
IG	IM	IDMG	-	IDMG	-	86.19	8
DG	IMG	IDMG	-	IDMG	-	86.17	9
D		IDMG	-	IDMG	-	85.60	10

To help read the table, the alternative ranked 1 is Domestic aircraft at Zumpango with International, Military, and General aircraft at Texcoco in 1975; and International and Domestic at Zumpango with Military and General at Texcoco in 1985 and 1995.

the two types of strategies outlined above. Of course, these expected utility evaluations depend on two types of judgmental inputs: probability and utility assessments. The ones we used were those of officials of SOP and presumably, if the same analysis were to be made with inputs from officials of the SCT, another ranking of strategic alternatives would result. But more about these reconciliation problems later.

8.7.3. Use of the Analysis

As we indicated earlier in the chapter, the original purpose of the work described here was to identify effective strategies - as measured by our six measures of effectiveness - for developing the airport facilities of Mexico City. It was not to indicate what action should be taken by the Government of Mexico in 1971 to meet its needs. Once the "effective strategies" had been identified, the problem shifted to this second question: What action should be initially implemented?

So far, the formal analysis has included only master plans defining actions for a thirty year period. A more appropriate course would seem to be to make some initial decision and then, based on subsequent events, to revise strategies as necessary. Furthermore, any study which is designed to aid in the selection of an airport development policy for Mexico City must include factors such as political preferences and community priorities. This was the task undertaken in a dynamic analysis of development strategies to be discussed in the next section.

8.8. THE DYNAMIC ANALYSIS

The purpose of the dynamic model was to decide what governmental action should be taken in 1971 which would best serve the overall objective of providing quality air service to Mexico City for the remainder of the century. This model assumed the second step in the decision process could be taken in 1975 or 1976, at the end of the current President's six-year term. The action taken then would depend both on the action taken now and the critical events which might occur in the interim. Our analysis of the dynamic model was much less formal than the one developed for the static model, primarily because of time pressures and the general complexity of the situation.

8.8.1. Alternatives for 1971

We first identified (using common sense) the reasonable alternatives available to the government in 1971. These alternatives differed in the degrees of commitment to immediate construction at the two sites. We chose only four levels of commitment (minimum, low, moderate, and high) giving us the 16 alternatives exhibited in Figure 8.6. Actually, each nominal case in the figure represents a class of specific alternatives. The idea was to do a

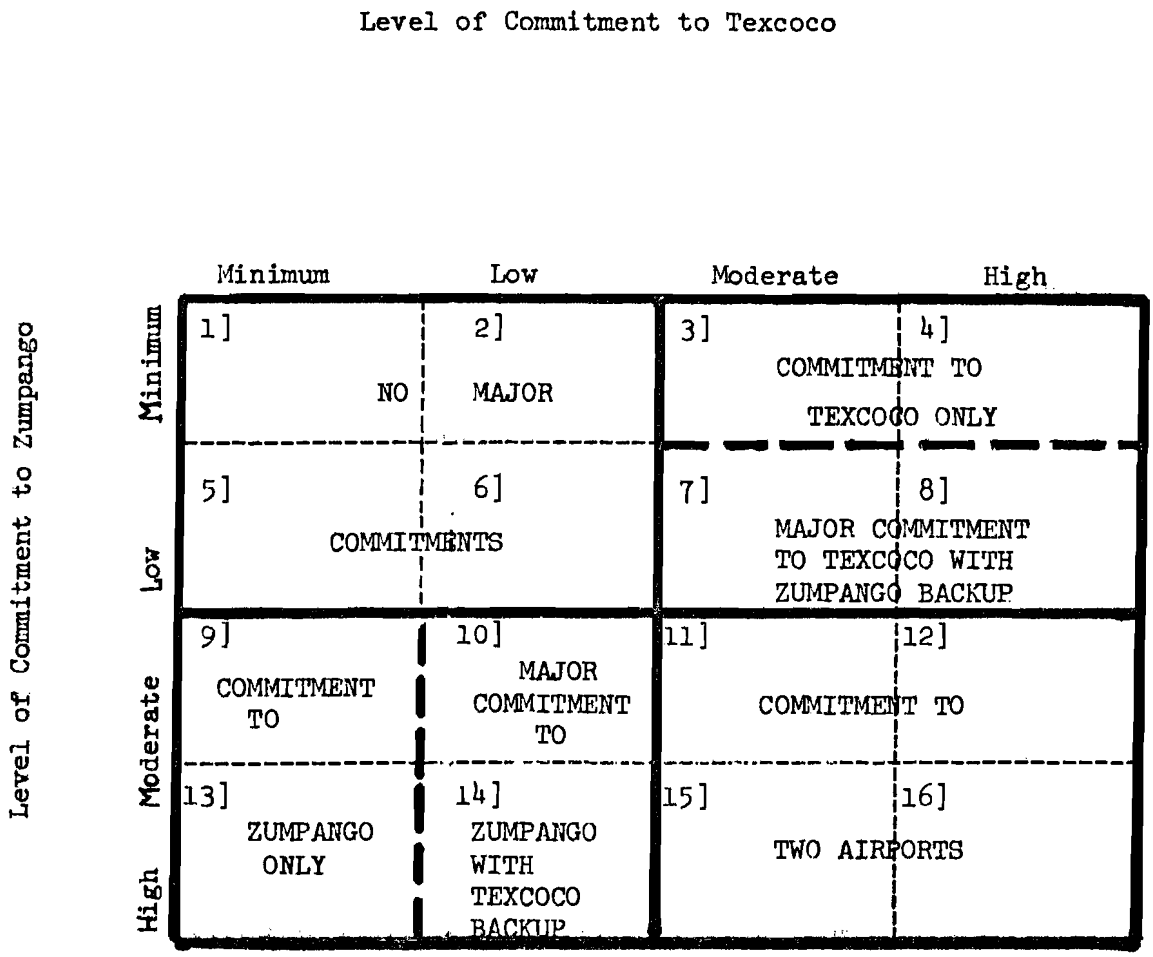


Figure 8.6: THE 16 NOMINAL DYNAMIC ALTERNATIVES FOR 1971

first-cut analysis to decide which classes of alternatives were sufficiently viable to be examined in more detail. It should be noted that the two strategies defined by the static analysis could be compatible with all the nominal dynamic options except 11,12,15, and 16.

The next step involved defining what was meant by the alternatives in some greater detail. Briefly summarized, the alternatives at Texcoco (for the period 1971-1975) were defined as follows:

Minimum - maintenance and introduction of safety equipment only;

Low - extend the runways, upgrade support facilities such as terminals, do all routine maintenance and introduce new safety equipment;

Moderate- in addition to that done with a low strategy, buy and prepare land for building a new runway and expand passenger facilities;

High - build a new runway and passenger facilities, improve the airport access - in short, build a totally new airport at Texcoco.

Similarly, for Zumpango, we defined the commitment levels:

Minimum - at most, buy land at Zumpango;

Low - buy land, build one jet runway and very modest passenger facilities;

Moderate- buy land, build a first jet runway and plan others, build major passenger facilities, and construct an access road connection to the main Mexico City highway;

High - build multiple jet runways, major passenger facilities, and access roads - that is, build a large new airport at Zumpango.

8.8.2. Objectives

We identified four major objectives that were important in choosing a strategy for airport development: effectiveness, political consequences, externalities, and flexibility of the various alternatives. The components of the "effectiveness" attributes are indicated by the six measures of effectiveness covered in the static model. The political consequences were those important to the President - since he was the principal decision maker - involving the political effects which would be felt by SOP, by SCT, and by the Presidencia. Flexibility concerned the range of options open to the President at the second stage of the decision-making process: what freedom would he realistically have at the end of his tenure in modifying his earlier 1971 stance after learning about the intervening uncertain events. Finally, all other important considerations were lumped together as "externalities." These included the amount of access roads needed, the distribution of federal expenditures between the Mexico City region and the rest of the country, the distribution of expenditures for airports and other uses, regional development away from central Mexico City, and the national prestige associated with new airport facilities.

8.8.3. Possible Scenarios

To gain insight into the meanings and implications of each of the classes of alternatives, detailed scenarios were outlined for each. These included: (1) the consideration of important and critical events which could occur in the period 1971-1976, and possibly affect the best strategy in 1976; (2) the likelihood of their occurrences; (3) the strategic reaction to each intervening event-complex; and (4) the possible eventual consequences for each act-event-reaction path. The events involved safety factors and air disasters; shifts in demand in terms of both passengers and aircraft; technological innovations, such as noise suppressors, better runway construction on marshy ground, etc.; changes in citizen attitudes toward the environment; and changes in priorities, such as national willingness to have government funds used for major airport construction. Figure 8.7 depicts a schematic representation of one possible scenario.

In each of the scenarios, the manner in which the 1971 strategy should be altered in 1976 to account for the critical events listed above was defined. For instance, if one originally chose strategy six, then a reasonable response to increased numbers of landings and thus decreased safety, in addition to increased consideration about the impact of noise and air pollution in Mexico City, would be to hasten the building at Zumpango and make it the

Table 8.3: PRELIMINARY EVALUATION OF PLAUSIBLE GOVERNMENTAL OPTIONS FOR 1971 BY RANK ORDER

ALTERNATIVE	ATTRIBUTES										
	Flexibility	Political Effects on				Externalities due to					Effectiveness
		Presi-dencia	SOP	SCT	Overall	Pres-tige	Reg. Dev.	Bal. of Fed Exp.	Roads	Overall	
2	1	1	8	2	3	4	4	1	1	3	7
3=4*	7	4	5	1	4	1	4	6	3	7	8
5	2	3	6	4	3	3	3	2	1	1	3
6	3	2	7	3	2	3	3	3	1	3	1
9*	4	6	3	6	5	2	2	4	2	2	4
10	5	5	4	5	1	2	2	5	2	4	5
13*	6	8	1	8	7	1	1	6	4	5	2
14*	8	7	2	7	6	1	1	7	4	6	6

*Alternatives dominated by 2, 5, 6, or 10 on overall ranking of four major attributes siderations and externalities, the assessments on the components were first carried out, and then the overall ranking for these attributes was established. The ranking of the alternatives according to effectiveness was provided directly by results of the static model.

The results of the first ranking effort are shown in Table 8.3, where the smaller numbers represent the better rankings. From this table it can be seen that alternatives 3=4, 9, 13, and 14 are each dominated by others on the basis of their overall rankings for the four main measures of effectiveness. Alternative 6, for instance, is better than alternative 14 in terms of all four of the measures. Hence alternative 14 — and likewise alternatives 3=4, 9, and 13 — can be dropped from further consideration. The alternatives which were not dominated are those represented by the nominal cases 2, 5, 6, and 10. It is important to note here, however, that before we actually discarded dominated alternatives we engaged in a devil's advocate procedure: we tried to give the benefit of reasonable doubt to the impending noncontenders to see whether they could be resurrected to a place of contention. They could not.

8.8.5. Final Analysis of Dynamic Options

To refine the analysis of the possible governmental decisions, it was necessary to define the remaining contending alternatives more precisely. This was done as follows:

2 - At Zumpango, do no more than buy land for an airport. At Texcoco, extend the two main runways and the aircraft apron; construct freight and parking facilities, and a new control tower. Do not build any new passenger terminals.

5A - Build one jet runway, some terminal facilities and a minor access road connection at Zumpango. Buy enough land for a major international airport. At Texcoco, perform only routine maintenance and make safety improvements.

5B - Same as alternative 5A, except buy just enough land for the current Zumpango construction.

6 - Extend one runway at Texcoco and make other improvements enumerated in alternative 2. Buy land for a major international airport at Zumpango, and construct one runway with some passenger and access facilities.

10 - Same implications for Texcoco as alternative 6. Build two jet runways with major passenger facilities and access roads to Zumpango.

These five alternatives were ranked in the manner previously described. The results are given in Table 8.4. Proceeding as before, we can quickly see that alternative 6 dominates 10, and alternatives 2, 5A and 6 all dominate 5B. Thus the three remaining viable alternatives are 2, 5A, and 6.

Table 8.4: FINAL EVALUATION OF GOVERNMENTAL OPTIONS FOR 1971

ALTERNATIVE	ATTRIBUTES			
	Flexibility	Political Effects	Externalities	Effectiveness
2	1	4	4	3
5A	2	3	3	2
5B*	4	5	5	4
6	3	1	1	1
10*	5	2	2	1

*Alternatives dominated by 2, 5A, or 6 on overall ranking of four major attributes.

The relative advantages of these three options were, finally, subjectively weighed by the SOP personnel as follows. Alternative 6 ranks better on effectiveness, externalities, and political considerations than either 2 or 5A. Although it is worse in terms of relative flexibility, it does allow the President to react effectively to all the critical events which might occur between 1971 and 1976, when the second stage of the airport decision could be made. Hence, in the opinion of the members of SOP working on this problem, alternative 6 was chosen as the best strategy.

8.9. IMPACT OF THE RESULTS

Based on the 1965-1967 study by SOP which recommended moving the International Airport to Zumpango as soon as possible, as well as indications early in this study, it was clear that some members of SOP held the opinion that a major move to Zumpango was still the most effective strategy. The static analysis, using SOP's own estimates and preferences, indicated a phased development involving a gradual shift toward Zumpango appeared equally as good. Once political considerations, flexibility of the policy, and externalities were accounted for along with effectiveness in the dynamic analysis of alternatives open to the government in 1971, it was evident that the "phased development at Zumpango" policy was better than an "all Zumpango" policy.

Looking at the implications of their evaluations, the SOP staff was very surprised and bewildered. Using their own preferences over measures of effectiveness they knew were relevant for a realistic set of options, they agreed that the two alternatives, thirteen and fourteen, which were most consistent with their so strongly held position, were completely dominated. Note also that the position of SCT, being most consistent with alternative 3=4 was also dominated.

This glaring inconsistency had a profound impact on many individuals within SOP. They rethought their position, analyzing in their own minds how this "strange" implication came about. As they understood the implication better, they gained some confidence in the result. With the final analysis of non-dominated alternatives and additional group discussions of the dynamic analysis, SOP adopted a new flexible position, exemplified by an initial choice of option six in 1971. Thus a very strange thing happened: an analysis undertaken for unabashedly advocacy purposes (i.e., to justify going all-out to Zumpango) turned out to convince the sponsors of the analysis that perhaps a more flexible stance was really in the best interest of Mexico.

8.9.1. The Ensuing Political Process

SOP recommended a "phased development" strategy to the President in December 1971. Specifically, it was suggested that land be acquired at Zumpango, that a major runway and modest terminal facilities be planned for construction during President Echeverria's term. It was also proposed that he reserve until 1976 a more detailed decision on how the airport facilities for Mexico City should be developed. This recommendation represented a major change in SOP's posture from the 1967 study. The previous recommendations of SOP were for master plans specifying what should be done at various points in time over the next thirty years without regard to the unfolding of relevant uncertain events. Thinking in terms of adaptive dynamic strategies rather than in terms of master plans played a pivotal role in our analysis.

As the last stage of our consulting activities, we, in collaboration with our clients, examined in some detail the steps that had to be taken in order to implement the newly developed stance of SOP. This required developing a strategy for the planning of technical documents, for informal presentations to key government agencies, for private meetings, and for possible announcements. Since we were not certain of the reactions of SCT and the Presidencia we mapped out some contingency plans which themselves were more in the spirit of an adaptive dynamic analysis than of a master plan. We are sure that you will understand that this chapter, however, is not the place to discuss the details of these politically sensitive considerations.

The analyses described in this chapter were completed in early September, 1971. In late 1971, Ing. Jauffred and

Ing. Dovali, together with Secretary Bracamontes of SOP presented the basic ideas of this study to the President of Mexico. Members of SCT and the Presidencia, including the respective secretaries of these ministries, were also present at this meeting.

The meeting, perhaps needless to say, did not eliminate all differences of opinion concerning the two basic points of view--remain at Texcoco or more to Zumpango--positions that had long been established. After the meeting, the President requested that SOP, SCT, and the Presidencia work out philosophical disagreements on the airport issue, as well as technical and financial details of further developing the airport facilities of the Mexico City area. Because of its complexity and importance, the process of 'working out the details' is very time consuming. By mid-1974, no concrete decision had been made. However, the winds seemed to blow a bit differently in 1974 than in 1971. In the earlier year, the basic issue was whether the main Mexico City Airport should be at Texcoco or Zumpango. In 1974, the issue seemed to involve when the Zumpango site would be the main airport--next year, in five years, or twenty. Support for this came from the fact that land for an airport at Zumpango was exappropriated by SOP, who holds this authority, in early 1974. Presumably, whatever decision evolves by the Government of Mexico will be done with greater awareness of the relative influence of the different attributes and of the dynamic issues.

ACKNOWLEDGEMENT

It was a pleasure for us, the authors, to work with F.J. Jauffred, Director of the Center for Computation and Statistics, and F. Dovali, Head of the Department of Airports, and their staffs in the Secretaria de Obras Publicas, and with Richard deNeufville of M.I.T., who added both engineering know-how as well as prodigious language skills to our consulting team. We felt that our colleagues in SOP contributed greatly on the integrity of the study by weighing seriously their answers to vexing value questions. Throughout they demonstrated a healthy skepticism at critical points, demanded meaningful explanations of the process of anlaysis, and were flexible enough to modify previously advocated positions.